Situating Spatial Templates for Human-Robot Interaction

John Kelleher and Robert Ross and Brian Mac Namee and Colm Sloan
Artificial Intelligence Group, Dublin Institute of Technology, Ireland.

Participants, Stimuli & Procedure

Participants for the study were recruited online and compensated. 42 participants were native English speakers and their data was retained for analysis. Participants were asked to rate their agreement with a series of paired linguistic and visual stimuli. Linguistic stimuli situated a trajector with respect to a landmark. Each linguistic stimulus was of the form ‘The A box is REL of the B box’, where A and B were substituted by color words (explained below) and REL was one of three directional spatial terms, i.e., ‘in front’, ‘to the right’, and ‘to the left’. Visual stimuli were 2.5 dimensional images of a scene consisting of a rectangular landmark and cylindrical trajector. The landmark object was 8 units wide by 6 units deep by 2 units high, while cylinders were one unit in diameter and one unit high. The landmark was situated obliquely to the participant’s viewing angle. While the landmark position and viewing angle were fixed, trajector position could be moved to one of ten locations. The landmark, trajector positions, and viewer angle are depicted to scale in Figure 1. Note that in order to reduce repetition effects, each scene configuration was produced in accordance with two different coloring schemes (trajector:yellow landmark:red and trajector:red landmark:blue).

After being given written instructions describing the procedure, but not priming for any discourse or spatial phenomena, participants were presented with a randomly ordered set of visual and linguistic stimuli pairings. For the ‘in front of’ linguistic stimulus, the visual stimulus could be drawn from

Copyright © 2010, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.
any possible trajector position. For the ‘to the right of’ and ‘to the left of’ linguistic stimulus, the visual stimulus was limited to location sets 4, 5, 6 and 1, 2, 3 respectively. For each stimulus pairing, participants were required to indicate their level of agreement through a 7 point Likert scale. Each participant rated each applicable visual stimulus for each linguistic stimulus, and color schemes for each pairing were selected randomly from the available pairings.

Results & Analysis

Table 1 presents the mean acceptance ratings and standard deviations for each stimulus pairing. As a control test of participant understanding, we expected, and found, high ratings for trajector positions 2 and 3 for stimuli including ‘to the left’, and high ratings positions 5 and 6 for stimuli including ‘to the right’. Also expected and found, was that participants rated the acceptability of ‘in front of’ higher for positions oriented with the long face of the landmark higher than the short face (compare set 1 and 4). We attribute this to the effect of the landmark’s intrinsic frame of reference.

In comparison with (Logan and Sadler 1996; Carlson-Radvansky and Logan 1997; Kelleher and Costello 2005), however, there are some anomalies in our data. Most striking is the fact that the acceptability ratings do not systematically drop with angular deviation from any plausible direction vectors, nor with distance from the landmark. For example, 4A, 4B, 4C have the same angular deviation from both the intrinsic and the viewer-centered direction vectors. However, their acceptance ratings vary and this variation does not correlate with distance from the landmark. The same holds for 1A, 1B, 1C. We posit that this is caused by the participants’ oblique perspective on the landmark causing a distortion in the spatial templates for the different frames of reference.

Conclusions & Future Work

In Human-Robot dialogue systems spatial template models bridge the grounding gap between spatial language and sensor data. To date, however, the effect of interlocutor perspective on spatial templates has not been systematically examined. To this end, we conducted an experiment where the landmark object was presented at an oblique angle to the participants. We interpret our results to indicate that interlocutor perspective on a landmark may distort directional spatial templates anchored on the object. While this is not in itself a surprising result, it does highlight an issue with current template-based accounts of spatial term semantics; namely, that these models focus on the geometric relationships between the trajector and the landmark and largely omit interlocutor perspective as a feature. As such they are incomplete. In future work we aim to take the computational models used by robotic systems and move them towards these more complete accounts.

References


