The Comprehension of "Left" and "Right" in a Referential Communication Task

Stanka A. Fitneva (Fitneva@Queensu.ca)
Department of Psychology, Queen's University
Kingston, ON K7L 3N6, Canada

Yi Song (7ys4@Queensu.ca)
Department of Psychology, Queen's University
Kingston, ON K7L 3N6, Canada

Abstract

The study examined the role of establishing local conventions in the interpretation of *left* and *right*. 32 adults participated in a referential communication game carrying commands like "Put the red cup to the left of the green." The results suggest that local conventions strongly influence the interpretation of the two spatial terms but global expectations about the usage of the terms also play a role.

Keywords: perspective taking; spatial cognition; eye tracking.

Introduction

Computing the meaning of spatial relation terms, especially of directional terms like *left* and *right*, is difficult and time consuming (e. g., Bryant & Tversky, 1991; Farrell, 1979; Franklin & Tversky, 1990; Maki & Braine, 1985; Maki, Grandy, & Hauge, 1979). *Left* and *right* are acquired late in development， after other conceptually similar terms like *front/back* and *above/below* (Fisher & Kamenzuli, 1987; Internicola & Weist, 2003). They take longer to produce and understand (Franklin & Tversky, 1990) and when other options are available, adults appear to avoid using *left* and *right* (Mainwaring, Tversky, Ogishi, & Schiano, 2003). The explanation of the left/right confusion has been historically attributed to the bilateral symmetry of the nervous system (i.e., the existence of projections into both hemispheres) and the conceptual complexity of the referential frames evoked by the terms (Corballis & Beale, 1976; Maki, Grandy & Hauge, 1979). Another explanation is the intrinsic ambiguity of *left* and *right* (Mainwaring et al., 2003; Schober, 1993). In face-to-face conversations, utterances like "Put the red cup to the left of the green" require establishing whose perspective should be considered: the speaker's or the addressee's?

Language users rely on a set of conventions in interpreting the meaning of linguistic expressions (Lewis, 1979/1991). There are global conventions, which are shared by the broader linguistic community and represent general patterns of word use. These conventions are reflected in dictionaries. There are also local conventions which develop in the course of communication within a subgroup of the community (e.g., family members) or with a single other individual (e.g., Brennan & Clark, 1996; Garrod & Anderson, 1987). As an example of the distinction, consider the word "mother." Its dictionary definition is "a female parent." Spoken amongst family members, however, "mother" identifies a specific individual and thus not just a female parent but also possibly an accomplished musician and a great cook. In the present study, we examined the extent to which adults' processing of *left* and *right* is influenced by local conventions established in conversation.

The influence of local conventions in conversation is illustrated by Brennan and Clark (1996). In their study, when participants saw a dress shoe together a high-heel shoe and a sneaker, they labeled it *a loafer* rather than *a shoe*, which is a more common label. This was expected because participants had to distinguish the dress shoe from the other shoes. However, when afterwards participants saw the dress shoe in the context of unrelated objects, they still referred to it as *a loafer*. Here, *a loafer* is overinformative and *a shoe* would suffice. People's tendency to follow and rely on the recently established conventions in conversation provides a compelling account of these findings.

Research has consistently shown that speakers tend to assume the listener's perspective in providing spatial descriptions (Mainwaring et al., 2003; Schober, 1993). For example, Schober (1993) presented participants with a modified version of the referential communication task (Glucksberg, Krauss, & Weisberg, 1966). They saw scenes depicting two identical objects (in addition to indicating the place from which the addressee is looking at the scene) and the participants' task was to guide the addressee to picking one of them. There was a limited number of ways to describe the scenes. The participants could either adopt their own perspective, saying something like "it's the one on my left," or the listener's perspective, saying something like "it's the one on your right." Schober (1993) found that most of the time, speakers took the addressee's perspective. This finding was replicated by Mainwaring et al. (2003) who also showed that this usage is modulated by the speaker's assumptions about their own and the addressee's cognitive load. Schober (1993) and Mainwaring et al. (2003) suggested that speakers take the addressees' perspective because it is easier to translate a visual scene into language (the speaker's task) than language to a visual representation (the addressee's task). Regardless of the explanation, the global convention for the usage and interpretation of spatial terms of reference in the referential communication task appears to involve the addressee's perspective.

Despite the prevalence of using *left* and *right* from the addressee's perspective, speakers mix spatial perspectives

about half of the time (Schober, 1993; Taylor & Tversky, 1996). Thus, while addressees may rely on global information about the use of *left* and *right*, locally established precedents could be particularly useful in their interpretation.

In the present study, we used the referential communication task to examine addressees' interpretation of *left* and *right* over time. On critical trials, the participants, who sat across a confederate, had to carry out commands like "Put the red cup to the left of the green." In the first third of the study – the Establish period – the physical display revealed whether the speaker was taking the listener's or her own perspective in using *left* and *right* (see Figure 1). In the second third of the study – the Test period – we examined whether addressees relied on the previously established local convention of interpreting *left* and *right* in the absence of physical supports. In the last third of the study – the Break period – we examined whether switching perspectives incurred cognitive costs for the addressees. There were two conditions in the study – Speaker and Listener – named after the perspective used in the Establish period. We refer to the point-of-reference cup (the green one in the above example) as the anchor cup.

The first question we asked is what are the temporal dynamics of establishing a local convention for the meaning of *left* and *right*. To examine this question, we compared participants' performance over the six trials in the Establish period. If addressees draw on the tendency of speakers to adopt the listener's perspective, then performance in the Listener condition would be faster. Performance in the Speaker condition would be slower because it requires deviating from the global convention. If addressees rely exclusively on local conventions in interpreting *left* and *right*, then any problems in the beginning would be due to the lack of precedent, which is common to both conditions.

The Test period tested whether the interpretation of *left* and *right* will be maintained in the absence of physical/perceptual constraints. This period contained ambiguous trials, in which the anchor cup was flanked by an empty slot on each side. We expected that the placement of the cup will show consistency with the previously established perspective. Our analyses focused on the time it took to carry out the commands, and the proportion of time participants spent looking in the region suggesting that they interpreted the command the same way as in the establish trials. We expected that both measures would show that the addressees expected that the speakers would maintain the same perspective in issuing the commands. Of critical interest was whether there was a difference in the degree to which this was true in both conditions. If local conventions have only transient effect, then the perseverance in perspective-taking might be greater in the Listener condition than in the Speaker condition. This is because the test trials allow the participants in the Speaker condition to exercise their default, listener-centered interpretation of the command, grounded in the global convention of *left/right* usage. If so, participants in that condition could spend less time looking at the consistent location and have longer response times.

If the interpretation of the spatial terms is entirely determined by the local conventions developed during the Establish period, then we would not expect any differences between conditions.

We also examined how participants would respond to a change in the perspective from which the speaker uses *left* and *right*. After the test trials, there were six more 'Break' trials. For the participants who received speaker trials in the Establish period, these were listener trials, and for the participants who received listener trials in the Establish period, these were speaker trials. If addressees rely on local conventions, the costs of switching from an established precedent to a new perspective should be similar in both conditions. However, if they rely on the global convention, there would be higher costs in switching from interpreting a term from one's own perspective to interpreting it from another's perspective than in switching from interpreting a term from another's perspective to interpreting it from one's own perspective.

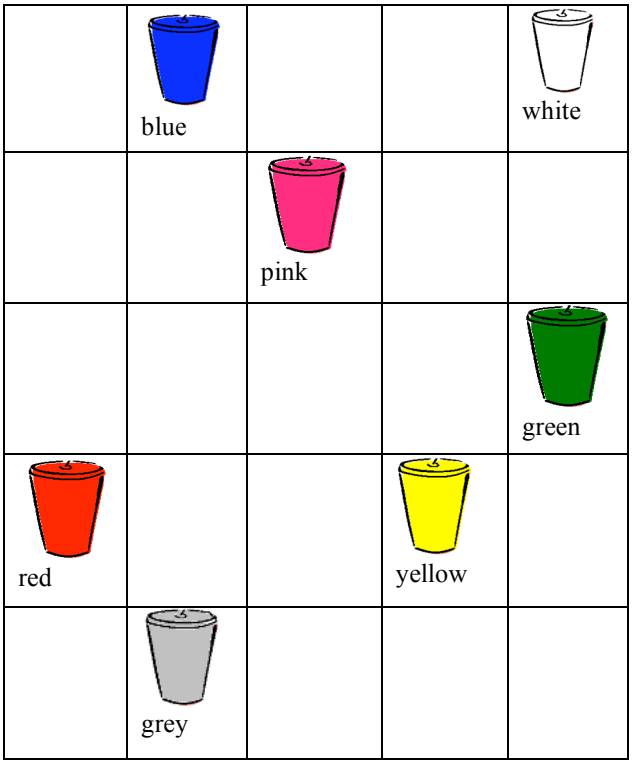

Figure 1. Illustration of the displays in speaker and listener trials. "Put the red cup to the *right* of the green" suggest a speaker-centered use of the spatial terms while "Put the red cup to the *left* of the green" a listener-centered use. When the display is rotated vertically 180°, the same commands demand a listener–centered and a speaker-centered interpretation respectively.

Method

Participants

There were 16 undergraduates, all monolingual speakers of Canadian English, in each of the two condition. Additional six participants were recruited but not included in the study because of equipment failure (failure to calibrate or capture video). All received course credit for their participation.

Apparatus

A 5 x 5 plywood structure was used to position seven different color cups (blue, green, pink, red, gray, white, and yellow). Each slot was approximately 21 cm by 25 cm. Eye movements were recorded using Applied Science Laboratories head-mounted eye-tracker with 60Hz sampling rate. A magnetic head-tracker allowed for the automatic extraction of the spatial coordinates of the point of gaze. We recoded the video from the scene image for later coding of the nonverbal behavior of the confederate. An additional camera provided a side view of the participant and the display and enabled coding the arm movement and the final placement of the cup.

Design and Procedure

Participants were randomly assigned to condition. The participant met the confederate in a waiting room and the experimenter ushered both into the lab. Both were given a letter of information about the study that described it as examining how people give and follow commands. The letter also outlined the roles of the speaker and the addressee. The pair was instructed to work as efficiently as possible. After signing the consent form, the confederate and the participant participated in a draw, which was fixed so that the participant always took the role of an Addressee. The participant was then taken to the testing room and the eyetracking equipment calibrated. The confederate sat across, wearing reflective sunglasses.

No specific instruction were given about communication during the study. To clarification questions, the confederate responded by nodding (reaffirming the guess of the participant). In very few cases, where no guess was apparent, the confederate responded by pointing because this most clearly disambiguated the command.

Each trial began with the experimenter positioning the seven cups in the display. Then a card was handed to the confederate. The card supposedly contained a picture of how the display should end up looking like. In reality, it contained a series of commands that the confederate had to give. The critical instruction, containing a 'left' or a 'right' term was always second. The other three commands contained instructions like "Move the red cup one slot down."

There were three types of trials. On listener and speaker trials, the anchor cup was in an end column so it was flanked by just one empty slot. On these trials the physical display determined the interpretation of the spatial term and suggested either a listerner-centered interpretation or a speaker-centered interpretation. On ambiguous trials, the anchor cup was flanked by an empty slot on each side.

There were 18 trials. The first six trials differed between conditions. For the participants in the Listener and Speaker conditions, they were respectively listener and speaker trials. The six 'test' trials were the same for both groups and were always ambiguous trials. The last six 'break' trials were either listener or speaker trials. They required a change in perspective in the interpretation of the *left* and *right* terms.

We constructed six displays and another six were added to the set by rotating the original ones around the vertical axis. Thus, a cup that was in the left half of a display would be in the right half of the rotated one. The participants saw one version of each display in the Establish part and one in the Break part. Two different random orders were used for presenting the trials in the Establish and Break parts and order was counterbalanced across subject.

The displays used in the test trials were different from those in the other parts of the study. An example of a command if the display in Figure 1 were used would be "Put the white cup to the left of the pink." There were three commands with *left* and three with *right* in each of the three parts of the experiment.

Results

A preliminary analysis targeted establishing that our dependent variables, especially the eye movement measures are driven by interpretation of the verbal command rather than the nonverbal behavior of the confederate. Two coders blind to the hypotheses watched the scene camera recordings of the ambiguous trials when the commands were given. They did not hear the commands but knew which cup was the anchor cup. They were asked to guess which flanker slot the confederate wants to get filled. The correlation between the two coders' guesses was not significant, suggesting that there were no reliable nonverbal cues that could influence participants' interpretation of the spatial terms.

Behavioral Data

Cup Placement In the Establish and Break periods there was only one slot where the cup being moved could be placed. Thus, naturally, cup placements were always consistent with an interpretation of the commands from the speaker's perspective on Speaker trials and an interpretation of the commands from the listener's perspective on Listener trials. In the Test period, cup placements were consistent with the precedent from the Establish period 100% of the time. After Listener trials, the participants showed listener-centered interpretation of the spatial terms and after Speaker trials, they showed speaker-centered interpretation of the spatial terms.

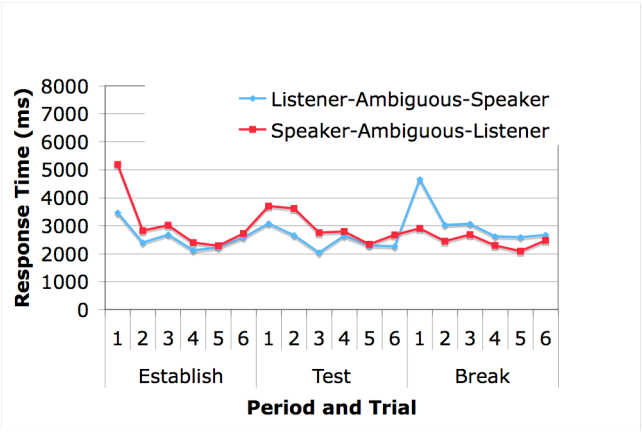

Figure 2. Response times in each condition by period and trial

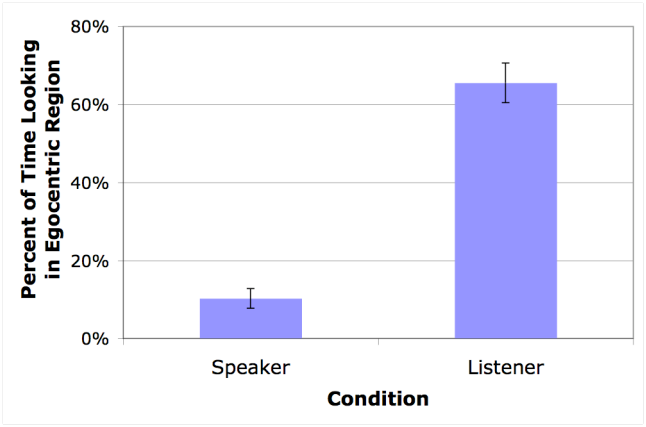

Figure 3. Percent of time spent looking in the region showing egocentric interpretation of *left* and *right,* in test trials.

Response Time Figure 2 shows the average response times in the three periods of the study by condition and trial. Response time was defined as the period from the offset of the last word in the command, which identified the color of the anchor cup, to the moment the cup being moved was placed. These time points were identified from the video from the side camera. The average response times in each period were submitted to a 2 x 3 repeated measures ANOVA with condition as a between-subject variable and period as a within-subject variable. One participant was dropped from this analysis because of missing data. The analysis revealed a significant interaction between period and condition, $F(2, 58) = 10.55$, $p < .001$. As Figure 2 shows, this interaction was due to the fact that the response times of the participants in the speaker condition were longer than the response times in the listener condition in the first two periods (Establish and Test) but shorter in the Break period. Post-hoc tests showed that only the difference between conditions in the Establish period was significant.

Looking Data

The analyses of the looking data focused on the ambiguous test trials because there were very few fixations outside of the displays during listener and speaker trials. We examined two looking measures: the location of first fixation following the end of the command and looking time. For both measures, fixations were coded as listener-centered or speaker-centered if they were in the columns immediately adjacent to the column in which the anchor cup was.

First Fixations The location of first fixations was analyzed using generalized estimating equations, which is a technique designed to handle repeated measures categorical data. The proportion of listener-centered first fixations was 29% and 20% in the Listener and Speaker conditions respectively. The difference between conditions was not significant $\chi^2(1) = .84$, $p > .05$. The lack of difference suggests that the precedents in the Establish trials did not significantly influence first fixations in the Test period.

Looking Time The second measure we examined was the proportion of time participants spent looking at the region suggesting listener-centered interpretation of the spatial term relative to the total number of fixations suggesting listener-centered and speaker-centered interpretation. Only fixations in the period from the end of the command to the placement of the cup being moved were considered. Figure 3 shows the average proportion of time participants spent looking at the slots consistent with an egocentric interpretation of the commands. The data were submitted to a repeated measures ANOVA with condition as a between-subject variable. Participants in the Listener and Speaker conditions spent 65% and 10% respectively looking at the slots showing an egocentric interpretation of the commands. The difference was statistically significant, $F(1, 29) = 101.46$, $p < .001$.

Discussion

The goal of this study was to examine the role of local conventions in the interpretation of the spatial terms *left* and *right*. The placement of the cups during the Test period clearly showed that addressees expected that the speaker would continue to use the two spatial terms in the same way as in the Establish period. Moreover, the analyses of the looking time data showed that addressees were more likely to look at the region consistent with a listener interpretation of the commands if they had received Listener trials in the Establish period than if they had received Speaker trials. These findings provide strong evidence for the role of local conventions in the interpretation of *left* and *right*.

The experiment also provided evidence for addressees' expectations based on the global use of the terms. Specifically, the response time data showed a significant difference between conditions in the time it took participants to execute the commands in the Establish period. Participants in the Listener condition were faster as predicted from the research showing that speakers usually take the listener's perspective in giving commands (Mainwaring et al., 2003; Schober, 1993). There was also no

difference between conditions in the proportion of first fixations in the regions consistent with a listener interpretation of the commands in the Test period. Although this finding contradicts the looking time data in suggesting that the Establish period had no effect on the interpretation of the spatial terms in the Test period, it should be interpreted with caution. First fixation were often not in the target regions and thus, it is based on only one quarter of all trials. Furthermore, while first fixations are sensitive to the processes of referential interpretation in reading, they may reflect visuo-spatial information in speech processing, e.g., the distribution of cups around the anchor cup, and thus provide a noisier measure of on-line speech comprehension.

In conclusion, the present study demonstrates the importance of local conventions in the processing of spatial terms. The picture that the results reveal, however, is one of interaction between global, stored representations and local, dynamically evolving ones.

Acknowledgments

We thank Bronwen Fullerton, Courtney Fitzgerald, Scott Ross, Myriam Leduc-Robert, and Ashley Frambach for help with data collection and coding.

References

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(6), 1482-1493.

Bryant, D., & Tversky, B. (1992). Assessing spatial frameworks with object and direction probes. *Bulletin of the Psychonomic Society*, 30(1), 29-32.

Bryant, D., Tversky, B., & Franklin, N. (1992). Internal and external spatial frameworks for representing described scenes. *Journal of Memory and Language*, 31(1), 74-98.

Farrell, W. (1979). Coding left and right. *Journal of Experimental Psychology: Human Perception and Performance*, 5(1), 42-51.

Fisher, C., & Camenzuli, C. (1987). Influence of body rotation on children's left-right confusion: A challenge to bilateral symmetry theory. *Developmental Psychology*, *23*(2), 187-189.

Franklin, N. (1995). Parsing surrounding space into regions. *Memory & Cognition*, 23, 397-407.

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition, 27*, 181–218.

Glucksberg, S., Krauss, R. M., & Weisberg, R. (1966). Referential communication in nursery school children: Method and some preliminary findings. *Journal of Experimental Child Psychology*.

Internicola, R., & Weist, R. M. (2003). The acquisition of simple and complex spatial locatives in English: A longitudinal investigation *First Language, 23*, 239 -248.

Lewis, D. (1979/1991). Scorekeeping in a language game. In S. Davis (Ed.), *Pragmatics: A Reader* (pp. 416-427). New York: Oxford University Press.

Mainwaring, S., Tversky, B., Ohgishi, M., & Schiano, D. (2003). Descriptions of simple spatial scenes in English and Japanese. *Spatial Cognition and Computation*, 3(1), 3-42.

Maki, R., & Braine, L. (1985). The role of verbal labels in the judgment of orientation and location. *Perception*, 14(1), 67-80.

Maki, R., Grandy, C., & Hauge, G. (1979). Why is telling right from left more difficult than telling above from below? *Journal of Experimental Psychology: Human Perception and Performance, 5*(1), 52-67.

Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition, 47*(1), 1-24.