Learning from speaker word choice by assuming adjectives are informative

Alexandra Horowitz
ahorowit@stanford.edu
Department of Psychology
Stanford University

Michael C. Frank
mcfrank@stanford.edu
Department of Psychology
Stanford University

Abstract

Pragmatic abilities are not only a component of efficient communication; they can also be an important learning mechanism for young children. We discuss four experiments and a corpus analysis to investigate whether children and adults can infer information about a speaker's knowledge based on the choice of an adjective. In Experiments 1 – 3, we found that adults are sensitive to adjective use as an indicator of intended contrast dimension (e.g. that people say "red" if an object could have been blue, but "tall" if it could have been short). In Experiment 4, we found developmental differences between older and younger 4-year-olds: older children were above chance at selecting the referential dimension of interest, while younger children exhibited some contrast inference but a strong color bias. This suggests that by preschool, children are beginning to make inferences from a speaker's word choices, but that there are differences between adjective types. We conducted an exploratory corpus analysis to investigate possible causes for this developmental difference.

Keywords: Pragmatics; adjectives; language development.

Introduction

A key feature of human language is its ability to convey information efficiently in context. For adults, the ability to make pragmatic inferences—extrapolations about meaning in context—can dramatically facilitate the exchange of information between conversational partners. For example, from "I can't find my left shoe," we can make the inference that the speaker probably knows where her right shoe is, or else she would have mentioned that both of her shoes were lost. For children, sensitivity to word choice information is instrumental in early word learning; pragmatic inference helps guide both language acquisition and comprehension. In general, recognizing that speakers have chosen to say something in a particular way *because of some communicative goal* is an integral part of understanding language (H. Clark, 1996).

In addition to aiding in the acquisition and comprehension of language, inferences about the pragmatic implications of speakers' wording decisions might also be an important learning mechanism for children. Following our shoe example further: Children who learn to infer implied information embedded in word choices can incorporate implicit knowledge (we are talking about one specific shoe) and move on to acquire additional information ("where was it last seen?") rather than spending time repeating and confirming implied details ("only one shoe is lost?"). Using speakers' word choices to make broader inferences allows children to learn from both what is stated and what are implied alternatives. The earlier and faster children can recognize implicit contrasts from word choice, the greater their opportunities to make use of this information.

The goal in our studies was thus to investigate whether children can learn from how speakers choose to express themselves. We focused on adjective use as a case study because adjectives are optional and may signify cues to contrast and noteworthy features. As an extension of the principle of contrast—that a contrast in form signifies a contrast in meaning (E. Clark, 1987)—children should interpret that referential expressions modified with adjectives convey different types of information than expressions with bare nouns. This makes adjective interpretation a useful domain for examining how children form implicit inferences from a speaker's word choices. We began by looking at color and size terms because they are some of the earliest-learned and most commonly-used adjectives.

Adults perceive adjective use as marking contrast, but there are differences between their comprehension and production. Grice's maxim of quantity—that speakers should be only as informative as is necessary—predicts that modifiers should be used selectively to disambiguate target referents from contrast sets (Grice, 1975). Indeed, in comprehension, visual search findings reveal that adults process color and size information as it comes online and are faster to locate a modified referent (e.g. "big comb") when a contrasting competitor item is present (e.g. a small comb) than when a distractor item is present (e.g. a spoon) (Sedivy, Tanenhaus, Chambers, & Carlson, 1999). This indicates that prenominally modified expressions may evoke a contrast set with the referent; adults are sensitive to implied contrast information embedded in adjective use and process prenominal modifiers incrementally to locate and disambiguate a speaker's intended referent. However, adults are not always Gricean in their production; although they rarely produce scalar modifiers without a size contrast set present, they frequently over-modify with color terms (Grodner & Sedivy, 2005; Sedivy, 2003). In all, adults are sensitive to the implications of adjective use, but they are not always maximally informative.

Children seem to be developing similar skills in their preschool-age years. By kindergarten, children can recognize the informativeness of adjective use in both comprehension and production; children are more likely to use an adjective to uniquely identify a big cup from a small cup than when only a single cup is present (Nadig & Sedivy, 2002), suggesting that they are able to consider what level of description is most useful to the perspective of an interlocutor. Preschoolers can also learn to produce unambiguous references for more complex scenes with feedback (Deutsch & Pechmann, 1982). In processing speech, 3-year-olds make use of adjective information as soon as it becomes available; they correctly look to

the bigger of two cars upon hearing "big", even before they hear the word "car" (Fernald, Thorpe, & Marchman, 2010).

In sum, preschoolers actively process adjective information in production and comprehension as cues for uniquely identifying referents. But can children apply this knowledge to make inferences about what implicit information is conveyed by a speaker's choice of adjective use? To our knowledge, this is the first investigation of children's abilities to use referential expressions to make inferences about the broader context.

In this paper, we outline four experiments and a corpus analysis investigating pragmatic inferences from information contained in speaker word choice. In Experiment 1, we used a novel task to examine whether adults use adjective information to infer referential contrast, and found that they performed equally strongly with color and size terms. We slightly modified the language in Experiment 2 to make implied contrast less salient, and found that performance only decreased slightly. In Experiment 3, we examined performance with other context-dependent and context-independent features, and found no differences across modifier types. In Experiment 4, we extended our task to 4-year-olds, and found a developmental change between children younger and older than 4-and-a-half: while older 4s performed above chance, younger 4s showed a strong color bias. We conducted an exploratory corpus analysis to examine possible causes for this developmental difference.

Experiment 1

Our goal was to examine what information listeners can infer about why a speaker chose to form an utterance in a particular way. Before studying behavior in children, we wanted to confirm our intuitions that adults interpret adjective information contrastively when more than one possible visual contrast is available. In Experiment 1, we used a novel task in which adjective choice was the only informative cue to referential contrast (Figure 1). If adults are able make inferences about a speaker's intended contrast set from adjective information, then they will be more likely to infer contrast along the referenced dimension rather than another visible but unstated dimension. This is precisely what we found.

Methods

Participants Seventy-two adult participants were recruited from Amazon's Mechanical Turk (MTurk). They were informed that the task was designed for children. All reported that they were native speakers of English.

Stimuli Participants viewed an online storybook with cartoon images of aliens, with one test trial. The test featured a set of three aliens: a training exemplar, and a test set consisting of an alien identical to the training exemplar except for size and an alien identical to the training exemplar except for color (Figure 1). We used the height dimension to reflect size because piloting revealed that scaling total size was construed as reflecting age rather than size.

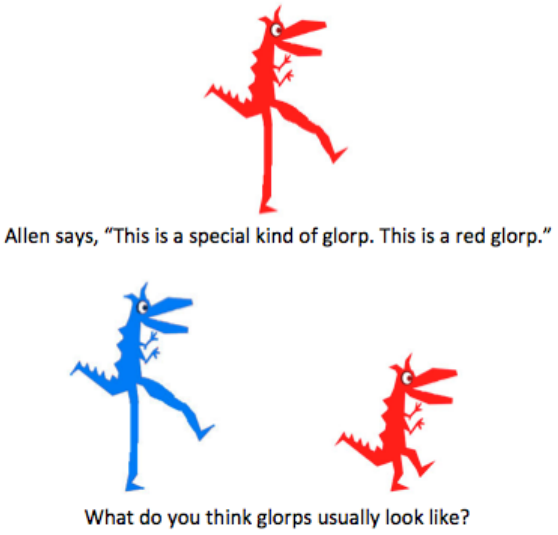

Figure 1: Participants saw a set of three aliens: a training exemplar (e.g. tall red) followed by a test set containing one alien that differed only by color (e.g. tall blue), and one that differed only by size (e.g. short red). Participants were told that a character uttered an expression modified by an adjective (color or size), and asked to predict which test exemplar represented what that kind of alien usually looked like.

Procedures Participants were introduced to a character named Allen the Alien. Allen presented the training exemplar, and said something about it, e.g. "This is a special kind of glorp. This is a red glorp." For a third of participants the reference was about color, and for another third the reference was about size. Participants then saw the test set and were asked, "What do you think glorps usually look like?", and prompted to select one of the two images. We measured whether adults used the adjective information to infer a contrast along the referenced dimension. For the remaining third of participants, we measured responses to a bare noun baseline in order to detect any response biases unrelated to adjective information.

Results and Discussion

The baseline control revealed that when adults had no access to adjective information, they were exactly at chance for the image they selected. Of the 24 baseline participants, they selected the size contrast half of the time (n=12) and the color contrast equally as often (n=12). Additionally, participants were equally likely to select an image on either the left (n=12) or right (n=12) side regardless of counterbalancing (n=6 for each combination of contrast type and side presentation). These results indicate that adults showed no selection bias with our stimuli when only visual cues were available.

For the experimental trials, responses were coded as correct if participants selected the alien that differed along the referenced dimension (i.e. chose the alien that differed by size

in size trials, and color in color trials). Participants selected the contrasting dimension more often than chance ($p < .01$ in an exact binomial test for both conditions) and performance did not differ across the two conditions ($\chi^2(1) = 0.10$, $p = .75$, Figure 2).

Adjective information was used informatively to denote contrast along a specified dimension; adults interpreted color or size reference as conveying a noteworthy or contrastive feature that distinguished different aliens. Although adults often have access to prosody and emphatic stress on prenominal adjectives as additional cues to contrastive focus in speech (Ito & Speer, 2008), they were sensitive to the implications of adjective use even in our written task. The overall high level of performance suggests that adults are attuned to the subtle informativeness of word choice information. Though the only difference across all conditions was the use of a size term vs. a color term to describe the training exemplar, adults were able to detect and apply this information when making generalizations about the broader populations. They effectively interpreted the single difference in word choice across trials as signaling either a color or size contrast.

Experiment 2

In Experiment 1, we tried to make cues to contrast highly salient by drawing attention to the unique referent (i.e., "this is a special kind of alien"). In Experiment 2, we ran the same procedure but used a more neutral introduction ("this is a glorp") in order to determine whether adults would still use adjective information to infer a contrast dimension. Performance decreased only slightly from Experiment 1.

Methods

Participants Forty-eight MTurk workers were recruited.

Stimuli and procedure Stimuli were identical to Experiment 1. The only change to the procedure was a modification of the referential expression. Instead of Allen saying, "This is a special kind of glorp. This is a red glorp," we changed the wording to, "This is a glorp. This is a red glorp."

Results and Discussion

Again, adult participants performed significantly above chance ($p < .05$ in an exact binomial test for both conditions) with no difference between conditions ($\chi^2(1) = 0.11$, $p = .74$, Figure 2). Adults used adjective information to infer a referential contrast dimension even when more overt cues to contrast were removed.

Experiment 3

In our next experiment, we tested whether inferences about contrast from adjective use would generalize to other properties that are less ubiquitous than color and size. We re-ran the procedure keeping color constant, but used the dimensions of texture (which, like color, remains constant across contexts) and width (which, like size, is relative to context).

Methods

Participants Ninety-six MTurk workers were recruited.

Stimuli and procedure Stimuli were similar to Experiment 1, but color was held constant and other contrasts were added. Participants either saw aliens that differed by height and texture (spiky vs. smooth), or by height and width (fat vs. thin). We again used language to highlight contrast salience, e.g. "This is a special kind of glorp. This is a [spiky] glorp." A height, texture, or width adjective was used.

Results and Discussion

Responses were coded as correct if participants selected the alien that differed along the referenced dimension. Adult participants performed significantly above chance in exact binomial tests ($p < .05$) for all conditions, and with no differences between conditions ($\chi^2(2) = 0.55$, $p = .76$, Figure 2). This result suggests that adults are able to infer relevant contrast information from adjective use across a variety of context-dependent and -independent dimensions. Our findings thus show that adults are sensitive to pragmatic inference from adjective use across a variety of adjective types. Our next experiment tests children's sensitivity.

Experiment 4

To assess children's ability to use adjective choice to infer contrast, we used a similar paradigm. We focused on color and size contrasts with 4 – 5 year-old children because this has been found to be an age of pragmatic development in other studies (Barner, Brooks, & Bale, 2011). If young children infer the potential informativeness of adjective use to convey cues to contrast, then they should be more likely to select the image that differs along the referenced dimension (i.e. color or size). If they are not sensitive to this information, then they should select an image at random or according to a baseline bias for one dimension or the other.

Methods

Participants We recruited 46 four-year-old children from the Bing Nursery School at Stanford University. Pilot testing suggested response differences by older and younger 4-year-olds, so we recruited two age groups: twenty-four children age 4.0 – 4.5 (mean age 4 years 2 months) and 22 children age 4.5 – 5.0 (mean age 4 years 10 months).

Stimuli We used a similar task design to the previous experiments, but printed a physical book that the experimenter read with children. Each child received two training and six test trials. Each test trial used a unique set of three aliens: a training exemplar alien and two test exemplar aliens that differed from the training alien each by only color or only size.

Procedures The experimenter read the book to each child individually in a quiet room at the Bing Nursery School. Children were introduced to Allen the Alien, and then completed two training trials with familiar items to get them used to the study design (e.g., "This is a special kind of milk. This is

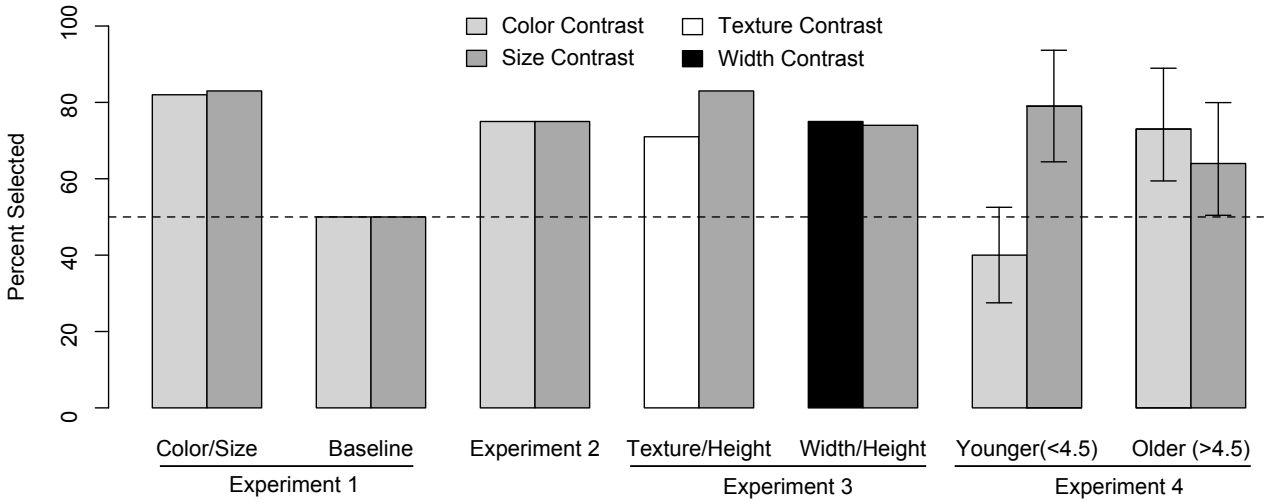

Figure 2: Mean percent correct performance across all conditions of Experiments 1-4. The dashed line represents chance (50%). Error bars represent 95% confidence intervals.

chocolate milk. What does milk usually look like?"). Training trials featured adjectives other than color and size and only one relevant contrast choice (e.g. plain milk vs. orange juice). If children did not select the correct training image, they were prompted until the correct image was chosen.

Training was followed by six test trials. For each test trial, the child was shown a picture of a single exemplar alien and told something about it, e.g. "This is a special kind of glorp. This is a tall glorp." Children were then shown two pictures, one of an alien that differed from the exemplar by size, and one that differed by color. They were asked, "What do you think glorps usually look like?" Children received six test trials: two trials using size adjectives (e.g. "this is a tall glorp"), two trials using color adjectives (e.g., "this is a red glorp"), and two trials with no adjective to serve as a baseline (e.g., "this is a glorp"). Adjectives were focused with contrastive stress. The experimenter averted her gaze while children pointed to their response. The sessions were video-recorded. Trial types and test pictures were counterbalanced across children, and alien sets were presented in one of two orders. The task took about 10 minutes to complete.

Results and Discussion

Preschool-aged children did show sensitivity to the implications of adjective use. Overall, preschoolers could pick out adjective information as marking an implied contrast dimension. Nevertheless, we saw an interesting developmental trend in the data (Figure 2).

We analyzed our results using a logit mixed model, predicting correct responses as an interaction between age (older vs. younger) and contrast type with random effects of participant and alien type. There was a significant effect of age, such that older children performed better than younger children ($\beta = -2.06$, $p < .001$). There was also a significant interaction between age and contrast (color or size) such that older children performed above chance for both color and size trials, and younger children responded above chance for size trials but were only at chance for color trials ($\beta = -3.25$, $p < .01$). Overall, this analysis suggested weak responding by the younger 4s with successes by the older 4s.

To ensure that performance differences were not due to unfamiliarity with the color and size terms, we ran a posttest with a subset of children for each age group (n=13 younger, n=12 older). Younger children produced the correct size term over 80% and color terms 95% of the time. Older children's production was 94% for size and 99% for color. These data suggest that younger children's lower performance on color trials was not a result of not knowing their color words.

Baseline responses also indicated a significant developmental change. While younger 4s chose color-matching targets 71% of the time on the baseline trials, older 4s chose color-matching targets 38% of the time. In fact, nearly half (n=10) of the younger children selected a color match for all trials in the study, while only one of the older children did. To examine this effect, we replotted our data by proportion of trials on which the color-matching target was chosen (Figure 3). A correct pattern of responding for a color trial would be choosing the size match (hence success on a color trial would be below 50% in Figure 3), while a correct response on a size trial would be choosing the color match (above 50%). Replotted in this manner, we can see that younger 4s are modifying their responses only slightly for color trials and not at all for size trials, while older 4s modify their responding slightly for color trials and considerably for size trials.

We captured this pattern with a second logit mixed model, this time predicting choice of color-matching target as a function of trial type (including baseline), age group, and their interaction. In this analysis, we saw that younger children had

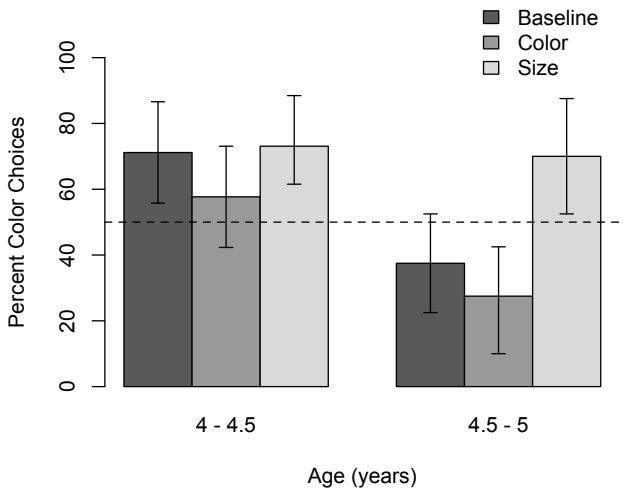

Figure 3: Mean percent color choices for all trials in the three trial types by the children in Experiment 4. Error bars show 95% confidence intervals.

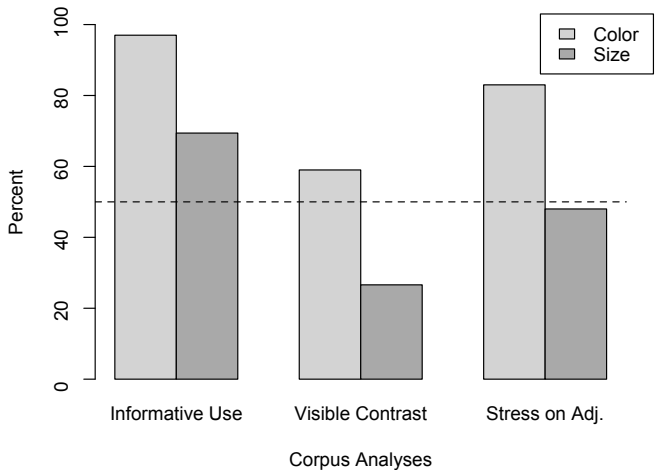

Figure 4: Percent of color and size corpus samples that were categorized by informative use, being in the context of a visible contrast, and exhibiting stress on the adjective.

a significant bias for color ($\beta = 1.32$, $p = .004$), and a trend towards differential responding in color trials ($\beta = -.83$, $p = .09$). There was a significant coefficient on older children's bias, indicating more size responding ($\beta = -2.02$, $p = .002$), as well as a significant interaction for size trials, indicating success in overcoming this baseline effect ($\beta = 1.76$, $p = .04$), but only for size trials. Thus, both groups showed some bias in their responding, but older 4s were better able to overcome that bias—at least for size—and make inferences about why a particular adjective was produced.

To summarize: Even within the narrow age range of our sample, there was a developmental difference between performance on color and size adjective trials. Although children at both ages were sometimes able to make contrast inferences, the ability to make these inferences clearly depended on the category of the adjective being used. Our next study looks to children's input to investigate one potential source for this developmental change.

Corpus Analysis

There are a number of possible reasons for developmental differences in color and size responding. One factor raised by Sandhofer and Smith (2001) was that distinctions in the ways color and size terms are used may promote production for color terms and comprehension for size terms. We conducted a corpus study to investigate how color and size terms are presented in natural speech to children, and whether there are differences in the types of contexts and contrasts in which these terms are typically used.

Materials Fourteen full speech files from the Providence corpus of CHILDES (Demuth, Culbertson, & Alter, 2006) were analyzed for adult-to-child speech containing color and size use. We used the last two or three samples for five children to most closely match the ages we used in Experiment 4 (age range 3;5.16 – 4;0.02, mean 3;8.06). A second coder blind to the hypotheses coded a random 40% of the samples. Inter-rater agreement was substantial (Cohen's $\kappa = 0.70$).

Procedures For each file, all uses of color and size adjectives throughout the written transcript were analyzed along with the video context. For each adjective use, we marked the category (color or size), adjective (e.g. red), position (prenominal, postnominal, or adjective used as noun), stress (whether or not the adjective was marked with emphatic stress), whether a visual contrast was present (e.g. whether a big and small item were both physically available during a reference to size), and whether the term was used in a contrastive or informative context (e.g. did the adjective convey a clear possible learning opportunity for the child).

Results and Discussion

We analyzed 330 speech instances of adjective use: 228 for size and 102 for color. The majority of the utterances used the adjective in the prenominal position (66% for color and 77% for size), but while 20% of size uses appeared in the postnominal, we only found 2 instances for color. The remaining third of color examples were used for naming or without a noun.

Emphatic stress was placed on color terms more often than size terms (83% for color, 48% for size). Color terms were also used more often with a visual contrast present than were size terms (59% color samples and 27% size samples). Results are plotted in Figure 4. It may be that adult speakers are selecting informative opportunities to use color labels with children at this age. Children may have a firmer grasp on size terms and infer contrast without a visible contrast set present.

For each of the samples, we examined whether or not the adjective was used in an informative context. Contexts were coded as informative if there was verbal reference to contrast, visual contrast, or marked salience to evoke an implied contrast with the prototype. Adjective use was considered uninformative when it did not provide qualifiable information to the child. In our samples, we found that 25% of size term use was uninformative, but this was almost never the case for color, suggesting that color and size terms are produced in different types of contexts. Color terms were nearly always used in instances that provided some type of information to the child (97%), whereas size was used informatively in 70% of instances. In addition, color was used to reference a visible contrast that was present in the immediate scene—as opposed to a reference to an unseen or implicit contrast item—44% of the time but only 16% for size.

It may be the case that children expect color terms to highlight the salience of a particular item in a visible contrast set instead of selecting an implied contrast set. In addition, perhaps the greater information contained in color utterances for children at this age may lead children to be biased to pair items on the basis of color, as in baseline performance in Experiment 4. Although more work is needed to understand the links between corpus distribution and behavior, the differences in use we observed might lead children to interpret color and size information in different ways.

General Discussion

Can adults and children learn from speakers' choice of a particular adjective? Our results suggest that adults are able to infer the general structure of a category based on the words chosen to describe a specific, anomalous example. Children also showed sensitivity to word choice, though we saw developmental differences between ages four and five. Children older than four-and-a-half sometimes succeeded in making inferences based on word choice, while younger children primarily exhibited a color bias. Our corpus analysis suggests that the language adults use with children around this age may mark color as implying a salient, immediate dimension, whereas size was used for a wider variety of functions.

The ability to make inferences from speakers' word choices may not only reflect more adult-like comprehension, but may also be an important learning mechanism for children. The earlier and faster children can go beyond what is stated at face value, the more opportunities they have to gain further knowledge through pragmatic cues. This kind of inference is consistent with work suggesting that pragmatic mechanisms can be used in a variety of different kinds of inferences: for inferring speaker meaning, learning words, and in this case, inferring facts about the world (Frank, Goodman, Lai, & Tenenbaum, 2009).

Children who are able to recognize that word choices can convey broader information will have greater opportunities for learning about the world because they can recognize both what is explicitly stated and implicitly implied. This ability allows children both to learn more from each utterance and to increase learning opportunities. Although pragmatic inferences are not always easy for children, our results suggest that these inferences may become an important source of background knowledge about the world.

Acknowledgments

Thanks to the staff and families at the Bing Nursery School.

References

Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition*, *118*(1), 84–93.

Clark, E. (1987). The principle of contrast: A constraint on language acquisition. *Mechanisms of language acquisition*, *1*, 33.

Clark, H. (1996). *Using language* (Vol. 4). Cambridge University Press Cambridge.

Demuth, K., Culbertson, J., & Alter, J. (2006). Word-minimality, epenthesis and coda licensing in the early acquisition of english. *Language and Speech*, *49*(2), 137–173.

Deutsch, W., & Pechmann, T. (1982). Social interaction and the development of definite descriptions. *Cognition*, *11*(2), 159–184.

Fernald, A., Thorpe, K., & Marchman, V. (2010). Blue car, red car: Developing efficiency in online interpretation of adjective-noun phrases. *Cognitive psychology*, *60*(3), 190–217.

Frank, M., Goodman, N., Lai, P., & Tenenbaum, J. (2009). Informative communication in word production and word learning. In *Proceedings of the 31st annual conference of the cognitive science society* (pp. 206–211).

Grice, H. (1975). Logic and conversation. *1975*, 41–58.

Grodner, D., & Sedivy, J. C. (2005). The effect of speaker-specific information on pragmatic inferences. *The processing and acquisition of reference. MIT Press: Cambridge, MA*.

Ito, K., & Speer, S. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language*, *58*(2), 541–573.

Nadig, A., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, *13*(4), 329.

Sandhofer, C., & Smith, L. (2001). Why children learn color and size words so differently: Evidence from adults' learning of artificial terms. *Journal of Experimental Psychology: General*, *130*(4), 600.

Sedivy, J. C. (2003). Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research*, *32*(1), 3–23.

Sedivy, J. C., Tanenhaus, M., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*(2), 109–147.