Acquisition of Scalar Implicatures: When Some of the Crayons will do the Job[*]

Anna Verbuk
University of Massachusetts at Amherst

1. Introduction

Scalar implicatures, henceforth SIs, are a species of Quantity-based conversational implicatures that were first introduced in Horn (1972). SIs are generated in the following manner: by using a weaker scalar item, the speaker implicates that he does not know that any of the stronger ones hold. An example of an SI is provided in (1).

(1) Tiger wanted to eat hot soup. Monkey made Tiger warm soup.
SI: it is not the case that Monkey made Tiger hot soup.

The SI in (1) is based on a Horn scale <boiling, hot, warm>. Two major classes of scales employed in generating SIs are Horn scales and pragmatic scales. Horn scales are based on the ordering relation of entailment. Some examples of Horn scales are <and, or>, <all, most, many, some> and <impossible, unlikely, uncertain>. Pragmatic scales are based on an ordering relation other than entailment, such as level of description: <spaniel, dog, domestic animal>, part/whole: <house, room>, and stages: <send the letter, write the letter>. In order for an SI based on a pragmatic scale to arise, the scale must be salient in the given context (Hirschberg 1985).

The rest of the paper is organized as follows. In section 2, I discuss the Default account of SIs and introduce my own Context-based question-under-discussion account of SIs. In section 3, I introduce my acquisition questions and discuss my experiment testing between the two accounts. Concluding remarks are in section 4.

2. Accounts of SIs

Currently, there are three major classes of SI accounts: Horn's (1989) version of the neo-Gricean account, the Relevance Theory account (Wilson & Sperber 2004) and different versions of the Default account (Levinson (2000) and Chierchia (2004)). In this paper, I will be concerned with testing between the Default account of SIs and my own version of a Context-based account of SIs.

2.1 The Default Account

According to Levinson's (2000) version of the Default account, SIs belong to the class of utterance-type-meanings, which are based not on reasoning related to speaker-intentions but rather on general considerations of language use. The crucial difference between Grice's original view and the Default account is that Grice conceived of conversational implicatures as speaker-meanings rather than utterance-meanings. Because on the Default account SIs are viewed as utterance-type-meanings, the role of utterance context is minimized in this framework. On Levinson's (2000) account, the use of a scalar item *automatically* triggers the computation of a generalized conversational implicature (GCI) to the effect that a higher item on the relevant scale does not hold. Crucially, the context of the utterance is not taken into account in computing the SI. In contexts where the content of an SI is not relevant, the SI gets cancelled after having been computed.

According to Chierchia's (2004) version of the Default account, SIs are conceived of as semantic rather than pragmatic inferences. The scale associated with a given scalar item is part of its semantics. When an SI is being derived, the strengthened value of the given scalar expression is provided by the grammar. If α is an English expression, $[[\alpha]]$ is its plain value and $[[\alpha]]^S$ is its strengthened value that gets assigned recursively. The strengthened value of the scalar item is computed as soon as possible; maximal projections are shipped to the semantic module, where the strengthened meaning of the scalar item is computed. On Chierchia's (2004) account, an SI gets canceled in the following case; "whenever its (SI's *A.V.*) addition to a given context results in

[*] I would like to thank Tom Roeper, Chris Potts, Jill De Villiers, Laurence Horn, and the audiences at BUCLD 31, UUSLAW and LSA '07 for their advice and helpful comments. I am grateful to the children, parents and teachers at schools in the Amherst, MA area.

inconsistency, one falls back on the plain value" (Chierchia 2004: 18). Thus the prediction is that canceling an SI requires additional pragmatic reasoning.

2.2 My Context-based QUD Account of SIs

The context-based QUD account of SIs that I will argue for here is closest in spirit to Horn's (1984, 1989) version of the neo-Gricean account of SIs. On Horn's account, SIs are treated as inferences that belong to the class of speaker-type-meanings that are derived in contexts where they are relevant for the current purposes of the exchange. However, the notion of what it means to be relevant needs further clarification. According to Horn's (1984) account, SIs are relevant in neutral and upper-bounded contexts. Horn defines upper-bounded contexts as contexts where the stronger scalar item is relevant. Neutral contexts are defined as the ones where no information about the weaker or the stronger scalar item is provided. However, if one takes a closer look at the neutral contexts, one finds that SIs often do not arise in neutral contexts, as (2) illustrates.

(2) A: Where were you in October?
B: I was in Fresno. It's **warm** there in October.
Horn scale: <hot, warm>
No SI: it is not hot in Fresno in October.

At the same time, SIs may arise in neutral contexts, as (3) shows.

(3) A: What is the weather like in Fresno in October?
B: It's **warm** there in October.
Horn scale: <hot, warm>
SI: it is not hot in Fresno in October.

Since SIs may or may not arise in neutral contexts, a revision to Horn's (1984) account is called for. Note that on Grice's (1975) account and according to Horn's (1989) neo-Gricean account of SIs, SIs are computed iff they are relevant. However, the maxim of Relevance is the least studied and the least understood of Grice's maxims (Relevance Theory excluded).

Following Roberts (1996), I propose to reinterpret Grice's maxim of Relevance as Relevance to the question-under-discussion (henceforth QUD). Following Stalnaker (1979), Roberts (1996) views discourse as being organized around a series of conversational goals, and views the attempt to discover the way things are as the primary goal of discourse. Roberts argues that discourse may be conceived of as addressing a series of hierarchically structured QUDs. Consider Roberts' definition of what constitutes a move relevant to the QUD in (4).

(4) A move *m* is **Relevant** to the question under discussion *q*, i.e. to *last (QUD (m))*, iff *m* either introduces a partial answer to *q* (*m* is an assertion) or is part of a strategy to answer *q* (*m* is a question).
(Roberts 1996: 15).

Roberts' move *m* may be either explicit or implicit, whereby her definition entails that Relevance governs both what is said and what is implicated. I would like to propose that conversational implicatures in general and SIs in particular are computed iff they are relevant to the QUD. By reinterpreting the maxim of Relevance as relative to the QUD, one can straightforwardly account for contexts where conversational implicatures, SIs being a species thereof, do and do not arise. In (2), reproduced below as (5), the potential SI does not arise because it is not relevant to the QUD.

(5) A's QUD: Where were you in October?
B: I was in Fresno. It's warm there in October.
No SI: it is not hot in Fresno in October.

Since A's QUD in (5) concerns B's location, the potential SI concerning the weather in Fresno does not arise.

In contrast, in the dialogue in (3), reproduced below as (6), an SI does arise precisely because it is relevant to the QUD.

(6) A's QUD: What is the weather like in Fresno in October?

B: It's warm there in October.
SI: it is not hot in Fresno in October.

Since A's QUD in (6) concerns the weather in Fresno, the SI concerning the temperature in Fresno does arise. In view of these data, I would like to postulate the condition in (7),

(7) SIs are computed iff they are relevant to the QUD.

3. The Experiment

Experiments on the acquisition of SIs have been done by researchers working within the Default account framework (ex., Guasti et al. 2005) and by Relevance Theorists (ex., Papafragou et al. 2004). Given space limitations, I will not attempt to discuss previous experimental work in this paper in any detail. To date, children have not been tested on contexts where SIs are *not* generated; this is what my experiment does. My experiment tests between the Default account of SIs and my own Context-based QUD account of SIs.

3.1 Acquisition questions

In this paper, I will address the following acquisition questions.

(i) Are children sensitive to the role of context in licensing SIs from the start or do they go through a "default" SI computation stage?

(ii) How big a role does lexical learning play in determining the child's success on computing SIs based on a given scale?

(iii) Is the child's success on computing SIs predictable from the class of scale (Horn vs. pragmatic)?

3.2 Hypotheses

The Default account of SIs and my own Context-based QUD account of SIs make contrasting predictions concerning the roles of context and the type of scale in children's computation of SIs. According to the Default account, SIs based on Horn scales are computed automatically. Canceling an SI that does not get projected requires additional pragmatic reasoning. Thus the Default account gives rise to H_1.

(8) **H_1: Children start out by computing both the relevant and irrelevant SIs based on Horn scales.**

According to the Context-based QUD account, SIs based on Horn and pragmatic scales are computed by Gricean reasoning and **only** in contexts where they are relevant to the QUD. Therefore, irrelevant SIs are never computed. The Context-based QUD account gives rise to H_2.

(9) **H_2: Children start out by computing SIs that are relevant to the QUD but not the irrelevant SIs based on Horn scales.**

Next, I will discuss the hypotheses related to the role that the type of scale plays in the child's success on computing SIs based on this scale. In my experiment, I employed two classes of scales: Horn and pragmatic. The pragmatic scales that I employed were based on the part/whole ordering relation (ex., <body, paw>, <house, kitchen>). As far as the Horn scales are concerned, I used the quantifiers, connectives and two gradable adjective scales. The Default and Context-based QUD accounts make conflicting predictions concerning the two classes of scales. On Levinson's (2000) version of the Default account, SIs based on Horn scales belong to the class of *utterance*-type-meanings that are computed automatically, while SIs based on the part/whole ordering relation belong to the class of *speaker*-type-meanings that are entirely context-dependent. On Chierchia's (2004) version of the Default account, Horn scales are part of the semantics of lexical items, while pragmatic scales are not. While SIs based on Horn scales are derived automatically in the semantics, the traditional (neo)-Gricean approach to SIs based on pragmatic scales is not being disputed. Thus it is presupposed that SIs based on pragmatic scales are derived in the pragmatics, which implies that the hearer needs to infer the relevant scale from the context. It needs to be noted

that Foppolo, Guasti & Chierchia (in press) justly argue that children experience problems with computing SIs based on Horn scales because they have not learned which lexical items constitute these scales. Next, consider what prediction this version of the Default account makes with respect to children's performance on Horn vs. pragmatic scales. It is safe to assume that children will receive more exposure in the input to instances of Horn scales such as the connectives scale <and, or> rather than to highly idiosyncratic pragmatic scales such as <send the letter, write the letter>. Thus the child is predicted to form the relevant Horn scales and be able to automatically compute SIs based on them earlier than he will learn to infer idiosyncratic pragmatic scales from the context and compute SIs based on these by Gricean reasoning. In sum, different versions of the Default account give rise to H_3.

(10) **H_3: Children do better on computing SIs based on Horn scales than those based on the part/whole relation pragmatic scales.**

As far as the Context-based QUD account, it can make two different predictions concerning the child's performance on part/whole relation-based pragmatic scales and Horn scales. First, consider a view on which lexical learning does not play a significant role in the child's success on computing SIs. I am using the notion of lexical learning here in a very specialized sense of constructing a scale, such as the quantifier scale <all, most, many, some>. On this view, once the child has mastered Gricean reasoning, he will be able to compute SIs based on any scale. This version of the Context-based QUD account gives rise to H_4.

(11) **H_4: Children do the same on computing SIs based on Horn scales and those based on the part/whole relation pragmatic scales.**

Next, it will be useful to take a closer look at the types of scales that I employed in my experiment. The pragmatic scales that I used were based on the part/whole relation. It is safe to assume that even the youngest children tested, who were aged 4;3-5;11, have a *non-linguistic* knowledge of the part/whole relation. As I have noted, Horn scales are based on the relation of entailment. It has been argued at length in the pragmatic literature that entailment is not a sufficient condition on scalehood (Gazdar 1979, Atlas & Levinson 1981, Horn 1989). Some of the other conditions on scalehood are that scalar items need to have the same polarity, which rules out scales like <warm, cold>, scalar items need to belong to the same semantic field, which rules out scales like <regret, know> and scalar items need to be equally lexicalized, which rules out scales like <iff, if> and <humongous, big>. In experiments on the acquisition of SIs, it is sometimes found that the child's knowledge that a unilateral entailment relationship holds between members of a scale is not immediately translated into the ability to compute SIs based on a given scale (ex., Foppolo et al. (in press)). My interpretation of these findings is that the child has the knowledge of the conditions on scalehood, and is aware that entailment is only one of these conditions. Thus the child is being conservative and postpones forming a scale until she has received evidence to the effect that all of the conditions on scalehood have been satisfied.

Next, consider the second view on which lexical learning does play a significant role. As was argued above, part/whole relation-based pragmatic scales are likely to be easier than entailment-based Horn scales. If lexical properties of the Horn and pragmatic scales in question are taken into consideration, the Context-based QUD account predicts H_5.

(12) **H_5: Children do better on computing SIs based on the part/whole relation pragmatic scales than those based on Horn scales.**

3.3 Methods, Subjects and the Experimental Technique

Three independent variables were employed: (i) relevance of implicature to QUD (relevant / not relevant); (ii) type of scale (Horn, pragmatic); and (iii) the child's age (younger vs. older group). The dependent variable was the child's performance on computing SIs. Three conditions were employed: (i) an SI based on a Horn scale is relevant to the QUD; (ii) a potential SI based on a Horn scale is not relevant to the QUD; (iii) an SI based on a pragmatic scale is relevant to the QUD. A mixed design was employed. A total of four Horn scales were used: <all, some>, <and, or>, <wonderful, good> and <hot, warm>. Six scenarios on the basis of the four scales were used, the total being 24 scenarios. In 12 of the 24 scenarios, the content of the implicature was relevant to the QUD and in 12 it was not; the two sets of scenarios were minimal pairs. Crucially, the test sentence that did or did not generate the SI was identical in each of the minimal pairs. Every child received 12 scenarios based on Horn scales in 6 of which the implicature was relevant and in 6 of which it was not. (Every child was given only one member of every

minimal pair). In group one, children received 3 scenarios where the <and, or> scale was employed and the SI was relevant; 3 scenarios where the <wonderful, good> scale was employed and the SI was relevant; 3 scenarios where the <all, some> scale was employed and the SI was not relevant; and 3 scenarios where the <hot, warm> scale was employed and the SI was not relevant. In group two, children received 3 scenarios where the <and, or> scale was employed and the SI was not relevant; 3 scenarios where the <wonderful, good> scale was employed and the SI was not relevant; 3 scenarios where the <all, some> scale was employed and the SI was relevant; and 3 scenarios where the <hot, warm> scale was employed and the SI was relevant. In addition, 6 scenarios in which pragmatic scales generate SIs were used; every child was given all 6.

A total of 40 children aged 4;3-7;7 who were native speakers of English were tested. All of the children attended child care centers and elementary schools in Amherst and Northampton, Massachusetts. For the purposes of analyzing the data, children aged 4;3-5;11 (*M*=5;03) will be referred to as the younger group and children aged 6;1-7;7 (*M*=6;10) will be referred to as the older group.

The experimental technique that I used was based on Papafragou & Tantalou's (2004) experiment. The storyline was as follows. Tiger gave different animals a series of tasks. If an animal had performed a task successfully, Tiger rewarded it with a jewel; if not, Tiger gave it a card as a consolation prize. The child's task was to answer the question, "What will Tiger give animal X?" and to justify her answer. In the condition where the content of the potential implicature was not relevant to the QUD, an animal successfully performed Tiger's task. The target answer was, "Tiger will give animal X a jewel." In the condition where the content of the implicature was relevant to the QUD and the SI arose, an animal failed to complete its task. The target answer was, "Tiger will give animal X a card." In (13) below, an example of a story where an SI based on the Horn scale <all, some> is relevant to the QUD is provided.

(13) Tiger said, "I feel like drawing a picture but I can't find my crayons. I need all of my crayons because I want to draw a rainbow. Monkey, I want you to find all of my crayons for me." Monkey found some of the crayons.
Q: What will Tiger give Monkey? Why?
Target answer: a card.
QUD: Did Monkey find all of the crayons for Tiger?

(14) is an example of a story where a potential SI is not relevant to the QUD, hence does not arise.

(14) Tiger said, "I feel like drawing but I can't find any of my drawing stuff. I want to draw a car. Monkey, I want you to help me find something to draw with." Monkey found some of the crayons.
Q: What will Tiger give Monkey? Why?
Target answer: a jewel.
QUD: Did Monkey find something to draw with for Tiger?

3.4 Results

H_2: "Children start out by computing SIs that are relevant to the QUD but not the irrelevant SIs based on Horn scales" made by the Context-based QUD account of SIs was supported. Significant differences were found between the groups that got relevant vs. those that got irrelevant scenarios. First, consider the older children's results.

(15) Older children (6;1-7;7-year-olds). *M*=6;10. N=20.
some: relevant: *M*=2.40, *SD*=1.26; irrelevant: *M*=0, *SD*=0. $F(1,19)=36$, $p<.001$
warm: relevant: *M*=1.50, *SD*=1.17; irrelevant: *M*=0, *SD*=0. $F(1,19)=16.2$, $p<.001$
or: relevant: *M*=2.10, *SD*=1.44; irrelevant: *M*=.50, *SD*=.70. $F(1,19)=9.85$, $p<.006$
good: relevant: *M*=1.20, *SD*=1.54; irrelevant: *M*=0, *SD*=0. $F(1,19)=6.0$, $p<.025$

Significant differences in the older children's performance on relevant vs. irrelevant scenarios were found on all four Horn scales above.

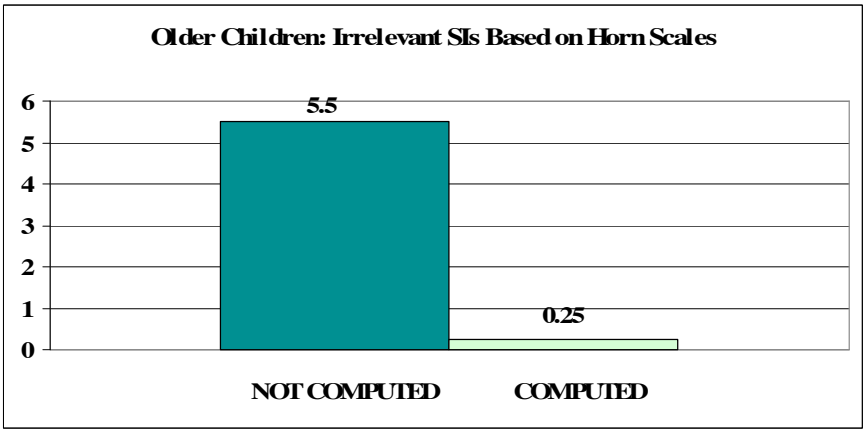

Figure One.

Younger children's results are provided in (16) below.

(16) Younger children (4;3-5;11-year-olds). (*M*=5;03). N=20.
some: relevant: *M*=2.50, *SD*=.97; irrelevant: *M*=.20, *SD*=.42 *F*(1,19)=47.13, p<0.000
warm: relevant: *M*=1.10, *SD*=1.10; irrelevant: *M*=.10 , *SD*=.31. *F*(1,19)=7.62, p<.013
or: relevant: *M*=.70, *SD*=1.25; irrelevant: *M*=.30, *SD*=.67. *F*(1,19)=.79, p<.385
good: relevant: *M*=.40, *SD*=.96; irrelevant: *M*=0, *SD*=0. *F*(1,19)=1.71, p<.207

In the case of the <*all, some*> and <*hot, warm*> scales, significant differences were found between groups that got relevant vs. irrelevant scenarios. In the case of the <*and, or*> and <*wonderful, good*> scales, no significant differences between groups that got relevant vs. irrelevant scenarios were found. While the younger children performed poorly on computing relevant SIs based on the <*and*, *or*> and <*wonderful*, *good*> scales, they did not compute irrelevant SIs based on these scales. Younger children's results are illustrated in Figure Two below.

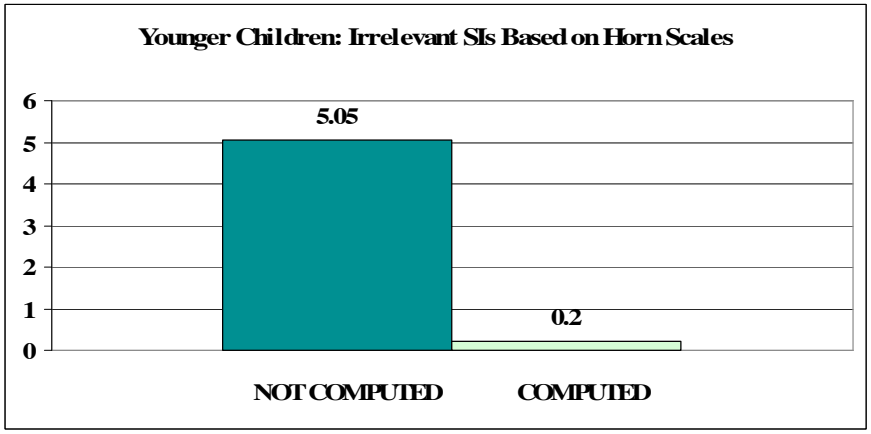

Figure Two.

Overall, 40 children of the younger and older groups did significantly better on pragmatic scales (*M*=3.87, *SD*=2.31) than Horn scales (*M*=2.15, *SD*=2.21); *F*(1,36)=25.834, p<.000. Three hypotheses concerning children's performance on Horn vs. pragmatic scales were postulated. H_3 made by the Default account and H_4 made by the Context-based QUD account on the assumption that lexical learning does not play a significant role were not supported. H_5**:** "Children do better on computing SIs based on the part/whole relation pragmatic scales than those based on Horn scales" made by the Context-based QUD account was supported. Children's results are provided in Figure Three.

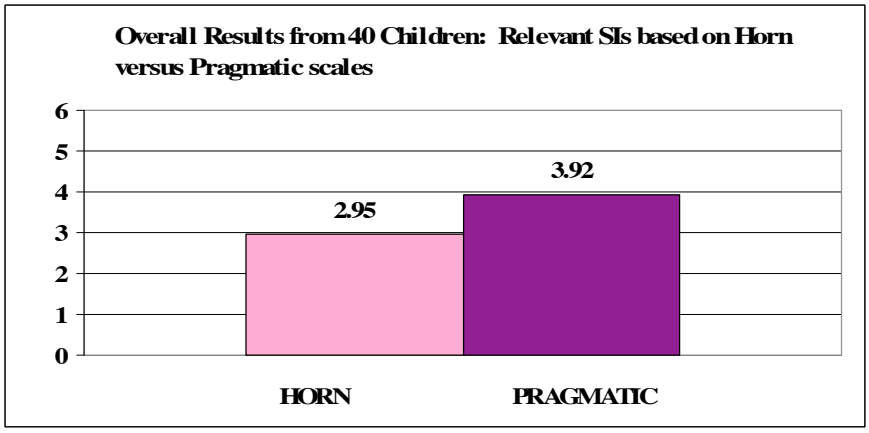

Figure Three.

3.5 Discussion

3.5.1 Role of Context

According to the Default account, the presence of a weaker scalar item triggers an SI automatically. Suppressing an SI in a context where it does not get projected requires additional pragmatic reasoning. As it was argued, the Default account gives rise to H_1: "Children start out by computing both the relevant and irrelevant SIs based on Horn scales." H_1 was not supported by the data. I found that children do not go through a stage where they compute SIs across the board. While some younger children did not compute any relevant SIs, they did not compute any irrelevant SIs either. While children's performance on computing SIs based on different scales varied, children computed only a negligible number of irrelevant SIs based on Horn scales. The group of the younger children computed irrelevant SIs in 4 instances out of 120, and the older children did so in 5 instances out of 120.

According to the Context-based QUD account, children are not predicted to compute irrelevant SIs because these cannot be arrived at by Gricean reasoning. Only the relevant SIs can be computed by Gricean reasoning. The Context-based QUD account predicts that an SI is computed iff it is relevant to the QUD. The Context-based QUD account of SIs predicted H_2: "children start out by computing SIs that are relevant to the QUD but not the irrelevant SIs based on Horn scales." H_2 was supported by the data. Consider S. P.'s (4;6) responses to the stories where SIs based on the <all, some> scale are relevant.

(17) S. P. (4;6): A card. Experimenter: Why? He didn't find all of them. [them=crayons].
S. P. (4;6): A card. Experimenter: Why? He didn't give him all of them. [them=ribbons].

Next, consider D. A.'s (6;5) responses to the stories where SIs based on the <all, some> scale are not relevant.

(18) D. A. (6;5): A jewel. Experimenter: Why? D. A.: Tiger asked him to find some drawing stuff. [=crayons].
D. A. (6;5): A jewel. Experimenter: Why? D. A.: Because Monkey gave him ribbons when he wanted some.

As the reader undoubtedly has noticed, children were not tested on pragmatic scales where the potential SI was not relevant to the QUD. This condition was not included in the experiment because no theory makes the prediction that children will go through a stage where they compute irrelevant SIs based on pragmatic scales, which are not part of the grammar and arise only as a result of having been made salient in the context.

One may raise the following objection to the claim that the presence vs. absence of a QUD that a potential SI is relevant to predetermines whether or not the SI gets projected. One may argue that in the stories in condition one where an SI based on a Horn scale is relevant to the QUD, one cannot tell if the SI arises as a result of the presence of a QUD or simply as a result of a stronger scalar item having been made salient, which, in turn, makes the pertinent Horn scale salient (13). I would like to argue that if the child were unable to compute the QUD that made the content of the SI relevant, he would not have been able to compute the SI by Gricean reasoning. The content of the Horn scale (ex., the quantifier scale in (13)) on its own does not tell the child that the content of the SI is relevant in the given context. Moreover, if a child is at a stage where she has constructed the quantifier scale, the presence of

a weaker scalar item such as *some* in (14) that is part of this scale could have, in principle, been construed as a signal that the quantifier scale is relevant. The data have shown this not to be the case. At the same time, in natural discourse, it need not be the case that whenever an SI is relevant to a QUD, the pertinent Horn scale has been made salient in the given context, as (19) illustrates. (See also (6) above).

(19) A's QUD: How many students came to John's class on Friday?
B: Some of them came to class.
SI: Not all students came to John's class on Friday.

In terms of children's performance on computing SIs, a possible follow-up experiment is to test children on contexts such as (19) vs. contexts where the scale has been explicitly mentioned in the prior discourse and the SI is relevant. While computing an SI in both types of contexts requires computing a QUD, it may or may not be the case that making a scale salient in the previous discourse aids the child in computing SIs.

3.5.2 Role of Type of Scale

One of the acquisition questions postulated in the beginning of the paper was,

(iii) Is the child's success on computing SIs predictable from the class of scale (Horn vs. pragmatic)?

On the one hand, it was found that children performed significantly better on computing SIs based on pragmatic part/whole scales than on computing those based on Horn scales. However, at the same time, taking a closer look at children's performance on the individual Horn scales makes it evident that the answer to (iii) is more complicated than it appears to be. Younger children provided the following percentages of target responses in the condition where the SI based on a given Horn scale was relevant: *or*: 23%; *good*: 13%; *some*: 83%; *warm*: 37%. Older children provided the following percentages of target responses in the condition in question: *or*: 70%; *good*: 40%; *some*: 80%; *warm*: 46;6%. Younger and older children's results are provided in Figures Four and Five, respectively.

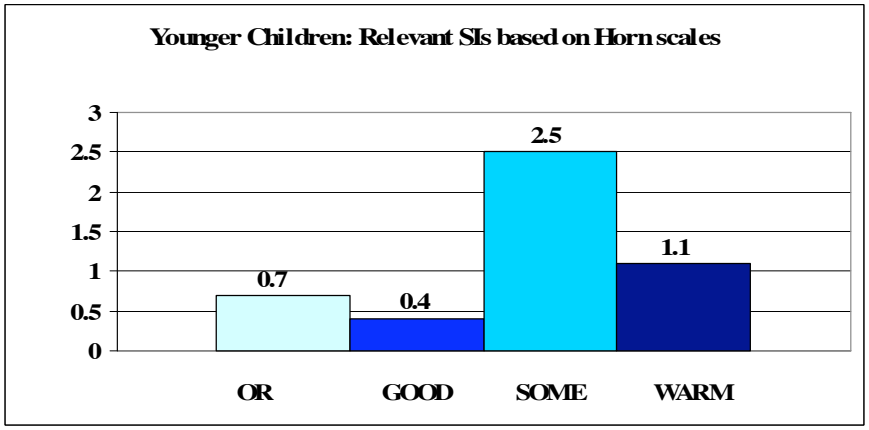

Figure Four.

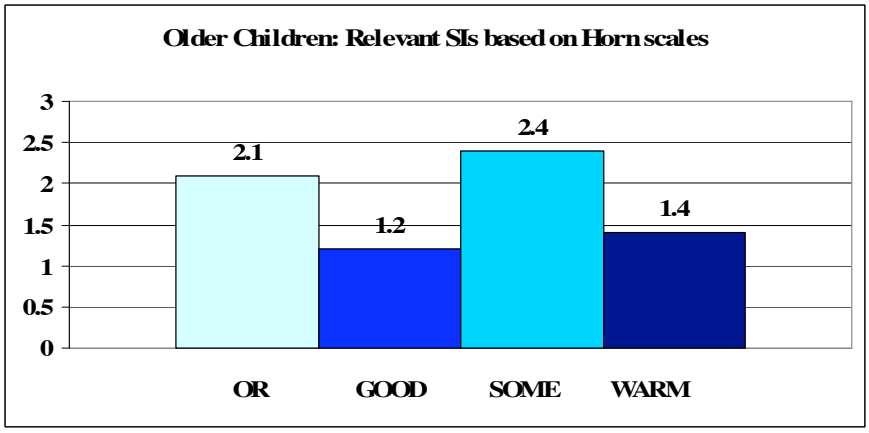

Figure Five.

The quantifier <all, some> scale is the only scale that the younger and older children have mastery of. By the groups of younger and older children combined, 81.6% of the relevant SIs based on the quantifier scale were computed. In the case of pragmatic scales, the groups of younger and older children combined computed only 65.4% of the relevant SIs. Given that the quantifier scale is a Horn scale, the answer to the acquisition question (iii) is a resounding "no" – the child's success on computing SIs is not predictable based on the class of scale.

The next question that naturally arises is why is the quantifier scale the least challenging one. As I have previously argued, entailment-based scales may be challenging in virtue of the fact that entailment is not the only condition on scalehood. I would like to hypothesize that children may at first conceive of the quantifier scale as being based on the part/whole rather than entailment relation. Note that the pragmatic scales that I employed in the present experiment were also based on the part/whole relation. The reason why children do better on the quantifier scale is that, in their input, children receive more exposure to SIs based on the quantifier scale than to those based on the part/whole relation pragmatic scales.

Another acquisition question that was postulated earlier was,

(i) Are children sensitive to the role of context in licensing SIs from the start or do they go through a "default" SI computation stage?

My experimental results have provided evidence to the effect that children keep track of what is and is not relevant for the purposes of the current conversational exchange or, on the Context-based QUD account, children keep track of QUDs, and are guided by the considerations of relevance in computing implicatures from the start.

Another acquisition question was,

(ii) How big a role does lexical learning play in determining the child's success on computing SIs based on a given scale?

My experimental results have shown that it plays a major role. The younger children demonstrated knowledge of one type of a Horn scale – the Quantifier scale – while performing worse than at chance on other Horn scales. The quantifier scale is lexicalized earlier than other entailment-based Horn scales.

Next, I will discuss children's performance on individual Horn scales in more detail. The connectives scale <and, or> has proven to be exceedingly challenging. This result is not surprising in the light of Grice's original conception of the implicature associated with *or* and Geurts' (2006) work on *or*. Grice's (1989) original argument was that *or* is used in contexts where one wishes to posit a non-truth-functional reason for accepting P v Q.

(20) a. The prize is either in the garden or in the attic.
b. Ignorance implicature: The speaker doesn't know for a fact that the prize is in the garden.
(Grice 1989: 44).

The context in (20) is an instance where affirming disjunct "P" is more informative than affirming a disjunction "P or

Q." In contrast, the use of *or* generates an SI in contexts where affirming a conjunction "P and Q" is more informative than affirming a disjunction "P or Q". The ignorance implicature in (20b) is not derived by making use of the <and, or> scale and, hence, is not an SI.

Geurts (2006) demonstrates that, contra the widely accepted view, the SI associated with *or* arises only in a very limited range of contexts. In a vast majority of cases, the SI is not derived and either an ignorance implicature is derived, as in (20), or a disjunctive reading of the sentence is computed but not via an SI. In order for a stronger SI of the form "speaker knows that not P" to be derived, the *competence assumption*, "the speaker is knowledgeable about the alethic status of the stronger statement," needs to be fulfilled (Geurts 2006: 2). Geurts demonstrates that, while in the case of *some*, the competence assumption is easily derivable, in the case of *or*, the competence assumption can be derived only in a very limited range of contexts, (21) being one example thereof.

(21) a. A to B: You may have an apple or a pear.
b. Stronger statement: You may have an apple and a pear.

In (21), the competence assumption is that A knows whether or not he will allow B to have an apple *and* a pear, whereby the use of *or* may give rise to an SI. (22) below is a case where an SI cannot be generated.

(22) a. A: John had an apple or a pear.
b. Stronger statement: John had an apple and a pear.

In (22), since speaker A does not know that disjunct P holds and does not know that disjunct Q holds, he cannot be assumed to know if the stronger statement "P and Q" holds, hence the competence assumption cannot be derived and an SI cannot be computed (Geurts 2006: 3).

All of my relevant SI condition scenarios based on the <and, or> scale were contexts where the competence assumption was satisfied, i.e., they were precisely the instances where the use of *or* generated an SI. In view of the above theoretical discussion, children's poor performance on the connectives scale is only to be expected. In the majority of uses of *or* that the child is exposed to, either an ignorance implicature is generated or a disjunctive reading of the sentence is generated but not via an SI. It is for this reason that the <and, or> scale takes a relatively long time to get lexicalized.

Next, I will discuss children's performance on gradable adjective scales. One could argue that children have non-linguistic knowledge of temperature differences or qualitative differences that are relevant for the <hot, warm> and <wonderful, good> scales, respectively. However, the following factors prevent the child from constructing gradable adjective scales early on. Interlocutors' use of gradable adjectives is highly context-dependent. A speaker may use the adjective *warm* to describe a +75F° weather in May and a +40F° weather in December in Massachusetts. In describing the May weather, the speaker may be referring to his own standards of warmth, while in describing the December weather, the speaker may be referring to the December standards of warmth. Moreover, different speakers may refer to a +80F° weather in July as either *warm* or *hot* depending on their personal perceptions. Being exposed to this kind of contradictory input, the child will experience difficulties in constructing gradable adjective scales.

An additional challenge that the gradable adjective scales pose for the child is the fact that the child may be unsure if the adjectives are lexicalized to the same degree. One of the conditions on Horn scales is that lexical items that constitute a Horn scale have to be equally lexicalized. As I have mentioned, I employed the <wonderful, good> scale in my experiment. It is likely that the term *wonderful* is much less frequent in the child's input than the term *good*. Thus the child may wonder if *wonderful* is as lexicalized as *good*. According to Levinson, "the lexicalization constraint requires that stronger items on a scale be lexicalized to an equal or greater degree than weaker items" (Levinson 2000: 80). *Wonderful* is, in fact, the stronger item on the relevant Horn scale. The lexicalization constraint specifically concerns the stronger scalar items because obeying the maxim of Manner cannot be the speaker's reason for eschewing a stronger scalar term. If a potentially stronger scalar item is being avoided for reasons of prolixity or obscurity, the speaker's use of a weaker scalar item will fail to generate an SI to the effect that the stronger one does not hold. In a nutshell, because the child may (justly) classify the stronger items on the "goodness" scale as less lexicalized than the weaker ones, this may prevent him from constructing this scale early on. Note that the lexicalization constraint is not an absolute rule – the "goodness" scale is operative in adult language despite the fact that the stronger items are less lexicalized.

Another stumbling block that children are likely to run into in constructing gradable adjective Horn scales is their perception of the degree of contrast between adjacent scalar items. The knowledge concerning the degrees of contrast between specific scalar items is language-specific information that is acquired from the input. Suppose that

for English-speaking adults, *fantastic* is stronger than *great* on the relevant "goodness" scale. However, the two items are close, if not adjacent, on the "goodness" scale, and *great* is fairly high on the scale. Thus it is likely that an adult would compute an SI from the speaker's use of *great* to "not fantastic / marvelous / fabulous etc." only in contexts where it has been made explicit that a greater degree of goodness than that designated by *great* is relevant to the QUD. That is, if a QUD is of the form, "Isn't x fantastic / marvelous / fabulous?" the SI "not fantastic" from the use of *great* may be computed; however, if the QUD is "How did you like x?" this SI will not be computed from the use of *great*. As a result, it is likely that the use of *great* would generate an SI only in contexts where a QUD contains a stronger scalar item.

It is easy to imagine that, given his limited experience with language use, the child is likely to be unsure concerning the degree of contrast between the given adjacent scalar items. There may be a stage where the child has constructed the "goodness" and "temperature" scales but perceives the contrasts between *good* vs. *wonderful* and *warm* vs. *hot* as being as subtle as the contrast between *great* and *fantastic* is for adults. Thus some children may experience difficulty in computing SIs based on gradable adjective scales not because they have not constructed the scales but because they are yet to infer the degree of contrast between the specific scalar items that exists in the adult language.

Moreover, speaker variation plays a role here as well. There are speakers who commonly use terms like *great*, *fantastic* and *fabulous*, and there are those who never use these terms but prefer to express the highest praise through using litotes like "not bad." A speaker who eschews the terms that are ranked high on the "goodness" scale may never intend an SI by using the term *good* because, in his I-language, *good* may be ranked as strongest. Such I-language uses related specifically to gradable adjective scales are likely to be part of the child's input. This type of input may delay the child's construction of the relevant scale because the child is exposed to uses of gradable adjectives that are not the strongest on the relevant Horn scales in contexts where the stronger scalar item is relevant. In sum, contextual factors, the issue of lexicalization, the degree of contrast between the given adjectives and I-language uses of gradable adjectives are the factors that make constructing gradable adjective scales challenging. Note that none of the complications posed by the connectives scale and gradable adjective scales are relevant to the quantifier scale, which was reflected in children's performance on the four Horn scales employed in the experiment.

4. Concluding Remarks

In the present experiment, I have provided evidence for the Context-based QUD account of SIs and evidence against the Default account of SIs. My experimental results have shown that from the start children compute SIs only in contexts where the content of the SI is relevant to the QUD. I have also found that the child's success on computing SIs is <u>not</u> predictable from the ordering relation that the scale is based on (entailment-based Horn scales vs. non-entailment-based pragmatic scales). The next question that arises is do Horn and pragmatic scales have different statuses in terms of the child's UG-based innate knowledge. As far as the UG-status of Horn scales, three options are tenable. The first option is that most or even all Horn scales are part of the child's innate knowledge, and a variety of factors related to language use that were discussed here make certain Horn scales more challenging than others. The second option is that Horn scales themselves are not part of UG, whereas conditions on scalehood (ex., unilateral entailment, adherence to the same semantic field) are part of UG. The third option is that neither Horn scales nor conditions on scalehood are part of the child's UG-based innate knowledge. I favor the second option over the first, in part, because experimental evidence has shown that children do not form Horn scales as soon as they have classified given lexical items as belonging to a category that participates in a Horn scale. In addition, many Horn scales are highly dependent on world knowledge, such as <ancient, old>, and are unlikely to be part of the child's innate knowledge. To date, all of the experimental results, including my own, have been compatible both with options two and three. As far as pragmatic scales are concerned, while these are not part of UG, children do have non-linguistic knowledge of some relations that pragmatic scales are based on, such as the part/whole relation and, by the age of 4, the level of description (Rosch et al. 1976). Thus it is plausible that children have UG-based linguistic knowledge of conditions on Horn scales and non-linguistic but also innate knowledge of ordering conditions that some pragmatic scales are based on. If option two is correct, it follows that, while there is a distinction between Horn and pragmatic scales, there is a continuum of Horn scales, with some being more grammaticalized and others less so, which makes them closer to their pragmatic cousins. If option three is correct, in terms of their UG-based innate knowledge status, there is no distinction between Horn and pragmatic scales. Adjudicating between options two and three is subject to future research.

5. Appendix

Test sentences: <all, most, many, some>: (1-3); <and, or>: (4-6); <wonderful, good>:(7-9); <hot, warm>: (10-12); pragmatic: (13-18).

(1) Monkey gave Tiger some of the ribbons.
(2) Deer found some of the knives.
(3) Monkey found some of the crayons.
(4) Monkey said, "I can teach you how to peel bananas or how to whistle."
(5) Deer said, "I can bring you some blueberries or strawberries from the forest."
(6) Monkey said, "I can lend you my motorcycle or my car."
(7) Deer made Tiger a good boat.
(8) Deer made a good wardrobe for Tiger.
(9) Monkey gave Tiger a good backrub.
(10) The soup was warm.
(11) Deer made warm apple cider for Tiger.
(12) Monkey started a fire and it got warm in the house.
(13) Deer cleaned the left corner of the kitchen table.
(14) Deer cut the nails on Tiger's pinky fingers.
(15) Monkey cleaned the kitchen.
(16) Deer scratched Tiger's left paw.
(17) Deer cleaned the right corner of the window.
(18) Monkey washed Tiger's right paw with a sponge.

References

Chierchia, G. (2004). Scalar Implicatures, Polarity Phenomena, and the Syntax/Pragmatics Interface. In A. Belletti (ed.), *Structures and Beyond*. Oxford University Press, Oxford.

Chierchia, G., Crain, S., Guasti, M. T., Gualmini, A. & Meroni, L. (2001). The Acquisition of Disjunction: Evidence for a Grammatical View of Scalar Implicatures, *BUCLD 25 Proceedings*, 157–168, Cascadilla, Somerville.

Chierchia, G., Guasti, M. T., Gualmini, A., Meroni, L. Crain & S., Foppolo F. (2004). Semantic and Pragmatic Competence in Children's and Adults' Comprehension of "or." In *Experimental Pragmatics*, eds. I. Noveck & D. Sperber, Palgrave Macmillan, New York.

Foppolo, F., Guasti, M. T. & Chierchia, G. (2004). Pragmatic Inferences in Children's Comprehension of Scalar Items. A paper presented at *Second Lisbon Meeting on Language Acquisition*.

Foppolo, F., Guasti, M. T., Chierchia, G. (in press). Scalar Implicatures in Child Language: Failure, Strategies and Lexical Factors.

Geurts, B. (2006). Exclusive Disjunction without Implicature. *Semantics Archive*.

Guasti, M. T., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., Meroni, L. (2005). Why Children Sometimes (but not Always) Compute Implicatures. *Language and Cognitive Processes*. 20 (5): 667-696.

Hirschberg, J. (1991) *A Theory of Scalar Implicature,* Garland, New York.

Horn, L. (1989). *A Natural History of Negation*. University of Chicago Press, Chicago.

Horn, L. (1984). Toward a New Taxonomy for Pragmatic Inference: Q-based and R-based Implicature. In D. Schiffrin (ed.), *Meaning, Form and Use in Context: Linguistic Applications* (GURT '84): 11-42. Washington, DC Georgetown University Press.

Horn, L. (2005). The Border Wars: a Neo-Gricean Perspective. In K. Turner & K. von Heusinger (eds.), *Where Semantics Meets Pragmatics*. Elsevier.

Levinson, S. (2000). *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Cambridge, MA: MIT Press.

Noveck, I. (2001) When children are more logical than adults: Experimental Investigations of Scalar Implicature', *Cognition* 78: 165–188.

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

Roberts, C. (1996). Information Structure: Towards an Integrated Theory of Formal Pragmatics. *OSU Working Papers in Linguistics*, vol. 49: *Papers in Semantics*: 91-136.