Measuring Scalar Implicature: Developing a Paradigm

Hannah Gogel

Tufts University

The development of the paradigm and data analysis were done with Noemi Hahn, Yi Ting Huang, and Jesse Snedeker. The running of participants was done with Noemi Hahn, Justin Morse and QingQing Wu.Abstract

Scalar implicatures are the inferences made by listeners when a speaker uses the weaker of two terms on a scale. This paper investigates the processing of the word *some* and under what contexts six to nine year olds interpret the term to be and not be consistent with the term *all*. Previous online processing tasks and offline judgment task research suggests that listeners start to understand *some* to mean only a subset and not a whole set sometime shortly after age five (Huang & Snedeker, 2009b; Huang, Spelke, & Snedeker, n.d.; Noveck, 2001; Papafragou & Musolino, 2003; Pouscoulous, Noveck, Politzer & Bastide, 2007). By creating new experiments based off of two previously run paradigms (Huang & Snedeker, 2009; Huang, Spelke, & Snedeker, n.d.) we compared online and offline data of six to nine year olds' interpretations of the word *some* to find that this age range of children does not make implicatures during processing, but seem to make them during offline tasks by age eight and possibly earlier. These results suggest that six to nine year olds either need specific contexts or extended time to make such implicatures and give us insight into what contexts may be useful for priming or teaching such pragmatic interpretations.

MEASURING SCALAR IMPLICATURE: DEVELOPING A PARADIGM

1. Introduction

1.1 Pragmatics

Pragmatics, an aspect of language, includes specifically the understanding and producing of social and situational cues. Within the category of pragmatics lies all information in a sentence outside of semantics and syntax including unsaid implications such as those in metaphor, irony, sarcasm, and hyperbole. The most significant attribute of pragmatics is that it requires the participants in a conversation to understand words to mean more than their definitions or semantic meanings. Snedeker (2009) highlights the importance of this characteristic in her definition of pragmatic skills as those that "allow us to go beyond the literal meaning of an utterance and figure out what it implies in a particular conversation or how we can use it to make our point" (p. 4).

For example even if a speaker says, "This dinner is going to be a train wreck," the listener should not be expecting there to be trains. There is nothing in the spoken words that directly states that the speaker is not being literal; the listener is assumed to be able to infer from context and previous experience that the term "train wreck" is being used as a metaphor.

Musolino and Lidz (2006) wrote, "Because quantificational phenomena involve the interplay of multiple levels of linguistic analysis (i.e., syntactic, semantic, and pragmatic), the grammar of quantification represents a paradigm particularly well-suited for investigating the development of a broad range of linguistic principles" (p. 818). In this paper we will be specifically looking into the case of scalar implicature, specifically the quantifier *some*, in order to gain insight into the complexities of pragmatic processing.

1.2 Scalar Implicature

A straightforward example of interpreting words to mean more than their technical definitions is the case of scalar implicature. A scalar implicature communicates the details of a situation with the listener by using the less informative of two words on a scale. For instance if a speaker asks “Is the tea too hot to drink?” and the listener responds “It’s warm”, the original speaker will drink the tea under the assumption that it is not *hot*; the tea is only *warm*. This assumption is an example of a scalar implicature: if a speaker uses the word *warm*, most adults would assume that the speaker did not mean *hot*. Another example of this type of implicature is *start* as opposed to *finish*. If a speaker states that he has *started* a movie, a listener assumes that the speaker has not yet *finished* it.

A possible explanation for the processing of making scalar implicatures is Grice’s maxim of quantity (Grice, 1975; Huang & Snedeker, 2009b). According to this maxim, the speaker will always state as much information as necessary, no more, no less. This means that a speaker leaves out all obvious information and relies on the listener to understand the obvious information to be implied. This maxim is relevant to the case of scalar implicature because if a speaker means the stronger term, in our examples *hot* or *finish*, he would have said it. Even though the speaker does not technically say that he doesn’t mean the stronger term, the listener here should understand that the speaker is giving all of the critical information and no less.

In this paper, we investigate the term *some* as opposed to the stronger term *all*. The term *some* has two interpretations: the first is the logical or semantic definition and the second is the inferred meaning, which requires pragmatic processing. The logical definition is broader, including the interpretation of *some* as a subset and as a whole set. The inferred meaning is more specific with *some* meaning strictly a proper subset and not a whole set (in this definition, *some*

cannot mean *all*). This second definition requires that there is a subset of objects in the specified set that are not included by the term *some*. *Some* of a set does not necessarily mean *all* of it, but the two terms can be consistent with each other as given by the first definition of the term. This is particularly clear in a conditional statement such as "If you finish some of your vegetables, you can have dessert." It is most likely also true that if the listener finishes *all* his vegetables, he can have dessert.

Some and *all* are not intuitively consistent because most listeners make the scalar implicature that the speaker would avoid using the word *some* and say *all* if he meant *all*; that is, most listeners assume the second definition of *some* to hold. For instance, if a speaker says, "She has some of the apples", she could still have *all* of them but most likely a listener would infer that there is a subset of the apples that she does not have.

1.3 Background: Offline Data

Various offline research tasks have been done with children to determine their ability to judge the consistency between statements including scalar implicatures and given situations. This includes Noveck's study (2001) where he presented children and adults with underinformative statements. Noveck found that children younger than eleven were not rejecting statements such as "Some giraffes have necks", which is logically true but pragmatically false, when adults were. Additionally, Papafragou and Musolino (2003) conducted an experiment in which children had to judge or correct a statement including the word *some* to describe a story. This study found that five year olds do not correct the use of *some* when it refers to a whole set. Pouscoulous, Noveck, Politzer, and Bastide (2007) administered a task in which subjects were asked to either accept or change an arrangement of tokens so that *some* boxes had tokens in

them. They found that children begin to reject the broader interpretation of *some* (only accepting *some* not to be including *all*) at around age seven.

Huang, Spelke, and Snedeker (n.d.) found matching offline results. The difference in their experiment is that each child was offered a chance to choose between the narrower and broader definition of the term *some* in addition to having the opportunity to reject the broader definition. The interpretation of the term was presented as a choice with several options rather than a chosen interpretation and the option to correct it. Within the trials of this study, children were presented with three boxes; one of them was always covered. The child was always instructed to point to a character with *some* of the object. The uncovered boxes either displayed the referred to character with (1) nothing in one, and a subset of the objects in another, with (2) a subset in one and the whole set in another, or with (3) a whole set in one and nothing in the other. The results from this task lead the researchers to conclude that three year olds do not reject the interpretation of *some* to mean *all* and do not favor either interpretation. When adults were run on this paradigm, they were found not to reject *all* as an interpretation for *some* when it was the only visible option (apart from interpreting *some* to mean *none*), but adults preferred the subset definition of the word to the whole set interpretation when both were shown.

During the condition in which the *none* and whole set options were visible, both adults and children chose the whole set. This was a situation in which the scalar implicature was cancelled for adults (Huang, Spelke, & Snedeker, n.d.).

Each of these experiments was a judgment task, asking children and/or adults to have an opinion on an already established situation, to interpret rather than produce. As Huang, Snedeker, and Spelke (n.d.) showed, there are some instances in which scalar implicatures can be cancelled.

In their study, when the only visible options for the interpretation of *some* is a whole set and *none*, the broader definition of *some* was accepted by adults. This is a possible explanation for the discrepancies between findings, as each one presented a different context for the interpretation of *some*. Though results are not consistent as to the age at which children start making scalar implicatures, all agree that it is shortly after the age of five.

1.4 Background: Online Data

Online eye tracking research has shown that there is a delay in processing implicatures in comparison to the processing of other quantifiers (ex. *all*, *none*, *two*, *three*) but that adults do infer that *some* does not mean *all* in most contexts (Grodner, Klein, Carbary, & Tanenhaus, 2010; Huang and Snedeker, 2009b). The delay is thought to come from making the inference, which is unnecessary for other quantifiers, as a result of choosing between the logical (including whole set) definition and the inferred definition (only subset) (Huang & Snedeker, 2009a; Huang & Snedeker, 2009b).

Huang and Snedeker (2009b) ran the same paradigm on both adults and five year old participants. We can see from this comparison that while children at age five process *some* to mean both a subset and *all*, adults interpret the term to mean only a subset.

1.5 Current: Online and Offline Processing of Scalar Implicatures, Ages 6-9

Adult listeners understand unspoken information all the time. Without being able to do so it would be impossible to carry a conversation with any sort of speed. The question of scalar implicature processing is a way of understanding how and when people make inferences during day-to-day conversation. By investigating the understanding of scalar implicatures in children in the years just after age five, we can look at how and when people are developing the ability to

make such pragmatic inferences. Musolino & Lidz (2006) deduced along with Crain and Thornton (1998) that "the study of grammatical development cannot proceed without a careful consideration of children's growing language processing and pragmatic abilities" (p. 3).

In this paper we replicated the study done by Huang, Spelke, & Snedeker (n.d.) with children from ages six to nine, as opposed to the previously tested three year olds with the expectation, drawn from previously discussed offline data, that children start making implicatures sometime shortly after age five.

This study also uses the visual world paradigm (Trueswell, Sekerina, Hill, & Logrip, 1999) and eye tracking data to examine the processing of the scalar term *some* by six to nine year old children, as opposed to the five year olds tested by Huang and Snedeker (2009b).

This online paradigm is based on the one used in Huang and Snedeker's (2009b) work. In addition to changing the age range of the participants, we also changed the phrasing of the sentences used in the task. This was in order to increase the length of the time period in which the participants can make the scalar implicature, or in other words, the time period between the onset of the quantifier and the phonological disambiguation between objects. We did this by referring to all objects in the study by phonologically overlapping compound nouns instead of just phonologically overlapping nouns. For example, instead of using a girl with "some of the matches" or "some of the maps" (Huang & Snedeker, 2009b), we used a girl with "some of the ice cream cones" or "some of the ice cream sandwiches" (The reasoning for this is further discussed in Experiment One Methods.).

By using online data we are able to examine the quantifier processing as it occurs as an implicit process, as it would in most day-to-day situations, rather than only having the subjects

reflect on statements or situations after the fact. The additional offline data allows us to see what happens when the subject is given the opportunity to make a judgment on the word *some* apart from natural processing. Offline measures also allow us to see in which cases the scalar implicature can be cancelled as well as show that the children are actually making the implicature. In the eye tracking data paradigm used in our experiment and in Huang and Snedeker (2009b), the participant is not required to make the choice between the subset and whole set definition because he is always eventually told which one to pick. By having offline data we can see the choice made by the participants when they are not given an answer.

We are also able to compare the online and offline data from the participants. Whether we find similar or conflicting results gives us insight regarding the different contexts in which implicatures are made and a timeline for when the listener is making the implicature, during or after listening, if he or she is.

2. Online Data

Online data was collected using a Tobii eye tracker that looks like a computer monitor and uses cameras to monitor where the participants' pupils are looking on the screen. A touch screen was attached to the eye tracker so that the children could respond by touching the screen while keeping their eyes on the monitor.

2.1 Experiment One

Experiment One collected online data on six to nine year olds' processing of the quantifiers *all*, *some*, *three*, and *two* with an edited version of the Huang and Snedeker (2009b) paradigm.

2.10 Participants

Twenty-one children between ages six and nine with a mean age of seven years and four months were subjects in this study (SD = 13.42 months). Each participant was given a toy for his or her time and the family of each participant was given five dollars as compensation for transportation costs. Recruitment for participants was done through fliers and mailings in the Boston area. In addition, researchers actively recruited at several Boston locations and events.

2.11 Methods

This experiment had two practice trials followed by sixteen experimental trials. The two practice trials did not include quantifiers to ensure the participants understood the task. During each trial the participant heard a sentence and was presented with a display such as the one below (Figure 1). The figures were presented on the screen of a Tobii eye tracker with an attached touch screen.

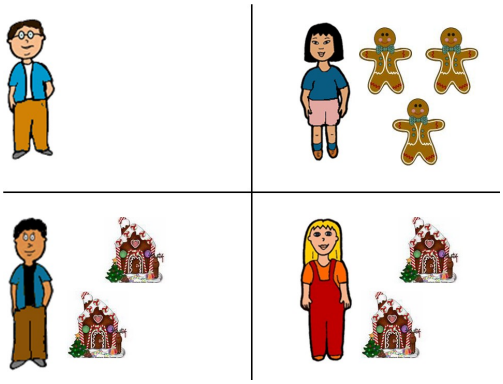


Figure 1 Display of an experimental trial for Experiment One. For an *all* trial, the target is top right and the distractor is the bottom right. For a *some* trial, the numbers of objects are flipped across the horizontal axis, leaving the top two characters with two gingerbread men each and the bottom left character with three gingerbread houses.

Participants were introduced to the four characters shown in Figure 1 during the practice period of the experiment as well as at the beginning of each experimental trial. The characters

were the same in every trial. At the start of each trial, an audio clip also identified the objects on the screen and to whom they belonged by telling a short story about the characters and the objects. Each girl-boy pair had the same type of object in the story and on the screen (unless a character had no objects, see Figure 1 top left).

After learning about the characters and the objects of the experimental trial, the participant would see an image such as Figure 1 and hear one of the following options (specific to Figure 1):

(1) “There is a girl who has all of the gingerbread men”
(2) “There is a girl who has some of the gingerbread men”

In an experimental trial looking for a character with *two* or *three* of an object the participant would hear one of the following options (specific to Figure 2):

(3) “There is a girl who has three of the gingerbread men”
(4) “There is a girl who has two of the gingerbread men”

Every display had a target and distractor (Figure 1, Figure 2). The target was the character described by the sentence, and the distractor was the character of the same gender as the target.

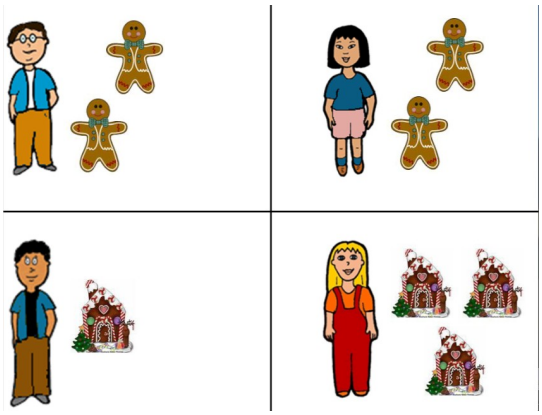


Figure 2 Display of an experimental trial for experiment One for a *two* trial. The target is top right; the distractor is bottom right. For a *three* trial, the number of objects would be flipped across the horizontal axis leaving the bottom two characters with two gingerbread houses each,

the top right character with three gingerbread men, and the top left character with one gingerbread man.

While the sentence played, the Tobii collected eye tracking data. At the end of each sentence, the participant was expected to touch the character on the screen that the sentence was describing, also known as the target character. Action responses were also recorded.

Compound nouns could be used to describe all of the objects in every trial. Pairs of compound nouns were used in order to extend the length of time allowed for the scalar implicature to be made before the options were phonologically disambiguated. The pairs of compound nouns in each trial shared the same first word (example from Figure 1 and Figure 2: *gingerbread* house and *gingerbread* man).

The use of compound nouns can be seen as significant in sentence 2 listed above where the participant has *of the gingerbread* to make or not make the implicature, where he would otherwise only have *of the* to make the implicature if the items were completely phonologically unique. The point of phonological disambiguation is pushed back from *of* to *–men* by using compound nouns.

The crucial part of the sentence is between the onset of the quantifier (in the case of sentence 2, *some*) and the point of phonological disambiguation (in this case, *-men*). If the participants make the implicature, their eyes should be looking towards the target character before the point of phonological disambiguation. If they are not making the implicature and accepting *some* to mean *all*, they should either be looking at the target or the distractor because the gender cue at the beginning of the sentence should guide them to eliminate the other two options.

2.12 Results

Action responses, the selection of a character after the whole sentence was heard, were close to 100% correct (> 99%), participants basically always selecting the target. Since the target is always unambiguously stated, this result shows only that the participants are paying attention to the task. We analyzed the proportion of looks to the target compared to the sum of all looks to target and distractor while participants listened to the sentence (Figure 3).

Figure 3 Looks to target divided by the sum of looks to target and looks to distractor during sentence processing in Experiment One.

In Figure 3 we can see that the participants favored looking at the character with three of an object or the character with all of an object, one of these characters, depending on which trial, is always the character with the most objects on the screen. Previous research has shown that children prefer to look at groups containing more objects (Kanan, Tong, Zhang, & Cottrell, 2009), which explains the looking preference shown before the critical sentence has even begun.

The analysis completed was a 2x2 ANOVA (strengths: weak vs. strong, quantifier: scalar vs. numeral). The only significant result from our looking patterns was in strength (*two* and *some* are weaker, *three* and *all* are stronger - $F = 30.389$; $p < .001$) which can be explained, as previously stated, by the children's preference to look at more objects. This brings us to what is more noticeable about this data: the looks to target in the *two* condition barely get to 60% by the point of phonological disambiguation. Since by age six, children should be able to rapidly interpret the quantifier *two*, we investigated this further.

In the five year old children's data collected by Huang and Snedeker (2009b) using a similar paradigm in the region equivalent to the highlighted region in Figure 4, from the onset of the gender cue to the onset of the quantifier, looks to target and distracter showed a significant increase. In our data, no significant increase is shown.

Figure 4 Looks to specific characters during sentences using the quantifier *two*. The region between the onset of the gender cue and the onset of the quantifier is highlighted.

From farther analysis shown in Figure 5, we can see that in every condition, many participants were not looking towards the correct gender. Even by the time the quantifier was heard, a large proportion, about 50%, of the participants had not reacted to the gender cue.

Figure 5 Looks to the target gender minus looks to the opposite gender in between gender cue and the hearing of the quantifier in all four conditions.

2.13 Discussion

If the children were making the scalar implicature, we should see a great increase in the looks to target in the *some* trials at some point after the quantifier and before the point of phonological disambiguation. We do not see this; the looks to target just reach 50% by the point of disambiguation (Figure 3).

It seems from Figure 3, that the children in this study were having difficulties with gender cues. After hearing the gender of the character, there are only two choices left for which character is being described. The participant should be able to eliminate the two characters of opposite gender (adjacent and diagonal from the target) from his options. As we can see from Figure 5, not only in the *two* trials are children not reacting to the gender cue, but in all other

conditions as well. So why is this lack of gender cue reaction affecting the *two* trials so dramatically (Figure 3)?

The stimulus presented in the *two* and *three* conditions is different than the ones presented for the *some* and *all* conditions. In Figure 6, we can see that if a participant were not immediately processing the gender cue, he or she would not be able to determine who was being described by a sentence about a character with two baseball hats (both options are circled in blue in Figure 6). This situation where the characters shown are only differentiated by gender only (not item or number of items) occurs only in the *some* and *two* conditions. In the *some* condition, there should be delay without gender cue confusion due to making or not making the implicature, leaving the only evidence of gender cue confusion to be during the *two* trials.

Figure 6 A comparison between the screen displays for the *two*/*three* conditions (left) and the *some*/*all* conditions (right). The characters that only differ in gender in the *two* trials are circled in blue.

It is possible that the delay in the *some* trials is due to a processing delay specific to *some* or to the same lack of gender cue processing as the *two* trials. Since the processing of the *two* quantifier is delayed similarly to the processing of the *some* quantifier, it is impossible for there to be a significant interaction between scale (numbers vs. *some*/*all*) and strength measures.

Possible difficulties processing the gender of characters may have resulted from the phrasing used. The possibility that this issue was brought about by ambiguous looking characters is small, since the characters are named and differentiated at the beginning of the task. Also, in Huang and Snedeker (2009b), there were no problems with gender interpretation with the same characters and younger participants. Huang and Snedeker (2009b) used the following phrasing

instead of the sentences used in Experiment One (phrasing here is in reference to what would be heard while an image similar to Figure 1 was displayed):

(1) “Point to the girl who has all of the gingerbread men”
(2) “Point to the girl who has some of the gingerbread men”

And the following sentences would be heard with an image similar to Figure 2:

(1) “Point to the girl who has three of the gingerbread men”
(2) “Point to the girl who has two of the gingerbread men”

The phrasing differences here are between the start “Point to the-” and “There is a-”. The original change from the 2009(b) paper to Experiment One was made in order to make sure the task was a test of scalar implicature, with the worry that “Point to the-” was too narrow of a command, implying that there is only one interpretation of *some*. As there seemed to be confusion with the current direction, we proposed that it might have been due to the new, more vague, phrasing “There is a-”.

“There is a-” brings up two important issues. First, it is existential phrasing using an indefinite article; a sentence starting this way does not imply that the character that follows is on the screen or relevant to the task at hand the way that the older phrasing does. It also does not contain a literal instruction; it is a declarative statement. The participant is just expected to learn that the statement starting with “There is a-” is describing a character that the participant should touch on the screen.

In an effort to eliminate the gender processing confusion, we created a second experiment. With the gender cue not being processed, the participants had no time or situation in which they were able to make or not make a scalar implicature. If the participants are not interpreting the gender cue, it is impossible to tell when they are interpreting the quantifiers. The

Huang and Snedeker (2009b) participants did not have a trouble with the gender cue; therefore, in Experiment Two we changed the phrasing back to that of the Huang and Snedeker (2009b) study.

2.2 Experiment Two

Experiment Two collected online data on six to nine year olds processing the quantifiers *all*, *some*, *three*, and *two*. The paradigm was slightly modified from Experiment One to correct for participants' gender cue confusion.

2.20 Participants

Twenty-two children between the ages of six and nine, with a mean age of seven years and eleven months (SD = 13.26 months), participated in this experiment. Participants were recruited and compensated in a similar fashion to Experiment One.

2.21 Methods

The procedure for Experiment Two was identical to the procedure for Experiment One except for the wording of the commands to the participant. The sentences in Experiment Two started with the words "Point to the-" rather than "There is a-" in order to ensure the processing of gender cues (see discussion for Experiment One for further reasoning and examples of new sentences).

2.22 Results

Action responses were more than 93% correct showing that the participants were paying attention to the task. There seemed to be a trend towards a time delay in the interpreting *some* compared to *two*, *three*, and *all*, but it was not significant.

However, we did notice that the compound nouns were split into two different groups. 25% of the compound noun pairs used were superordinate nouns. For example, the pair *gingerbread house* and *gingerbread man*, showed an instance where all of the pictures were of gingerbread whether it be in the form of a house or a man. 75% of the noun pairs did not have this issue. For instance pictures of *buttercups* and *butterflies* do not include an image of anything that could be considered butter. Since the participant could process the sentence as a whole after the phrase "Point to the girl who has some of the gingerbread –" we looked into this further.

The graph displayed in Figure 7, shows the target-distractor proportions in sentences that did not include superordinate compound nouns. Here the interaction (determined by a 2x2 ANOVA) between the scale and strength of quantifiers is not significant ($p > .05$, $F = 1.40$). The data in Figure 8, of the superordinate noun trials, shows drastically different looking patterns than in Figure 7. Though the noisiness in Figure 8 is partially due to the limited amount of data analyzed (only 25% of the data, compared to Figure 7 which has 75%), Figure 7 displays looking patterns that trend together in all trials where Figure 8 has looking patterns that seem not only

different, but also random with regard to the processing of the implicature.

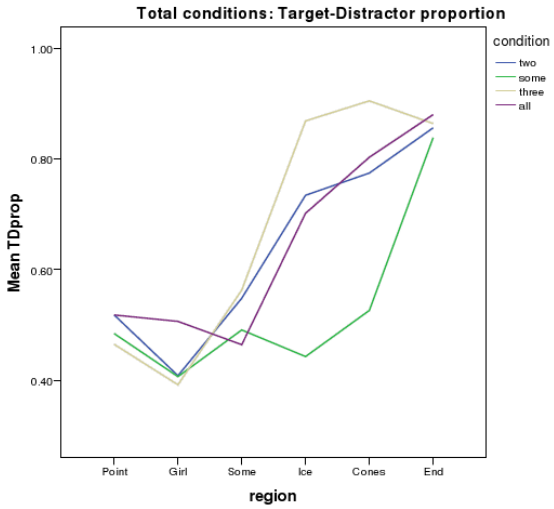


Figure 7 Looks to target divided by the sum of looks to target and looks to distractor during sentence processing for non-superordinate trials in each condition.

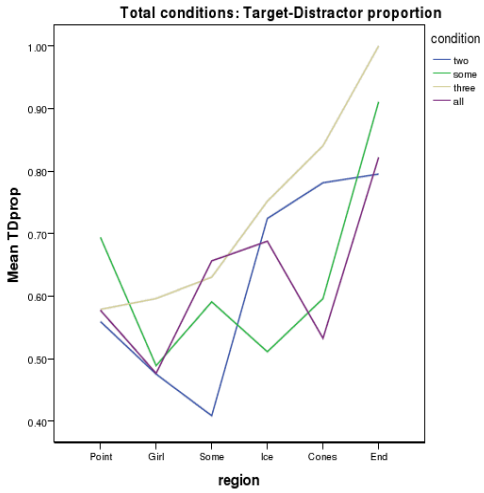


Figure 8 Looks to target divided by the sum of looks to target and looks to distractor during sentence processing for superordinate trials in each condition.

2.23 Discussion

It is clear that a quarter of the data was noticeably different than the rest of the data collected. The sentence referred to by Figure 7 and Figure 8 is "Point to the girl who has some of the ice cream cones" which asks for the character with ice cream cones as opposed to the distractor character who is holding ice cream sandwiches. We hypothesize that while listening to this sentence, the participant could start processing "Point to the girl with some of the ice cream-" as an entirety. These superordinate noun trials don't allow the interpretation of *some* as a whole set until the end of the sentence. Even if a subject is making an implicature, the girl with *all* of the ice cream cones only has a subset of the ice cream. Since all characters, in this example, have a subset of the ice cream, whether it be a cone or a sandwich, there is no possibility to interpret *some* to mean the whole set until the entire sentence is heard. If there is no opportunity to interpret *some* as meaning *all* there is no way to tell from the eye tracking data whether the participant is making the implicature. Also, since all characters have *some* of the ice

cream, the participant doesn't know which character to look at, which is a probable explanation for the different eye tracking patterns in the superordinate noun trials.

Though Figure 7 does show a delay until after disambiguation in the processing of the term *some*, we created a third experiment to eliminate the confusion brought about by the superordinate noun trials.

2.3 Experiment Three

Experiment Three collected online data on six to nine year olds processing the quantifiers *all*, *some*, *three*, and *two*. The paradigm was slightly modified from Experiment Two to correct for the differences in trials with superordinate and non-superordinate compound noun pairs.

2.30 Participants

Fourteen children between the ages of six and nine, with a mean age of seven years and five months (SD = 10.61 months), participated in this experiment. Participants were recruited and compensated in a similar fashion to Experiment One and Experiment Two.

2.31 Methods

The procedure for Experiment Three was identical to the procedure for Experiment Two except in the choice of objects within the trials. Since items such as the ones in Figure 1 include phonologically identical words (in Figure 1 this is *gingerbread*), the eye tracking data was thrown off to answer the command as ending with the overlapping segment (ex. "Point to the girl who has some of the gingerbread-" instead of "Point to the girl who has some of the gingerbread men"). Only a subset of the phonologically overlapping compound noun pairs were troublesome because the overlapping segment also defined a word found in both pictures (both the image of *gingerbread men* and *gingerbread house* contain *gingerbread* whereas *buttercups* and *butterflies*

do not both contain *butter*, see discussion of Experiment Two). These superordinate nouns were excluded from Experiment Three and replaced with other compound nouns with phonologically similar beginnings.

2.32 Results

Action responses were above 91% meaning that the participants were paying attention to the administered task. There was no significant interaction (determined by a 2x2 ANOVA, scale vs. strength) in looks to target (Figure 9).

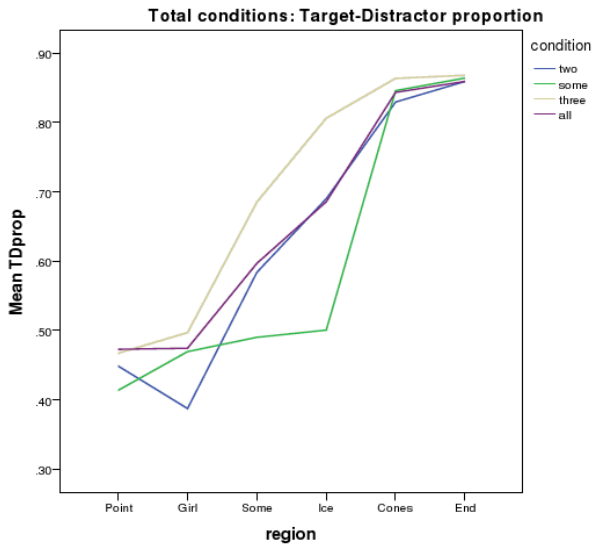


Figure 9 Looks to target divided by the sum of looks to target and looks to distractor, during sentence processing in Experiment Three.

Though this data seems to have no problems with compound noun structures or gender cues, there is a significantly accelerated interpretation of the quantifier *three* as compared to *two*, *some*, and *all*. There is a definite delay in the processing of the *some* quantifier and the scalar implicature does not appear to be made, as the looks to target do not significantly increase before the objects are phonologically disambiguated.

2.33 Discussion

With the acceleration of the interpretation of the *three* quantifier it is difficult to find a significant effect between strength and scale that would signify the delay in the implicature of *some*.

The gender cue confusion is no longer a problem with the new phrasing, and the superordinate noun objects have been eliminated leaving eye data that seems to respond appropriately at all points in the task: eliminating the incorrectly gendered characters, and then the distractor. From this data, the participants do not seem to be making the scalar implicature that adults do for the word *some* (Huang & Snedeker, 2009b).

3. Offline Data

Similar to the online data, the offline data was collected by the Tobii with an attached touch screen. Eye tracking data was not analyzed, only action responses. Offline and online data were collected to see the timeline and contexts for making and not making scalar implicatures.

3.1 Experiment One

Experiment One collected offline data on six to nine year olds' interpretation of the quantifier *some*. This experiment was identical to the one run in Huang, Spelke, and Snedeker (n.d.), apart from the age of the subjects run (in the previous study, the subjects were three year olds). The children are older in our study to match our online data and to cover the age range in which previous studies suggest scalar implicatures begin to be made.

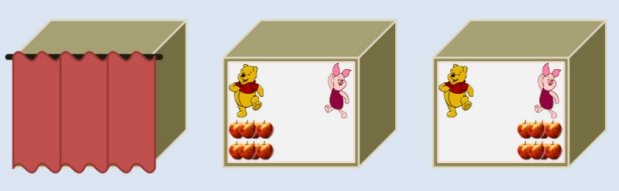
3.10 Participants

Thirty-eight children between the ages of six and nine, with a mean age of seven years and six months (SD = 14.01 months), participated in this experiment. Participants were recruited and compensated in the same way as for the online experiments.

3.11 Methods

In every trial the participant was presented with three boxes, one covered by a curtain and two showing their contents (Figure 10). The pictures were displayed on a Tobii eye tracker with an attached touch screen.

All vs. None

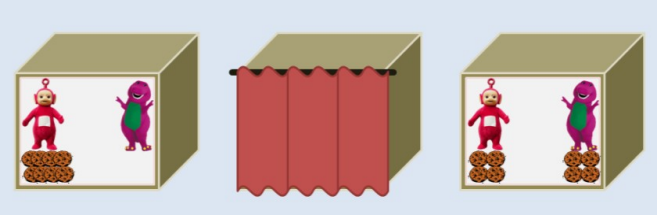


"Can you choose the box where Winnie the Pooh has some of the apples?"

Subset vs. Whole Set

"Can you choose the box where Big Bird has some of the lollipops?"

Subset vs. None



"Can you choose the box where Barney has some of the cookies?"

Figure 10 A display of all three trial types in the order they are shown in every condition.

Each participant started with a training period of several trials to make sure that he understood the idea of the covered box: that it held different contents than the other two boxes and that it could be, just as the other two uncovered boxes, selected as a response.

After the training period, the subject has three critical trials in which he or she would choose a definition of *some*: whole set versus none, subset versus whole set, and subset versus none (Figure 10). To introduce each trial, the child was introduced to the characters and the object on a separate screen. When the boxes were displayed, the child was asked whether he could choose the character with *some* of the object (phrasing shown in Figure 10). At this point, he or she was expected to touch the appropriate box on the screen. Within each subject each trial displayed different characters and different objects.

This experiment had three conditions. Each condition had the covered box in a different consistent location (left, middle, or right) in all three trials. In all conditions, the first trial always has the *none* option and whole set options uncovered, the second trial always had the subset and whole set options uncovered, and the last trial always had the subset and *none* option uncovered as a control trial.

3.12 Results

In trial one (all versus none) 76.32% of participants chose the box containing the specified character with *all* of the object, 2.63% chose the box where the character had *none*, and

21.05% chose the covered box. In trial two, (subset versus all) 52.63% of children chose the box with the character having *all* of the objects, 36.84% chose the character with the subset, and 10.53% chose the covered box. In trial three, (subset versus none) 73.68% of participants chose the box where the character had the subset, 5.26% chose the box where the character had none of the object, and 21.05% chose the covered box (Figure 11).

Figure 11 Percentages of response choices in all trials.

When separated by age (Figure 13, N(6) = 14, N(7) = 14, N(8) = 3, N(9) = 7), six and seven year olds preferred *all* to *none*, *all* to subset, and subset to *none* as definitions for *some*. For the subset versus *none* condition the six year olds preferred the covered box most of all and significantly more than seven year olds as established by a chi square test (χ^2 (1, N = 28) = 6.3, p <.05). The six year olds also chose the covered box significantly more than seven year olds in the subset versus *all* condition (χ^2 (1, N = 28) = 4.667, p < .05). Six year olds made up the age group with the most preference for the covered box across all trials (Figure 12). Eight and nine year olds differed from seven year olds in preferring the subset to *all* in the subset versus all condition (χ^2 (1, N = 24) = 6.171, p < .05, Figure 12).

Figure 12 A distribution of responses to all of the conditions in the box task divided in to age groups.

3.13 Discussion

37% of participants in this experiment were six year olds. The six year olds showed the highest preference towards the covered box through all three trials. The six year olds had a

tendency to perseverate on the covered box; if they selected the covered box on the first trial, they were likely to continue to choose it throughout the entire experiment.

With the remaining data of 7-9 year olds, there was a significant difference between ages. Seven year olds preferred the definition of *some* to mean the whole set, or the semantic meaning, where eight and nine year olds preferred the definition of *some* to mean the subset, or the inferred meaning. This suggests that implicature is being made at around age eight.

An issue with this task is that it did not show that the participants had a good understanding of the quantifier *all*. Without being sure that the participants understand *all*, we cannot be sure that the implicature is truly being made since a participant can't assume *some* and *all* are inconsistent if he doesn't understand the term *all* to begin with. If the quantifier *all* was correctly understood, this eight and nine year old implicature behavior is also cancelled in the *all* versus *none* trials like the adults' behavior in Huang, Spelke, and Snedeker's data (n.d.).

Though this data shows a difference between ages and implicature making, it is important to get rid of the perseveration of the six year olds in order to see whether they are actually accepting the whole set to mean some over the subset, or whether they just prefer selecting the covered box. It could be that since these six year old participants are younger, they prefer selecting the box that is different, or prefer the covered box simply because it holds unknown contents. Also during the training block, participants were encouraged to select the covered box. This could have resulted in perseverative behavior in six year old participants since they are the youngest of our participants.

We also had some question as to whether or not the participants understood the task as three separate boxes or one large picture. This is crucial to determine because a character can

have a whole set of an object within a box, but only a subset of it if compared to the other uncovered box as well. This is also a difference in design between the offline and online experiments. In the online experiments, the four quadrants holding characters and objects were compared to one another to determine if a character had a subset or whole set of an object, where in the offline experiments, the separate sections or boxes are meant to be considered as separate spaces. In order to get rid of the perseveration effects and ambiguousness of the design, we designed a second experiment to look at offline data.

3.2 Experiment Two

Experiment Two collected offline data on six to nine year olds' interpretation of the quantifier *some and all*. Experiment Two was similar to Experiment One, edited in order to assure that the participants understood the quantifier *all* and the task itself. This experiment was also designed to discourage perseverative participant behavior.

3.20 Participants

Eight children between the ages of six and nine, with a mean age of seven years and two months (SD = 12.03 months), participated in this experiment. Participants were recruited and compensated in the same way as all other experiments discussed in this paper.

Since this paradigm had not been tested previously, we also had seventeen adults participate. Each adult was compensated with five dollars for his or her time.

3.21 Methods

The methods of Experiment Two were similar to those of Experiment One. The experiment was now divided into two blocks (the order of the blocks was counterbalanced). The

three conditions in each block were differentiated by the location of the target, the same way as the three conditions were set apart in Experiment One.

Each condition had a *some* block and an *all* block. In the *some* block, questions were asked just as they were in Experiment One. In the *all* block, the questions remained the same, except instead they were asking for *all* of an object throughout the block instead of *some* of an object. The *all* block was put in place not only to make sure the participants understood the meaning of the quantifier itself, but also to make sure they understood that each box was to be treated as a separate entity. For example, a character can hold *all* of an object if he holds the whole set of it within that box, even if this is not the entire set of the object if we include other boxes on the screen.

The *some* and *all* trials displayed the same options in the same order as Experiment One. Between each of the three *some* or *all* trials in the block, were two number trials. These trials would ask for *two* or *three* of an object. If the participant selected the covered box on the first trial (*all* and *none* displayed, subset option covered), the number trial that followed would have the correct answer displayed in an uncovered box. Alternatively, if the subject selected the uncovered box on the first trial of a block, the second trial would have the correct answer be in the covered box to discourage repetitive selection of the covered box, a problem found in Experiment One.

3.22 Results

Both kids (96%) and adults (98%) were close to ceiling during the *all* block trials. showing that the participants all understood the task and the set up of the study, that each box represents a different entity. All following results are pertaining to the *some* block where

participants were asked to select the box in which a character had *some* of an object rather than *all* of it.

During the *some* vs. *none* condition both adults and kids preferred the box that showed a character with the subset of an object. There was no difference between the responses of children and adults (χ^2 (1, N = 25) = .324, p >.05). In the subset vs. whole set condition, both kids and adults chose the subset definition of the term *some* (χ^2 (1, N = 25) = .324, p > .05). For the *all* vs. none trial, the *all* box and the covered box were chosen about an equal amount of the time (around 50%) by both children and adults (χ^2 (1, N = 25) = 0.172, p >.05, Figure 13).

Figure 13 A distribution of responses of children and adults to each of the conditions in Experiment Two.

3.23 Discussion

Both children and adults responded close to perfectly during the *all* block trials showing that the participants all comprehended the task and that there was no confusion as to the boundaries between the boxed groups of objects and characters. Each participant understood that a character could have *all* of an object if the character had *all* of that object in that box, even if there was more of that object in another box on the screen. This also confirms that all participants understood the quantifier *all*.

From each condition, we can see that child behavior is similar to that of adults for this task. Both adults and children seem to prefer the subset definition of the word *some.* The confusion with this paradigm is with the *all* vs. *none* condition (Figure 14).

Figure 14 A comparison of child and adult responses to the *all* vs. *none* condition of the box task when the *all* block was presented first (left; Adults: N = 8, Kids: N = 5) and when the *some* block was presented first (right; Adults: N = 9; Kids: N = 3).

Though we are now left with even fewer participants represented in each chart (Figure 14), we can see that the trend to choose the covered box is higher when the *all* block is administered first. This result is different from Experiment One as well as Huang, Spelke, and Snedeker's (n.d.) work which showed that this set up should cancel out the implicature resulting in the preference of selecting the *all* definition of *some* rather than the covered box.

A likely explanation for this discrepancy is that having the *all* block administered first would prime the user to learn that display to be matched to the term *all* therefore using process of elimination (one box is now known as the *all* set up from the previous block, one box is known as the *none* set-up by definition) selecting the covered box. This could either mean that this order forces the implicature, reminding the participant that *all* could be said if that is what was meant, or that the implicature is avoided all together and the task becomes a matching game where the picture for *all* is already known and therefore the implicature process is avoided altogether and we are no longer measuring inferences.

The results of the subset versus whole set trial were consistent across block orders, showing that the kids and adults were both making the implicature not as a result of being primed by the *all* block. The only data result altered by the ordering of the blocks seems to be where we predicted the implicature would be cancelled in the whole set versus none trial from Huang, Spelke, and Snedeker's previous work (n.d.).

4. Conclusion

4.1 Our Data in the Context of Previous Findings

According to previous research, children start making scalar implicatures sometime shortly after the age of five (Noveck 2001, Papafragou & Musolino 2003, Pouscoulous 2007, Huang & Snedeker 2009a, Huang & Snedeker 2009b, Huang, Spelke, & Snedeker, n.d.). In order to investigate this further, we did offline and online analysis of scalar implicature processing with children just after age five (ages six to nine).

Our different paradigms found different results. The online experiment found that children through the age of nine were not making a scalar implicature for the word *some,* while our offline experiments showed evidence that children were making the implicature by age eight and possibly even earlier.

Possible explanations for this discrepancy include the division of the data as well as the differences in set up of each task. The online eye tracking data requires a greater amount of participants to be analyzed. A single participant's eye tracking data alone doesn't yield any coherent results, only the average eye tracking patterns of multiple participants is possible to evaluate. This is very different from the offline data, which can tell us about each individual participant as well as the whole group. Though the offline data does require many participants to yield significant results, it is easier to obtain enough data to split into age groups. Now that the online paradigm is finalized, this can be done with the eye tracking data once more data is collected.

Since we were able to divide the first offline experiment's data up into different age groups, we were able to see at which age implicatures might start being made. In the online data, the average age for the finalized experiment was seven, younger than the age at which the offline

data suggests implicatures start to be made. If scalar implicatures start to be made at age eight, the online data does not necessarily conflict with this result.

Another possibility is that the set up of each of these studies brings up different contexts for the implicature process. Just like the differences in the findings of the previous experiments discussed earlier (Huang & Snedeker, 2009a; Huang & Snedeker, 2009b; Huang, Spelke, & Snedeker, n.d.; Noveck, 2001; Papafragou & Musolino, 2003; Pouscoulous, Noveck, Politzer, & Bastide, 2007) can be brought back to the set up of each task, so can the findings in our two types of paradigms.

The box tasks are judgment tasks. Though the idea is to choose between the definitions of *some* as including the whole set or as limited to the subset, it can also be seen as accepting or rejecting *some* as the correct term in three stories, but only allowing the participant to say it is the correct term in one of them. Looking at it in this light, apart from the tasks our box tasks were modeled after, they are most like Papafragou and Musolino (2003) and Pouscoulous's (2007) works that agree that five year olds do not make these implicatures while seven year olds do. This is close to consistent with our offline data showing that implicatures are made at age eight if not earlier.

The online paradigm does not require the participant to reflect on the use of the implicature whatsoever; it only relies on natural processing. Perhaps without extra attention, younger children do not limit the definition of the word *some*. It is possible that children through age nine don't limit the interpretation of the word unless limitation is forced. Through experience, adults could have learned that it is more efficient to limit the definition, and then

cancel the implicature if necessary, whereas children still allow the broader definition of the term unless situations compel them to limit it.

Another reason for the different results may be the timing of the implicature. Though we allowed extra time for the implicature to be made in our online paradigm by using compound nouns, there is more time to make the implicature after the sentence is heard in the offline paradigm since there is no eye tracking data collected after the end of the sentence in the online experiment. Since the implicature is made in the offline paradigm and not during the online experiments, it is possible that it just takes kids much longer to make the implicature than adults.

4.2 Implications and Future Directions

Our data confirms previous suggestions that the context of scalar implicature has a significant impact on whether or not the implicature is made (Huang & Snedeker, 2009b). For instance, interrogative questions usually do not induce the making of an implicature. For example, if a speaker asks, “Did you take some of the cards?” the answer is still yes even if the listener took the whole deck. This opens up the possibility that younger children could have the capability of making such an implicature, or other implicatures, under the correctly primed or contextualized situation and allows insight into more ways to force the implicature in specific instances or teach these and other implicatures.

We are currently running these finalized paradigms on six to nine year old participants with autism spectrum disorder (ASD). According to Tager-Flusberg (2000), those diagnosed with autism have a variety of “spared and impaired” areas of language abilities, with one of the impaired areas being pragmatics. Even within the area of pragmatics it is unclear where people on the spectrum have difficulties. In previous research, Pijnacker et al. (2009) found that adults

on the autism spectrum made an equal amount of scalar implicatures as typical adults in the context of judging a set of underinformative sentences such as “Some sparrows are birds”. Yet, adults diagnosed with Asperger’s Syndrome (ASD with no language delay) gave more pragmatic interpretations than all other participants (understanding “some sparrows are birds” to be false).

The upcoming data will allow us to not only better understand populations on the spectrum, but will also allow us to delve deeper into what the differences are between our two experiments. Will participants with ASD perform better on one task than the other confirming that the two paradigms tap into two different contexts for the making of scalar implicature? Does the forced choice in the offline experiments have the same effect on participants with pragmatic deficits, or is that effect only seen with typically developing children? Or will they treat the two paradigms similarly avoiding pragmatics and using a strict definition for *some*?

The development of these paradigms allows us to look at the implicit and explicit processing of the term *some* with six to nine year olds. Our data suggests that when pushed, children can make scalar implicatures earlier than they do in a natural setting. That what they understand in everyday conversation is different than what adults understand, unless they are forced to or given more time to reflect on their interpretations.

References

Crain, S., & Thornton, R. (1998). Investigations in universal grammar: A guide to research on the acquisition of syntax and semantics. Cambridge, MA: MIT Press.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.). Syntax and semantics (Vol. 3, pp. 41–58). New York, NY: Academic Press.

Grodner, D., Klein, N. M., Carbary, K. M., & Tanenhaus, M.K. (2010). "Some," and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition*, 116, 42-55. doi:10.1016/j.cognition.2010.03.014

Huang, Y., Spelke, E., & Snedeker, J. (n.d.). *"What exactly do numbers mean?"*. Manuscript in preparation. Retrieved from http://www.unc.edu/~ythuang/papers/Number%20Semantics%20Full%2002Feb11.pdf

Huang, Y., & Snedeker, J. (2009a). Online interpretation of scalar quantifiers: Insight into the semantics–pragmatics interface. *Cognitive Psychology, 58(3)*, 376-415. doi: 10.1016/j.cogpsych.2008.09.001

Huang, Y., & Snedeker, J. (2009b). Semantic meaning and pragmatic interpretation in five-year olds: Evidence from real time spoken language comprehension. *Developmental Psychology, 45(6),*1723-1739. doi: 10.1037/a0016704

Kanan, C., Tong, M. H., Zhang, L., & Cottrell, G. W. (2009). 'SUN: Top-down saliency using natural statistics', Visual Cognition, doi: 10.1080/13506280902771138

Musolino, J., & Lidz, J. (2006). Why children aren't universally successful with quantification. *Linguistics* 44(4): 817-52. doi: 10.1515/LING.2006.026

Noveck, I. A. (2001). When children are more logical than adults: Experimental investigation of scalar implicatures. *Cognition*, *78*, 165–188. Retrieved from http://www.cs.columbia.edu /~vh/courses/LexicalSemantics/Scalar/noveck.children.pdf

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics–pragmatics interface. *Cognition, 86,* 253–282. Retrieved from http://repository.upenn .edu/cgi/viewcontent.cgi?article=1032&context=ircs_reports&sei-redir=1#search = %22noveck+2001%22

Pijnacker, J., Geurts, B., van Lambalgen, M., Kan, C. C., Buitelaar, J. K., Hagoort, P. (2009). Pragmatic inferences in high-functioning adults with autism and asperger syndrome. *Neuropsychologia, 47(3), 644-651*. doi: 10.1007/s10803-008-0661-8.

Pouscoulous, N., Noveck, I. A., Politzer, G., & Bastide, A. (2007). A developmental investigation of processing costs in implicature production. *Language Acquisition, 14,* 347–375.

Snedeker, J. (2009, Feb 13). *Prosodic and pragmatic processes in highly-verbal children with autism*. Grant Proposal. Cambridge, Massachusetts.

Tager-Flusberg, H. (2000). Understanding the language and communicative impairments in autism. *International Review of Research on Mental Retardation*, *Special Issue on Autism*. L.M. Glidden (Ed.) New York, NY: Academic Press.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten path effect:

Studying on-line sentence processing in young children. *Cognition, 73,* 89 –134.