
Testing Calculation of Scalar Implicatures in English and American Sign Language

KATHRYN DAVIDSON

University of California, San Diego
kdavidson@ling.ucsd.edu

IVANO CAPONIGRO

University of California, San Diego
ivano@ling.ucsd.edu

RACHEL MAYBERRY

University of California, San Diego
rmayberry@ling.ucsd.edu

ABSTRACT.

Scalar implicatures have been a subject of much debate concerning their place as part of either the grammatical computation of the semantics of a sentence or as post-compositional pragmatic reasoning. This paper describes three experiments addressing the semantic/pragmatic question of scalar implicatures from a developmental perspective, one in English and two in American Sign Language (ASL). The purpose of this work is three-fold: first, to test previous results in spoken language adult populations with a new experimental methodology to quantify scalar implicature calculation in ASL; second, to provide the first direct experimental comparison between constructions triggering standard scalar implicatures and constructions interpreted exhaustively, using English and ASL as testing languages; and third, to lay the baseline for and present results investigating the developmental issue of whether delayed first language acquisition affects semantic/pragmatic skills by testing deaf non-native signers on the ability to calculate scalar implicatures. Results show that adult native speakers of English and deaf adult native signers of ASL both interpret underinformative sentences in the constructions studied as having a scalar implicature. Underinformative sentences were accepted more often in constructions interpreted exhaustively than constructions which trigger standard scalar implicatures. Preliminary results are also presented for proficient deaf adult non-native signers of ASL, who behaved like deaf adult native signers of ASL, showing that the delayed first language acquisition of this group did not affect ability to interpret sentences with scalar implicatures in a native-like fashion.

1 Introduction

In his William James lectures of 1967, Grice outlined two separate types of inferences that can be made from a statement such as (1): logical entailment, as in (2a), and a separate type of inference, as in (2b)(Grice 1991).

- (1) Bob swims or Jim swims.
- (2) a. Somebody swims.
b. Bob and Jim don't both swim.

The inference in (2b) is known as an *implicature* to differentiate it from logical entailments. In Grice's formulation and subsequent reformulations the implicature is seen as a pragmatic non-grammatical enrichment to the sentence that arises through reasoning about the behavior of participants in a conversation (Horn 1989; Sperber and Wilson 1995; Levinson 2000). This view of how this type of meaning arises has been recently challenged by an account which argues that scalar implicatures are part of the grammatical semantic composition of a sentence (Chierchia 2004; Chierchia, Fox, and Spector 2008). This has led to a lively debate concerning the place of scalar implicatures in the semantics and pragmatic of natural language.

The main question to be addressed by this study is how deaf signers who are born to hearing families, and so acquire their first language in a nonnative manner, behave on this task of drawing implicatures that is at the border between semantic and pragmatic skills. This question was inspired by previous studies showing that normally hearing children have difficulty on such tasks (Noveck 2001). However, children and adults could show differences on scalar implicature calculation for any number of reasons, only some of which have to do with linguistic skills. In a review of the development of conversational reasoning, Siegal and Surian (2004) suggest that deaf children born into non-signing families could help researchers understand what developmental skills are needed for semantic/pragmatic tasks such as scalar implicature.

Before scalar implicature calculation can be studied in this population of non-native signers, it is necessary to set a baseline of how scalar implicatures are computed by deaf native-language users of ASL. Furthermore, a quantitative investigation of scalar implicatures in ASL requires a new testing paradigm for the visual/manual modality. Therefore, an experimental paradigm was devised and was tested on three separate populations: hearing native speakers of English, deaf native signers of ASL, and deaf non-native signers of ASL. Testing hearing speakers of English in the paradigm allows us to show that our methods replicate results of scalar implicature calculation found in previous research. Testing deaf native signers of ASL determines how scalar implicatures are calculated in the language by its native users. Finally, testing deaf non-native signers of ASL allows us to investigate how well people who learn their first language in a non-native manner are able to do on this task.

In terms of cross-linguistic generality, Grice's principles were not specific to English, but were intended as general rules of social interaction. Chierchia (2004) specifically states that we expect implicatures to be cross-linguistic, and all evidence points to this being the case. Thus, we predict that native speakers of English and native signers of ASL will both show similar patterns of scalar implicature calculation.

Deaf children who grow up in non-signing families show a variety of linguistic difficulties, not only related to their abilities in sign language, but also oral language, as well as reading (see Mayberry 2002 for overview). Understanding better the skills and limitations of this

population is important because the great majority of deaf people are born to hearing parents. Delayed development of Theory of Mind has also been reported, showing that not until ages 13-16 do most children who were deaf and raised in hearing families pass a standard Theory of Mind test (Russell, Hosie, Gray, and Hunter 1998). As adults, this population generally struggles with linguistic tasks but has plenty of social and educational opportunities for advancement, although it’s important to note that there is an extremely wide degree of variation in the backgrounds and abilities of deaf non-native signers. Because these participants were born deaf but raised in hearing families, they did not learn their first language in the usual native manner. Thus, it was hypothesized that the non-native signers, like young native speakers, would show more difficulty computing scalar implicatures. That is, they are predicted to accept more underinformative descriptions than native signers.

This study also compared speakers’ and signers’ calculation of implicature on different standard Horn scales, as well as sentences that contain lists which can often trigger exhaustivity implicatures, allowing for a first quantitative investigation comparing general exhaustivity with standard scalar implicatures. It was hypothesized that exhaustivity would pattern similarly to standard Horn scales, based on a variety of scalar implicature analyses that all treat exhaustivity using the same analysis used for standard scalar implicatures (Schultz and Van Rooij 2006; Spector 2007; Chierchia, Fox, and Spector 2008). Since the type of listing used in this set of stimuli in ASL involves locational classifiers, it is further hypothesized that scalar implicature calculation will be even more frequent (that is, there will be more rejections) in the general exhaustive case in ASL than in English, because the locational classifiers leave a part of the space unspecified, in addition to an unspecified object.

2 Stimuli and Design

This experiment contained two independent variables: *match condition* and *scale type*. ‘Match condition’ refers to whether the linguistic description matched the situation provided. This variable had three levels: Match, Mismatch, and Test, only the last of which contained an environment in which scalar implicatures are triggered. The ‘scale type’ variable had three levels: Quantifiers, Number, and Exhaustive¹. A fourth level, Logical Operators, was treated as filler and omitted from analysis due to pilot findings which showed that in ASL disjunction is not typically marked with a manual sign in declarative clauses, so no sentences using disjunction were used in the ASL stimuli.

Stimuli were created by crossing the two main factors (match and scale type) within a *situation template*, which consisted of 2 videos and 2 pictures. An example of a situation template is the ‘red can’ template, which was used in the $\langle all, some \rangle$ scale:

- An ‘All’ **photo**: a picture of three cans, all of which are red.
- A ‘Some’ **photo**: a picture of three cans, two of the cans are red and one is blue.

¹We use the term *exhaustive* for convenience to refer to the type of sentences seen in (5) and do not mean to imply that they form a standard scale.

- An ‘All’ **video**: a video in which the actor says “All of the cans are red.”
- A ‘Some’ **video**: a video in which the actor says “Some of the cans are red.”

Out of this template, three possible trials were created. The **match** trial consists of the ‘All’ photo presented with the ‘All’ video (the video description matches the photo). The **mismatch** trial consists of the ‘Some’ photo presented with the ‘All’ video (the video description does not match the photo). The **test** trial consists of the ‘All’ photo presented with the ‘Some’ video (the video is an underinformative description of the photo). An example sentence from each scale type is seen in (3)-(6). The scale types and sentences used for each scale were selected based on ease of comprehension, imageability, and naturalness both in English and ASL.

- (3) *Quantifiers*: All/Some of the cans are red.
- (4) *Numbers*: There are three/two bears.
- (5) *List w/Exhaustive Reading*: There is a wallet, (/a candle), and a globe.
- (6) *Logical Operators*: There is a spoon in the mug and/or the bowl.

There were a total of 48 ‘situation templates’, 12 for each scale (Quantifier, Number, and Exhaustive, as well as Logical Operators, which were eliminated from analysis), and thus a total of 144 possible trials (48 templates x 3 match conditions per template). The experiment was conducted as a within-subjects study. Trials per template were counter-balanced between participants so that each match condition in the template was seen by a third of the participants, but each participant saw only one match condition per template. Each participant saw a total of 48 experimental trials, one from each situation template. Photos were all taken with a Canon Digital Rebel XTi camera, while Video were recorded with a Flip Video digital video camera in artificial light.

The same experimental design was used in English and in ASL, except for one modification: for each participant in the ASL version, the four test trials in the Logical scale *< and, or >* were removed and replaced with 2 Match trials and 2 Mismatch trials in the Logical scale. Because both the Match and Mismatch trials used a video description with (implicit) conjunction (differing in whether the pictures matched or not), this eliminated possible issues concerning the naturalness of declarative sentences with overt disjunction discussed above. Since there was no test condition, these 12 Logical scale trials were treated as fillers and do not factor into the analysis below.

The actor in the English videos is a 25 year old male volunteer and a native speaker of English. The actor in the ASL videos is a 36 year old male who is hearing but grew up with deaf parents in a signing household and considers ASL his first language. He works as an interpreter in the San Diego area, and was paid for his time. All ASL videos were shown to a second native signer of ASL, who was Deaf, to confirm that they were natural and grammatical sentences of the language.

3 Procedure

The experiment was carried out on an Apple Intel Macbook with a 13"3 monitor, in a quiet room, with the participant sitting comfortably at the laptop. The experimenter was present in the room but not watching the participant's responses. In the English version sound was conveyed through Sony Koss over-ear headphones. The entire experiment was conducted within the free software program Psyscope.

Instructions were given in a video on the computer screen by the same actor that appeared in the trials in each video. Besides the usual disclaimers and directions, the instructions asked the participants to judge whether 'you think the video description matches the scene in the picture, you are satisfied' or whether 'you do not think the description matches the picture, you are not satisfied.' After the initial set of instructions, participants were given three practice trials to become familiar with the task.

A single trial consisted of a picture in the left side of the computer screen, and a video on the right side of the screen. Each trial participants used the 'space bar' to play the video, which would only play once. After watching the video, and determining if they were satisfied with the description of the sentence, they pressed a key covered with a smile face. If they felt the description was wrong, they pressed a key covered with a drawing of a frown face. Each participant saw 48 total trials. Responses and reactions times were recorded to a data file at the end of the experiment.

4 Participants

Hearing English speakers were recruited from the undergraduate Psychology, Linguistics, and Cognitive Science class experimental pool at the University of California, San Diego. Participants ranged in age from 18-30, and all were raised in the United States and were native speakers of English. There were a total of 12 hearing English speaking participants.

Deaf native signers were recruited by word of mouth in the San Diego metropolitan area. All were profoundly deaf, raised in the United States, and had deaf parents with whom they communicated in ASL. Since a great majority of deaf children are born to hearing parents, the number of deaf adults who are native signers of ASL is comparatively small, and thus there were only 5 deaf native signing participants. Participants ranged in age from 24-59 and all had at least some university education. Each were paid twenty dollars in gift certificates or cash for their time in this study.

Deaf non-native signers were also recruited by word of mouth in the San Diego metropolitan area. They learned to sign between ages 7 and 18 and learned to communicate with hearing people in English earlier, such that all of them considered themselves dominant in English at some point in their history, but currently sign as their primary language for everyday communication both at work and at home. All had some advanced education beyond high school, and were either still in college or were working as staff at a high school. There were a total of 5 non-native signing participants, who ranged in age from 28-59, and all were raised in the United States.

5 Results

5.1 Native Speakers of English

For the native speakers of English, the first dependent measure is the acceptance rate of the descriptions, i.e. the proportion of smile responses. The acceptance rates were entered into an analysis of variance (ANOVA) with one factor: match condition (levels: Match, Test, Mismatch). Each level represents the mean for each participant across the three scale types (Number, Quantifier and Exhaustive). The analysis revealed a strong effect of match condition ($F(2, 12) = 145.1, p < 0.0001$) (See Figure (1.1)). Bonferroni's Multiple Comparison Test revealed a significant difference between each pair of match conditions.

Acceptance Rates for English Speakers (n=12)

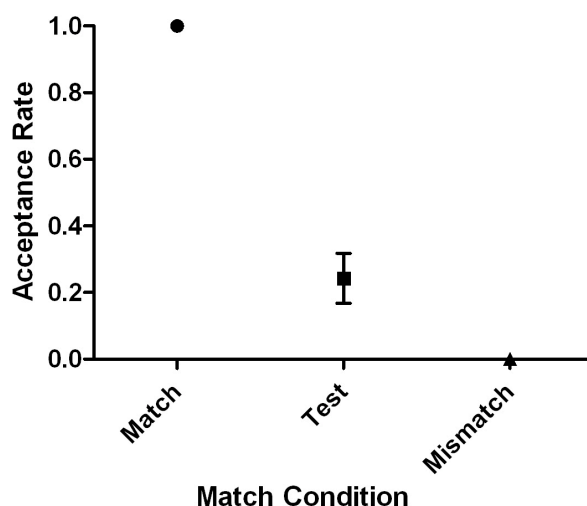


Figure 1.1: Acceptance Rates of Native Speakers of English Averaged Across 3 Scales: Numbers, Quantifiers, Exhaustive

A second analysis of variance was conducted for the factor of scale type (levels: Number, Quantifier, Exhaustive). This was not combined with the previous analysis in a two-way ANOVA because it was not felt that comparing scale types across Match and Mismatch conditions was meaningful. This is because Match conditions were always accepted and Mismatch conditions were always rejected, and there is no theoretical basis for assuming that scale type would interact with whether true (Matching) or false (Mismatching) sentences are judge to be true or false (native speakers should interpret all true and false sentences correctly, as they did). Instead, a second analysis of variance (ANOVA) was conducted using only the data in the Test conditions with only one factor: scale type (levels: Number, Quantifier, Exhaustive). The analysis revealed an effect of scale type

($F(2, 12) = 6.429, p < 0.01$)(See Figure 1.2).

Test Condition Acceptance Rates for English Speakers (n=12)

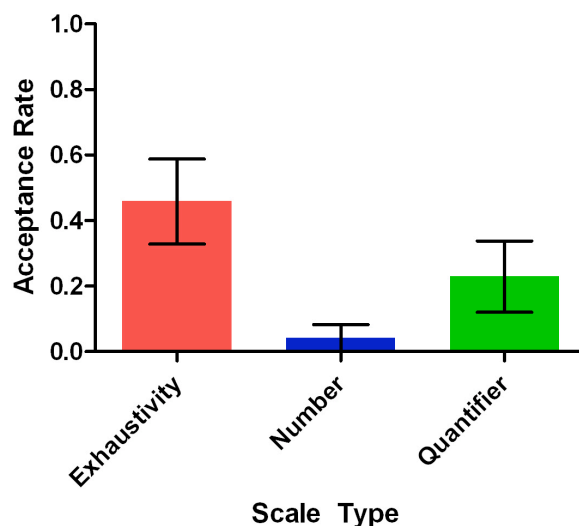


Figure 1.2: Mean acceptance rate for the test conditions in each of the three scale types, where acceptance indicates lack of scalar implicature calculation

Bonferroni's Multiple Comparison Test revealed a significant difference between the Exhaustivity and Number scales, but not the Exhaustivity and Quantifier scales. There was a trend towards significance between the Number and Quantifier scales, as seen in Figure (1.2).

Although this experiment was not designed as a reaction time study, an analysis was also conducted using the reaction time as a dependent variable. Reaction times were recorded in Psycscope as the time between the video onset and the time that a smile or frown response is recorded. Because the sentences in which the different types of scales were used were of different lengths, reaction times between scales cannot be compared directly. However, it is possible to look at the reaction times within a scale for rough comparisons between match conditions. Reaction times can show if scalar implicatures were indeed more difficult to process, as we expect based on previous research (Bott and Noveck 2004).

Because we cannot compare reaction times between scales, three separate one-way ANOVAs were conducted, one for each scale. In each scale, there were three levels corresponding to the Match Conditions: Match, Test, and Mismatch. Results for the reaction time ANOVA in the quantifier scale were significant, ($F(2,12) = 3.533, p < 0.05$). Results for the ANOVA on reaction times in the Number scale were not significant ($F(2, 12) = 2.063, p > 0.05$). Results for the reaction time ANOVA in the Exhaustive condition were also not significant, but there was a trend towards significance ($F(2, 12) = 2.695, p =$

Scale	Match	Test	Mismatch
Quantifiers	3.66 (0.98)	*4.15 (1.31)	3.62 (0.98)
Numbers	3.59 (1.61)	3.85 (1.21)	3.33 (0.95)
Exhaustive	4.45 (1.45)	4.85 (1.58)	4.15 (1.20)

Table 1.1: Mean reaction times for each scale are given in seconds, with standard deviation in parenthesis. (* $p < 0.05$)

0.089). Although only reaching significance in one scale, it is the case that the test condition took longer than the other two conditions in every scale, as can be seen in Table (1.1).

5.2 Native Signers of ASL

Acceptance Rates for Native Signers of ASL (n=5)

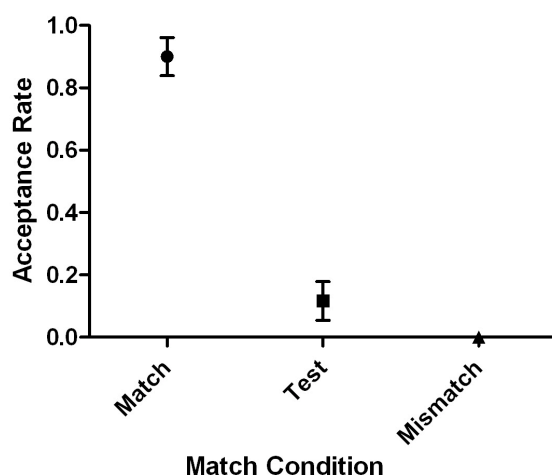


Figure 1.3: Mean acceptance rates of native signers of ASL for the three scales together: Numbers, Quantifiers, Exhaustive

As for native English speakers, the first dependent measure for native signers of ASL is the acceptance rate of the descriptions, i.e. the proportion of smile responses. The acceptance rates were entered into an analysis of variance (ANOVA) with one factor: match condition (levels: Match, Test, Mismatch). Each level represents the mean for each participant across the three scale types (Number, Quantifier and Exhaustive). The analysis revealed a strong effect of match condition ($F(2, 4) = 81.59, p < 0.0001$) (See Figure 1.3). Bonferroni's Multiple Comparison Test revealed a significant difference between Match vs.

Test conditions and between Match vs. Mismatch conditions, but no significant difference between Test and Mismatch conditions.

Because the sample size for native signers was small ($n=5$), there were no significant differences in an analysis of variance with a factor of scale type. Though not significant due to the large amount of variance, there was a small trend for the exhaustive scale to be accepted more often than the other two scales. The small sample size also prevented reaction times between match conditions or scales from showing any significant differences.

5.3 Non-Native Signers of ASL

Acceptance Rates for Non-Native Signers of ASL ($n=5$)

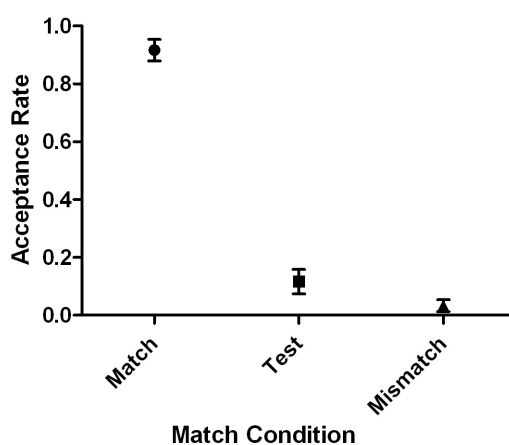


Figure 1.4: Mean acceptance rates of deaf non-native signers of ASL averaged across three scales: Numbers, Quantifiers, Exhaustive ($n = 5$)

As above, the first dependent measure for non-native signers of ASL is the acceptance rate of the descriptions, i.e. the proportion of smile responses. The acceptance rates were entered into an analysis of variance (ANOVA) with one factor: match condition (levels: Match, Test, Mismatch). Each level represents the mean for participants across the three scale types (Number, Quantifier and Exhaustive). The analysis revealed a strong effect of match condition ($F(2, 5) = 144.7, p < 0.0001$)(See Figure 1.4). Bonferroni's Multiple Comparison Test revealed a significant difference between Match vs. Test conditions and between Match vs. Mismatch conditions, but no significant difference between Test and Mismatch conditions.

Because the sample size for non-native signers was small ($n=5$), there were no significant differences in an analysis of variance with a factor of scale type except that the Number scale was never accepted in the Test condition. The small sample size also prevented reaction times between match conditions or scales to show significant differences.

6 Discussion

As stated at the outset, this study has three goals: first, a cross-linguistic, cross-modality investigation of scalar implicature calculation in ASL, which required a new experimental paradigm; second, a quantitative comparison of list constructions interpreted exhaustively compared to constructions that trigger standard scalar implicatures; and third, testing calculation of scalar implicatures by deaf non-native signers of ASL, who acquire their first language in a non-native manner.

Since many factors have been shown to influence behavior on a task such as scalar implicature (eg. linguistic and non-linguistic context, experimental instructions, stimuli, etc.) it was important to determine that this experimental paradigm accurately replicates findings of adults on scalar implicatures. The combination of the hearing speakers of English and the deaf signers of ASL show this to be the case. Deaf native signers of American Sign Language reject underinformative sentences which are true as confirmed by a consultant, but which are infelicitous if a scalar implicature is calculated. Thus, these participants compute scalar implicatures, as all theories would predict but which has not been shown in the visual-manual modality before this study. Not only does this study confirm that native signers in ASL do reject underinformative descriptions/compute implicatures, but they do it in a very similar pattern to native speakers of English, as seen in the overwhelming rejection of underinformative descriptions involving the Number Scale, compared to the Exhaustive and Quantifier scales.

As for the exhaustive interpretation of the list constructions, what we can tentatively conclude is that in both English and ASL, underinformative list descriptions tend to be accepted more often than standard scalar underinformative descriptions. It was hypothesized that the Exhaustive scale in ASL might be different than the one in English because it used locational classifiers, but this did not seem to have an effect. The requirement to fully describe a spatial situation does not seem to be any stronger than the general tendency for exhaustive interpretations of lists in the given context. However, because this experiment did have a small sample size, more participants should be included in the study before drawing definite conclusions on the issue of exhaustivity with locational classifiers.

Although the sample size was small and will be expanded in future research, the data shows that adults who grew up in non-signing families interpret statements using the quantifier scale $\langle all, some \rangle$, numbers $\langle three, two \rangle$ and lists $\langle 3 - items, 2 - items \rangle$ with the same pattern as native signing adults. This includes the Test Condition trials, which were rejected at rates significantly different from the rejection rates of the Match Condition, showing that non-native signers do seem to compute simple scalar implicatures. Differences between scales did not reach significance, but the trend is towards exactly the same pattern seen in the native signers, in which the Test condition of the Number scale is universally rejected. This fits with previous results showing that even young children that accepted underinformativity in other scales, such as quantifiers, reject it when seen in numbers (Papafragou and Musolino 2003). The fact that even these differences in accepting and rejecting underinformativity are mirrored in native and non-native signers show that these non-native signers are competent in interpreting these constructions. It would be

important for future research to investigate the abilities of younger deaf non-native signers, who have weak language skills, to determine if there is a further delay in acquisition of implicature above the delay seen in hearing children acquiring a spoken language and deaf children natively acquiring a sign language.

Another issue lurking beneath the surface of experimental investigation in ASL is bilingualism. Because both the deaf native signers and the deaf non-native signers are living in the United States and have had many years of education, they are all bilingual to some extent in English. This could be relevant to the results above because young Slovenian/Italian bilingual children were shown to have an advantage over their monolingual counterparts in tests of pragmatic conversational skills in what seems to be the first study of the effect of bilingualism on scalar implicature calculation (Siegal, Iozzi, and Surian 2009). Some of the non-native adults tested in this study considered themselves dominant in a signed form of English, because they learned ASL at such a late age. Therefore, because the experiments in this study all tested bilingual adults, it would be important to see if bilingual adults behave any differently than monolingual adults on this type of task by testing hearing bilingual adults, such as Spanish-English speakers.

Younger non-native signers and bilingual hearing adults are just two of the groups in which scalar implicature calculation should be better understood before truly understanding the results of a study of non-native signers in an investigation such as this one. Furthermore, it would be useful for future research to have independent evidence for participants' linguistic deficits beyond self-reports and general patterns seen in the population. However, the results of the current study show that it is not necessary to have acquired a language in a native manner to have the appropriate interpretations for sentences that trigger scalar implicatures. If the results seen here hold after more participants are added to the study, we can conclude that this supports a theory in which the difficulty with scalar implicature calculation in children is due to a non-linguistic difficulty such as short term memory, pragmatic reasoning, or the ability to compare alternatives online and not necessarily the linguistic complexity of a scalar implicature construction.

Acknowledgments

Many thanks to Brandon Scates and Jeff Zaremba for being extraordinarily patient and willing to contribute to this project. Thanks also to David Barner and Matt Hall, and Susan Fischer for helpful discussion and advice about doing experimental work in signed languages. This project was funded by the National Institute of Health predoctoral training grant #T32DC000041 through the Center for Research in Language at UCSD.

Bibliography

- Bott, L. and I. Noveck (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language* 51, 437–457.
- Chierchia, G. (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In A. Belletti (Ed.), *Structures and Beyond*, Volume 3 of *The Cartography of Syntactic Structures*, Chapter 2, pp. 39–103. Oxford University Press, Oxford, UK.
- Chierchia, G., D. Fox, and B. Spector (2008). The grammatical view of scalar implicatures and the relationship between semantics and pragmatics.
- Grice, P. (1991). *Studies in the Way of Words*. Harvard University Press.
- Horn, L. R. (1989). *A Natural History of Negation*. Chicago: University of Chicago Press.
- Levinson, S. C. (2000). *Presumptive Meanings*. MIT Press, Cambridge MA.
- Mayberry, R. I. (2002). Cognitive development in deaf children: the interface of language and perception in neuropsychology. In S. J. Segalowitz and I. Rapin (Eds.), *Handbook of Neuropsychology*, Volume 8. Elsevier Science B. V.
- Noveck, I. (2001). When children are more logical than adults: experimental investigations of scalar implicature. *Cognition* 78, 165–188.
- Papafragou, A. and J. Musolino (2003). Scalar implicatures: experiments at the semantics-pragmatics interface. *Cognition* 86, 253–282.
- Russell, P., J. A. Hosie, C. D. Gray, and N. Hunter (1998). The development of theory of mind in deaf children. *Journal of Child Psychology and Psychiatry* 39(6), 903–910.
- Schultz, K. and R. Van Rooij (2006). Pragmatic meaning and non-monotonic reasoning: The case of exhaustive interpretation. *Linguistics and Philosophy* 29, 205–250.
- Siegal, M., L. Iozzi, and L. Surian (2009, Jan). Bilingualism and conversational understanding in young children. *Cognition* 110(1), 115–122.
- Siegal, M. and L. Surian (2004, Dec). Conceptual development and conversational understanding. *Trends in Cognitive Sciences* 8(12), 534–538.
- Spector, B. (2007). Scalar implicatures: Exhaustivity and gricean reasoning. In M. Aloni, A. Butler, and P. Dekker (Eds.), *Questions in Dynamic Semantics*, Current Research in the Semantics/Pragmatics Interface. Elsevier.
- Sperber, D. and D. Wilson (1995). *Relevance*. Oxford: Blackwell.