Bilingual Children’s Integration of Multiple Cues to Understand a Speaker’s Referential Intent

W. Quin Yow and Ellen M. Markman
Stanford University

1. Introduction

In most natural situations, the literal meaning of words are just one among many cues to a speaker’s communicative intent (Bloom, 1997; Clark, 1996; Sperber & Wilson, 1986). To accurately interpret a speaker’s utterance, adults often have to take into account multiple aspects of the communication, such as the linguistic context and semantics of what were being said, as well as nonlinguistic cues, e.g. tone of voice and facial expression (e.g. Kreuz, 1996; deGroot, Kaplan, Rosenblatt, Dews, & Winner, 1995; for a review, see Pexman, 2005).

Young children too can use simple linguistic and nonlinguistic cues to understand a speaker’s communicative intent (e.g. Diesendruck, Hall, & Graham, 2006; Behne, Carpenter, Tomasello, 2005; Baldwin, 1993; Fernald, 1993). For example, 3- to 4-year-old children are able to use linguistic cues, such as syntactic information to understand a speaker’s reference to a novel adjective (Diesendruck et al., 2006). Fourteen-month-olds can use an experimenter’s pointing and gaze direction to locate a hidden toy when both cues were directed to the same hiding location (Behne, Carpenter, & Tomasello, 2005). However, difficulty arises when young children have to attend to multiple cues that conflict with each other or when these cues have to be construed differently according to context (e.g. Ackerman, 1982; Milosky & Ford, 1997; Freire, Eskritt, & Lee, 2002; Nurmsoo & Bloom, 2008; Povinelli, Reaux, Bierschwale, Allain, & Simon, 1997; Jaswal & Hansen, 2006; Morton & Trehub, 2000). Hancock, Durham, and Purdy (2000) suggested that as children become more sophisticated in attending to multiple aspects of communication simultaneously, they adjust their responses based on singular message elements (e.g. words, face, voice) to integrating the different information and responding as communicative wholes. Children who regularly experience communicative challenges that demand greater attention and flexibility may develop a greater capacity to make use of multiple cues to infer the communicative intent of a speaker. One such population is children who grow up in a dual language environment. There is a growing body of evidence that bilingual children may be more sensitive to and better able to integrate sources of information as to a speaker’s communicative intent (e.g. Ben-Zeev, 1977; Comeau, Genesee, & Lapaquette, 2003; Cummins & Mulcahy, 1978; Genesee, Tucker, & Lambert, 1975; Siegal, Iozzi, & Surian, 2009; Yow & Markman, 2009a, 2009b). To take an example from pragmatics, Siegal and colleagues (2009) found that bilingual
children showed an enhanced ability to detect violations of Gricean conversational maxims.

In terms of tone of voice, Morton and Trehub (2000) asked children and adults to judge a speaker’s emotion while the speaker uttered sentences describing happy and sad situations in either a happy or sad voice. When the content and paralanguage matched (e.g. a happy sentence said in a happy voice), both children and adults could accurately identify happy and sad sentences. When the cues conflicted (e.g. a happy sentence said in a sad voice), adults overwhelming relied on how speakers sounded, while 4-year-olds almost exclusively judged speakers’ emotion from what they said. Yow and Markman (2009a) found that 4-year-old bilingual children were more tuned-in to the speaker’s affective intent and were better able to use the speaker’s tone of voice to judge the speaker’s emotion compared to monolingual children, especially when content of the utterance conflicted with tone of voice.

In terms of eye gaze and gesture, in Povinelli et al. (1997), an experimenter pointed to or looked at a baited box while seated either centered between two boxes (no conflict) or directly behind an empty box (conflict). They found that 2.5-year-old children were able to use either point or gaze direction to locate hidden rewards when there was no conflict between gesture and body position. However, when gesture was conflicted with body position, children were able to use the more straightforward gesture, the experimenter’s point, to locate hidden rewards, but failed when the subtler gesture, gaze direction, was provided. Yow and Markman (2009b) adapted Povinelli et al. (1997)’s procedure with 2- to 5-year-old monolingual and bilingual children. They found that young bilingual children made use of the gestures more successfully to locate hidden objects than the monolinguals, especially when the gestures conflicted with the experimenter’s body position cues.

A hiding game devised by Nurmsoo and Bloom (2008) required multiple sources of information to be integrated to infer a speaker’s referential intent. Two novel objects were placed in a box such that a child could see both objects while a speaker could only see one. The speaker fixed her gaze upon the visible object and said either “There’s the [novel-word]!” or “Where’s the [novel-word]?” In this situation, the speaker’s eye gaze provides information about which object she is looking at. The context is that she knows there are two objects, but she can see only one but not the other. Finally, the semantics of the utterance “there” conveys information about reference to a specific target object, while the semantics of “where” conveys a questioning intent about the location of an object. Integrating all the cues means we should expect the speaker to refer to the mutually visible object when she looked at it and said “There” but that the speaker was actually looking for the object she could not see when she asked “Where”. Nurmsoo and Bloom (2008) found that this task was challenging for young children: 2.5-year-olds were more likely to pick the mutually visible object when the speaker said “There” but were not more likely to pick the nonvisible object when the speaker said “Where”. Yow and Markman (2009b) replicated Nurmsoo and Bloom (2008)’s study with 3-year-
old monolingual and bilingual children and found that the bilingual children were better able to differentiate the “there” from the “where” questions.

The focus of the current work is to ask whether this bilingual advantage would persist even among four-year-olds who succeeded in differentiating “there” from “where” questions in Nurmsoo and Bloom (2008), when tone of voice, yet another source of information about communicative intent, needs to be taken into account. Tone of voice can provide information such as whether the speaker is serious or merely playing pretend with them (Reissland & Snow, 1996). A related pragmatic difference signaled by tone of voice is to distinguish serious questions about where an object is located when the speaker is trying to find something versus the more pedagogical playful tone of voice commonly used in picture book reading where an adult might ask a child “where” something is but the object is mutually visible and the goal is for the child to display their knowledge.

In this study, we made use of Nurmsoo and Bloom (2008)’s procedure to examine whether 4-year-old children can still succeed when tone of voice needs to be integrated with the speaker’s perceptual access, nonlinguistic context, and semantics of the question. In both conditions speakers asked “where” an object was but either in a serious tone of voice which indicates the speaker is sincerely wondering where the object is or in a playful/pedagogical tone of voice that implies the speaker knows where the object is but wants the child to reveal that knowledge. Integrating all these cues means we should expect that the speaker is looking for the one that she cannot see when she asks “where” in a serious, genuine manner, as in Nurmsoo and Bloom’s study, but that the speaker is referring to the object mutually visible when she asks “where” in a playful, pedagogical manner. We predicted that 4-year-old bilinguals would be better than monolingual children at differentiating these two questions. That is, monolingual and bilingual children alike should succeed at the serious “where” questions as in the previous work but bilinguals should outperform monolinguals when the interpretation of the question needs to be modulated by tone of voice.

2. Experiment
2.1. Method
2.1.1. Participants

Fifty-eight 4-year-old English monolingual and bilingual children participated in this study. All children were recruited from the same university lab school, lived in its neighboring areas, and were mostly middle-to-upper class. Twenty-nine were monolinguals (15 males; mean age = 4.40; range = 4.02-4.91) and 29 were bilinguals (15 males; mean age = 4.49; range = 4.07-4.96). A language questionnaire was sent to the parents via the school that asked for information about the language first acquired by the child, the language used by the parents and caregivers, and the amount of time (average percentage of exposure per week) the child was exposed to each language.
Children were determined to be bilingual if they had at least 30% exposure to one of two languages weekly. The 29 bilingual children in the study were reported to have regular exposure to another language besides English, such as Spanish (n=11), French (n=4), Mandarin (n=4), Korean (n=3), Portuguese, Hebrew, Japanese, Swedish, Russian, Italian, German, (n=1 per language), mainly either from parents or a nanny.

25 adults who received introductory psychology course credit for their participation (12 females; 13 males) were also recruited to validate the stimuli used in the study.

2.1.2. Materials

The materials consisted of an opaque cardboard box (32 cm x 22 cm) and a bag of toys. The box had two compartments, each with a window located in the center of the compartment. A movable screen covered one of the windows. The box was placed between the child and the experimenter so that the child could see into both compartments but the experimenter could only see into one through the uncovered window. There were two familiar toys (a teddy bear and a toy car) and eight novel objects (uncommon objects or parts of a bigger object). The four pairs of novel objects were used with four novel labels, *spoodle*, *nurmy*, *flurg*, and *gorp* as per Nurmsoo and Bloom (2008). The target object, its location relative to the window and the left/right position of the screen were all counterbalanced.

Children were shown video clips of the study procedure. In the familiarization phase, the two familiar toys were placed in each of the compartments of the box. Children were first asked to identify which toy they thought the experimenter could and could not see. The box was then rotated so that children now had the experimenter’s perspective. They were then asked to identify which toy they could and could not see. In the experimental phase, a puppet placed pairs of novel objects in the box while an actor was away. When the actor returned, she fixed her gaze on the object that she could see and asked the test question “Oh! Where’s the [novel word]? Where is it?” either in a serious manner (“serious-where” condition) or in a playful manner (“playful-where“ condition). In the “serious-where” condition, the question was asked in a conventional, serious, questioning tone-of-voice. In the “playful-where” condition, the question was asked in a playful, pedagogical manner, similar to how adults behave when reading picture books to children. In both conditions, the actor then looked up, reached out her hand, and asked, “Can I have the [novel word]?” There were two pre-determined orders for each condition, counterbalanced for the target object’s location relative to the window and position of the window.

In addition, audio clips of all the test questions were abstracted from the video files using video-audio converter software. A total of 16 audio clips were obtained.
2.1.3. Procedure

The adults were asked to listen to each of the audio clips in a pre-determined random order and rate whether the speaker sounded serious or playful and also how confident they were in each of their rating (1 = not confident at all; 5 = very confident). This was to ascertain that the “serious-where” and “playful-where” video clips were indeed distinguishable in their prosodic qualities.

Children were tested individually in a quiet room in their preschool. The videos were presented to each child on a LCD monitor using a Macintosh laptop. Children first saw the familiarization video and asked to identify which toy the actor could see and which she could not see. Next, the experimenter showed each of the four videos in the experimental phase and asked the child which one of the two objects the actor was looking for. The child’s response was recorded. This procedure was repeated until all four experimental trials have been completed.

2.1.4. Other measures

2.1.4.1. Socio-economic status (SES)

To verify that the monolingual and bilingual children were drawn from the same socioeconomic status (SES) population, we followed the procedure reported by Westenberg, Siebelink, Warmenhoven & Treffers (1999), Furth et al. (2000), Buck, Small, Schisterman, Lyon & Rogers (2000), Rathore et al. (2006), and Ward (2008) and used the participants’ residential addresses to obtain an estimated value of each family’s dwelling from an internet website that provides real estate information such as home prices and home values (www.zillow.com). Using this method, we then calculated the median, mean, and variance property valuation for the monolingual and bilingual children in order to determine whether the two groups of children differ in SES.

2.1.4.2. Peabody Picture Vocabulary Test IV (Dunn & Dunn, 2007)

This is a test of receptive vocabulary where each child was to select one picture from a set of four that depicts the word that was being spoken by the experimenter. The test continued until the child made eight or more errors in any set of 12 items. Raw scores were converted to standard scores using normalized tables based on age.

2.1.4.3. Digit-Span task (Wechsler, 1997)

This task was adapted from the Wechsler Intelligence Scale for Children-Revised as a test of short-term memory. A list of pre-determined random numbers ranging from two to nine digits was read out loud. Each child was to repeat all the numbers verbally in the same order. There were two trials for each
digit length. The test began with two numbers, increasing until the child committed errors on both trials of the same digit length. The child’s digit span score was the total number of trials completed correctly.

2.1.4.4. Day-Night task

This task was adapted from the day-night task used in Gerstadt, Hong, and Diamond (1994). It involves instructing children to say the word “day” when they see a card depicting a nighttime sky and to say “night” when shown a picture of the daytime sky. This task requires remembering the two rules and inhibiting a response to the visual cues. There were two training cards and 16 testing cards used in this study. Half of the cards showed a yellow sun in a light blue background and half showed a white crescent moon and stars on a black background. The instructions and presentation of cards were adapted from Siegal, Iozzi, and Surian (2009). The experimenter first showed each child a card with the moon and said, “We are going to play a funny game. When you see this card I want you to say day. Can you say day?” The experimenter continued to show a card with the sun and said, “Now, when you see this card I want you to say night. Can you say night?” The child was then shown the first test card with the sun and asked, “Now, what do you say when you see this card?” The child was shown a card with the moon next and asked, “What do you say when you see this card?” If the child got either of the first two test trials wrong, these two trials were counted as practice trials. The child would then be told of the rules again and the test trials would start all over again. If the child responded correctly to the first two trials, these were counted as trials 1 and 2 and the child proceeded with the remaining trials. The total number of correct responses was scored on a 2-16 scale.

2.2. Results
2.2.1. Preliminary analyses
2.2.1.1. Validation of tone of voice stimuli

Adults were given a score of from 0 to 16 that reflects the number of times they correctly rated the speaker as sounded serious or playful. They were also given an average score of from 0 to 5 that reflects their average confidence level in their ratings. The average total number of items correct (out of 16) = 14.68, $SD = 1.18$, and the average confidence score (out of 5) = 4.10, $SD = .46$, both of which were significantly above chance performance (all $ps < .01$). Adults were able to distinguish between the “serious-where” and “playful-where” questions with relatively high confidence.

2.2.1.2. Measures of SES

In order to determine whether the monolingual and bilingual children came from similar SES background, statistical analyses were conducted on the ratios
of the mean, median, and variance property valuation between monolingual and bilingual children. The ratio of the median property valuation between monolingual and bilingual children was 1: 1.09 and Mann-Whitney U-test confirmed that these two groups of children came from the same SES backgrounds, \( Z = .023, P > .10 \). The ratio of the means was 1: 0.85, and \( t \)-tests showed no significant differences between these two groups of children based on the estimated property valuations, \( t(45) = -1.334, p > .10 \). The ratio of the variances was 1:0.96 and the Levene test of equality in variances confirmed that the two group variances of estimated property valuations did not differ significantly from each other, \( F(1,45) = .01, p > .10 \).

2.2.1.3. Measures of vocabulary, memory span, and inhibitory control

The mean scores and standard deviations for the PPVT, Digit-Span, and Day-Night Stroop tasks are shown in Table 1. A one-way ANOVA was conducted to compare children from the two language groups in these three tasks. No significant effects were found for PPVT and Day-Night task, all \( p > .10 \), but there was a marginal significant effect of language status in the Digit-Span task, \( F(1,52) = 2.86, p = .097 \). Bilingual children, on average, scored marginally higher than the monolingual children in the Digit-Span task (however, note below that there was no significant correlation in performance between the Digit-Span and the experimental task).

Table 1. Mean Scores and Standard Deviations (in parentheses) of Children on Peabody Picture Vocabulary Test (PPVT), Digit-Span Task (DS), and Day-Night Task (DN) in Study 1, and 2.

<table>
<thead>
<tr>
<th>Mean Age</th>
<th>Language Status</th>
<th>PPVT</th>
<th>DS</th>
<th>DN</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.40</td>
<td>Monolingual</td>
<td>120.93 (11.77)</td>
<td>6.29 (1.33)</td>
<td>11.86 (3.58)</td>
</tr>
<tr>
<td>4.49</td>
<td>Bilingual</td>
<td>117.40 (15.22)</td>
<td>7.12 (2.03)</td>
<td>11.38 (3.18)</td>
</tr>
</tbody>
</table>

For the experimental trials, children were given a score of from 0 to 4 that reflects the number of times they successfully selected the mutually visible object. There were no significant correlations between scores in the experimental trials and SES, PPVT, Digit-Span task, Day-Night Stroop task (all \( p > .10 \)).

Thus, the monolingual and bilingual children were drawn from identical SES populations and were very comparable in terms of standard measures of vocabulary, inhibitory control, with the only difference a marginal advantage in digit span. Moreover, none of these measures were correlated with success on the experimental measures of interest.
2.2.2. Main results

A univariate ANOVA was conducted with language status (monolingual vs. bilingual) and condition (serious vs. playful) as the fixed factors. As predicted, there was a significant interaction effect between language status and condition, \( F(1,54) = 4.20, p < .05 \) (Figure 1). Planned comparison \( t \)-tests revealed that bilingual children’s performance was significantly better than that of monolingual children in the “playful-where” condition but children were equally successful in the “serious-where” condition \( t(1,26) = 2.16, p < .05; t(1,26) = .80, p > .10, \) respectively. As predicted, monolingual and bilingual children were equally likely to select the hidden object when asked “where” in a serious, genuine manner, but bilingual children were more likely than monolingual children to select the mutually visible object when asked “where” in a playful, pedagogical manner. No other significant effects were found.

We also compared performance against chance. Monolingual children were at chance picking the visible object both in the “playful” and “serious” conditions \( t(13) = -1.53, p > .10 \) and \( t(14) = -.75, p > .10 \) respectively). While bilingual children were at chance in picking the visible object in the “playful” trials, they were marginally below chance in picking the visible object in the “serious” trials \( t(13) = 1.53, p > .10 \) and \( t(14) = -1.92, p = .076, \) respectively.

**Figure 1.** Average total number of times children picked the visible object (out of 4) (+SE) by condition for monolingual children (n = 29) and bilingual children (n = 29).
3. General discussion

Our results provided evidence that children growing up bilingual may be more adept than monolingual children at integrating multiple cues when interpreting a speaker’s referential intent. Our modified version of Nurmsoo and Bloom (2008) required children to integrate a speaker’s eye gaze, the nonlinguistic context, the semantics of a “where” question, and tone of voice to correctly infer the speaker’s communicative intent. Children saw two novel objects placed in a box such that a speaker could see only one of them while the children could see both (context). The speaker looked at the object she could see (eye gaze) and said “Where’s the [novel-word]?” (semantics), either in a serious, genuine manner or a playful, pedagogical manner (tone of voice). Integrating all these cues thus implied that the speaker was sincerely looking for the object that she could not see when she asked in a serious, genuine manner but that she wanted the child to pick the object she was looking at when she asked in a playful, pedagogical manner. We found that while monolingual and bilingual four-year-olds were equally likely to select the hidden object when asked “where” in a serious, genuine manner, bilingual children were better able than monolingual children to select the mutually visible object when asked “where” in a playful, pedagogical manner.

Past research has shown that young children can interpret simple and straightforward cues to understand a speaker’s communicative intent but it is often not until years later that they are able to integrate multiple communicative cues successfully. Regular exposure to and experience with challenging communicative situations may promote the development of such social-cognitive skills. Bilingual children regularly experience challenging communicative situations to a greater extent than those monolingual children have to cope with. There is the complexity of learning two language systems simultaneously. All children face the problem of referential indeterminacy when acquiring language, but the demands of multilingual communications compound the problem further. Bilingual children may achieve communicative effectiveness amidst these challenges by frequently monitoring the context and utilizing verbal and nonverbal cues available in the situation to better understand the speaker’s communicative intent. Children growing up in a bilingual environment may become better able to make use of various communicative cues to infer a speaker’s intent due to their heightened awareness of the social, pragmatic and communicative contexts surrounding language use.

References


