UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

An Empirical Investigation into Spatial Reference Frame Taxonomy using Dialogue

Permalink

https://escholarship.org/uc/item/96r8s236

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 28(28)

ISSN

1069-7977

Authors

Branigan, Holly P.
Watson, Matthew E.
Pickering, Martin J.

Publication Date

2006

Peer reviewed

An Empirical Investigation into Spatial Reference Frame Taxonomy using Dialogue

Matthew E. Watson (s0129081@sms.ed.ac.uk)
Department of Psychology, University of Edinburgh,
7 George Square, Edinburgh EH9 7HZ

Martin J. Pickering (martin.pickering@ed.ac.uk)
Department of Psychology, University of Edinburgh,
7 George Square, Edinburgh EH9 7HZ

Holly P. Branigan (holly.branigan@ed.ac.uk)
Department of Psychology, University of Edinburgh,
7 George Square, Edinburgh EH9 7HZ

Abstract

Reference frames are abstract representations that enable people to describe the location of objects using projective descriptions such as, *above, below, left,* and *right.* There have typically been three broad classes of reference frames identified in the literature: Those that are based upon the environment, those based upon an object and those that are egocentric. However, the exact details of reference frame classification have been disputed. In this paper we present two experiments that used a confederate priming paradigm that set out to distinguish between the Traditional taxonomy and Levinson's taxonomy of reference frames. The results provide evidence supporting the Traditional taxonomy of reference frames. We suggest possibilities for further investigations of reference frame alignment.

Introduction

When a speaker describes the location of an object using a *projective* spatial term (one that describes the location of an object in a region of space projected from a second object e.g. *above, below, left* and *right*), it requires the imposition of a *reference frame* on the scene. The reference frame is an abstract representation that parses space around a *reference* object (the object the figure object is related to) into regions, so that the location of a *figure* object (the located object) can be determined in relation to the reference object.

According to Logan and Sadler (1996), a reference frame is a three-dimensional, axial co-ordinate system that defines origin, orientation, direction and scale parameters. The settings of each of these parameters determine the appropriate description of the location of the figure object. For example, in Figure 1 the ball can be described as *in front of the car,* because it is located in the region projected from the car's headlights, which determine the car's front. In this case, the orientation and direction parameters are set according to the intrinsic sides of the reference object: the car. The orientation is set so that the horizontal, front-back axis is the car's major axis of elongation. The setting of the direction parameter then determines which ends of this axis correspond to the *front* and *back*. The origin of the reference frame is situated upon the reference object. The scale parameter is not explicitly given in the spatial description. However, there is evidence that suggests the scale parameter is set even when not explicitly stated (Carlson & van Deman, 2003), in which case it may be set according to the relative sizes of the figure and the reference object (e.g. *a*

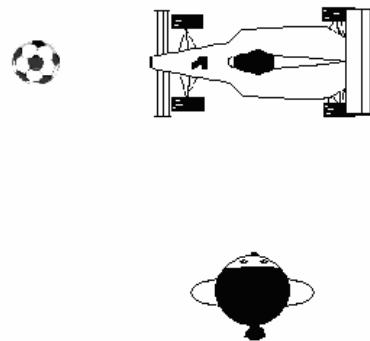*plane above a house* causes the scale parameter to be set larger than *a clock above a bookcase*).

Figure 1: The ball can be described as *in front of the car* or *to the left of the car*

The ball in Figure 1 may, however, also may be described as *to the left of the car*. This description is as felicitous as the description *the ball is in front of the car*; however, the parameters of the reference frame have different settings. In this case, the orientation and the direction parameters are set according to the reader's own directional axes. The horizontal front-back axis is set according to the front and back of the reader. The second horizontal axis (the left-right axis) is then set as orthogonal to the front-back axis. The direction parameter of the left-right axis is then set according to the anchoring system provided by the front-back axis. The origin of the reference frame could be situated upon either the reader or the reference object; it is not clear which of these is the case as both settings would be consistent with the description. The scale parameter is not explicitly given in the description, but may be set using other processes (Carlson & van Deman, 2003).

The different settings of the parameters of a reference frame yield different descriptions of a scene from a speaker, and different interpretations of a locative description from an addressee. The different settings of the parameters also lead to the reference frames being described as different types. For example, the first description given is an object-centered reference frame, whereas the second description given is a person-centered reference frame. But there has

been little agreement in the literature as to how reference frames should be categorized. Many psycholinguistic studies have divided reference frames into *absolute, intrinsic* and *deictic* (e.g. Carlson-Radvansky & Irwin, 1993, Miller & Johnson-Laird, 1976); we shall describe this as the *Traditional taxonomy*. Levinson (1996, 2003), however, used a different system of classification that divides reference frames into *absolute, intrinsic* and *relative*.

According to the Traditional taxonomy, the defining characteristic of reference frames is the object that sets the orientation and direction parameters. If these parameters are set by the reference object, it is an *intrinsic* reference frame. If the parameters are set by either the speaker or the addressee, it is a *deictic* reference frame. The deictic reference frame is often thought of as an *egocentric* reference frame, whereas the intrinsic reference frame is an *object-centric* reference frame. For example, take three different descriptions of the scene shown in Figure 1: *the ball is in front of the car*, *the ball is to the left of the car* and *the car is in front of me* (if the reader assumes the position of the person in the scene). According to the Traditional taxonomy, the first of these descriptions uses an intrinsic reference frame because the orientation and direction parameter are determined by the car. The second two descriptions use a deictic reference frame because the orientation and direction parameters are set egocentrically.

According to Levinson's taxonomy (1996, 2003), the defining characteristic of a reference frame is the argument structure of the description. If the description is a two-place, *binary* description, with a figure object and a reference object as the arguments, then it is an *intrinsic* reference frame. If the description is a three-place, *ternary* description, with a figure object, reference object and a viewpoint as the arguments, then it is a *relative* reference frame. A spatial relationship using a relative reference frame is accurate relative to a viewpoint on the scene.

If we consider the three descriptions of Figure 1, given above, using Levinson's taxonomy then the reference frame that each of the descriptions uses is different to the Traditional taxonomy. The description *the ball is in front of the car* uses an intrinsic reference frame because it is a binary function, with the figure object (the ball) as one argument, and the reference object (the car) as the second argument. The description *the ball is to the left of the car* uses a relative reference frame because it is a ternary function, with the figure object, reference object and the viewpoint of the reader as the three arguments. The description of the scene is only valid if the viewpoint remains constant. The description *the car is in front of me* uses an intrinsic reference frame because it is a binary function with the figure object (the car) and the reference object (the ego) as the two arguments. Note that despite this description being egocentric it still uses an intrinsic reference frame: Levinson's taxonomy affords no special status to an egocentric reference frame.

The grouping of reference frames differs according to the two taxonomies. The Traditional taxonomy groups together the descriptions *the ball is to the left of the car* and *the car is in front of me* as using the same reference frame: a deictic one. Levinson, however, argued that the descriptions *the ball is in front of the car* and *the car is in front of me* use the same reference frame: an intrinsic one. Despite these differences there has been little empirical evidence to distinguish between the two reference frame taxonomies.

In this paper, we present two experiments that used a confederate priming paradigm to investigate which taxonomy correctly categorizes reference frames.

Experiment 1

One of the differences between the two taxonomies is their categorization of sentences of the form *the ball is in front of me*; the Traditional taxonomy argues it is a *deictic* reference frame, whereas Levinson's taxonomy argues it is an *intrinsic* reference frame. These positions can be empirically distinguished using a confederate priming paradigm, in which a confederate who purported to be a naive participant and who followed a script, and a genuine naïve participant, took it in turns to describe pictures. Watson, Pickering, and Branigan (2004) found that naïve participants tended to use the same reference frame to describe a picture as the confederate had used to describe a previous picture. In the experiments reported below, a confederate described the location of a figure object using a scripted description of the form *the X is in front of us* (we refer to this as an *ego-referent* description) and the naïve participant chose which of two pictures matched the description. The participant then described the location of a figure object to the confederate, in the belief that she would have to also match a picture to this description. Following Watson et al. (2004), participants should tend to use the same reference frame for their descriptions as they had just heard from the confederate. The Traditional taxonomy and Levinson's taxonomy make different predictions about which reference frames count as 'the same', and hence which reference frame participants should use. Specifically, if the Traditional taxonomy is the correct way to categorize reference frames, then following hearing an ego-referent description, participants should be significantly more likely to use a relative/deictic reference frame of the form *the X is to the left of the Y* than an intrinsic reference frame of the form *the X is in front of the Y*. However, if Levinson's taxonomy is correct, then following hearing *the X is in front of us* participants would be more likely to use an intrinsic reference frame of the form *the X is in front of the Y* than a relative/deictic reference frame of the form *the X is to the left of the Y*.

Method

Participants: 16 native English-speaking students at the University of Edinburgh were paid to participate in the experiment. The confederate was a female postgraduate at the University of Edinburgh.

Materials: The experiment was run using two E-prime computer programs. One computer program presented locative sentences upon a computer screen and was used by the confederate. A second program was created for the participant which presented pictures for the match phase and the describe phase of each trial.

Each of the scenes used in the experiment involved a picture of two people (participants were told that they represented the participant and the confederate), one object that had intrinsic sides and was viewed from the side, termed a *tri-axial* object (e.g. a car), and one object that had no horizontal intrinsic axes, termed a *bi-axial* object (e.g. a tree) (Carlson-Radvansky & Jiang, 1998).

All the *match* and *target* scenes contained the two people at the centre and bottom of the screen, one tri-axial object and one biaxial object. The tri-axial and bi-axial objects were selected from three 'land' bi-axial items (mast, tree, and skyscraper) and three 'land' tri-axial objects (car, catapult, and train) or three 'sea' bi-axial objects (buoy, lighthouse and island) and three 'sea' tri-axial objects (ship, windsurfer, and hovercraft). The sea and land items were never paired together in a picture. One of the bi-axial or tri-axial objects was presented in the centre of the screen and one to the left or the right of the central object. All of the match and target scenes in the experiment therefore appeared in an arrangement as in Figure 2.
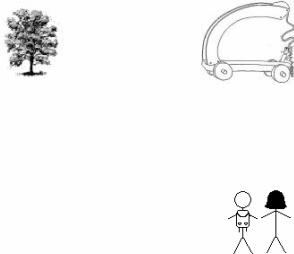

Figure 2: The arrangement of objects in Experiment 1.

Design: There was one within-participants and within-items factor of Prime reference frame (*ego-referent*, *intrinsic*, and *relative/deictic*). The ego-referent description of Figure 2 would be *the catapult is in front of us*. The intrinsic description of Figure 2 would be *the tree is in front of the catapult*. The relative/deictic description of Figure 2 would be *the tree is to the left of the catapult*. Each object used in the experiment appeared as the figure object an equal number of times and as the reference object an equal number of times. There were 108 trials each consisting of a *prime sentence,* where the confederate read a sentence which described the location of one of the objects on one of the scenes on the match scene, a *match phase,* where the participant saw two scenes on their screen and had to decide

which matched the confederate's description, and finally a *target phase,* where the participant was told, via a word on the screen, to describe the location of one of the objects on the target scene.

Fifty-four of the trials were experimental trials, 18 in each of the three conditions. The match scenes were created by having one scene matching the confederate's description exactly and a second scene with the same objects, but in a different spatial arrangement. These scenes were positioned so that one was on the left of the screen and one on the right; the position of the matching scene was counter-balanced.

The target scenes were created from the same pool of objects as the objects used in the match scenes. Again the 18 target scenes in each condition were formed from combinations of the tri-axial and bi-axial objects. The target scene always contained different objects (except for the two people who were present on all scenes) from the match scenes, with the stipulation that the objects were of the same type as in the match scene (e.g. if the match scene used the sea objects then so did the target scene). On experimental target scenes the tri-axial object was always in the centre of the screen and the bi-axial object was to the left or right. For the experimental trials, the participant always had to describe the location of the bi-axial object. The spatial relationship between the tri-axial and bi-axial objects was always different on the target scene from the match scene.

The remaining 54 trials were filler trials. 18 of the filler trials had a relative reference frame prime sentence, 18 had an intrinsic reference frame prime sentence, and 18 had the ego-referent prime sentence. On the filler trials the figure object on the target scene, which the participant had to describe the location of, was always the tri-axial object. This meant that on the filler trials there was no possibility of using the intrinsic reference frame because bi-axial objects have no horizontal intrinsic axes.

Procedure: The confederate and naïve participant were introduced as if they were both naïve participants and seated at computers on two desks side by side; a divider prevented them from seeing each others' screens. The participant and the confederate were told that they would be taking it in turns to describe the location of objects to each other and then decide which of two scenes on their screen matched their partner's description. They were also told that all the pictures contained two people at the bottom of the screen which represented them both, and that they should treat themselves as part of the scene. They were also told that all the pictures would contain two other objects, one in the centre of the screen and one that would be one side or the other of the central object. Participants were told not to use the screen to describe the location of the objects; such as *the mast is to the left of the screen.*

The participants then pressed the space bar to begin the practice session. The practice session lasted six trials, one for each of the three experimental conditions, and three

fillers. Instructions on the screen signalled the end of the practice session and the start of the experiment.

Each trial proceeded as follows: After participants pressed the space bar to begin, they were presented with the match scenes. The confederate then followed her script to describe the location of an object that was present on the participant's match scenes. The participant then decided which match scene matched the confederate's description, pressing the Z key if they thought it was the left hand scene and the M key if they thought it was the right hand scene. A fixation cross then appeared for 1000ms, before being replaced by a word. This word stayed on the screen for 1500ms and named the object which was to be the figure object on the target scene. When the word disappeared it was replaced by the target scene. Participants then had to describe the location of the named object. After describing the location of the figure object on the target scene participants pressed the space bar and the target scene was replaced with a fixation cross. The fixation cross remained for 500ms, before being replaced with the next match scene. The confederate pressed the space bar after hearing the participant's description; this caused the next prime sentence to appear on her screen. This was the procedure for all practice, experimental and filler trials.

Results

We analyzed participants' first responses (i.e. if participants used one reference frame, but then switched to an alternative reference frame the response was classified according to their first response). Responses were coded as using either an *intrinsic* or *relative/deictic* reference frame or *other* response (3.5% of responses) if the response did not use a reference frame (e.g. *the train is travelling towards the skyscraper*). Analysis was performed only on the relative/deictic and the intrinsic responses. This was done by dividing the percentage of intrinsic responses by the percentage of relative/deictic responses plus the number of intrinsic responses to give a reference frame index score (RI). Table 1 shows the mean RI for each of the three conditions. The RI was analysed using two within-participants one-way ANOVAS, one for participants (F1) and one for items (F2), with the factor Prime Reference Frame (*relative/deictic*, *intrinsic* and *ego-referent*). There was a significant main effect of Prime Reference Frame ($F1$ $(2,30) = 3.46$; $p < .05$; $F2$ $(2,34) = 11.47$; $p < .01$). Planned comparisons showed that there was a significant difference between the *intrinsic* prime condition and the *relative/deictic* prime condition ($t1(15) = 2.54$; $p < .05$, $t2(17) = 3.76$; $p < .01$). Participants were more likely to use an intrinsic reference frame after hearing an intrinsic reference frame utterance, than after a relative/deictic reference frame utterance. There was also a significant difference in the RI between the *intrinsic* prime condition and the *ego-referent* prime condition ($t1(15) = 2.15$; $p < .05$, $t2(17) = 3.8$; $p < .01$). Participants were more likely to use an intrinsic reference frame after hearing an intrinsic reference frame utterance, than after an ego-referent

reference frame utterance. Finally, there was no significant difference in the RI between the *relative/deictic* prime condition and the *ego-referent* prime condition ($t1(15) = 0.13$; $p > .05$, $t2(17) = 0.16$; $p > .05$) Participants were just as likely to use a relative reference frame after an *ego-referent* reference frame utterance, as after a relative reference frame utterance.

Condition	Relative /deictic	Intrinsic	Ego-referent
Mean Reference frame index	0.23	0.34	0.24
Mean percentage Intrinsic responses	20.5%	31.6%	23.3%
Mean percentage Relative responses	69.1%	61.8%	72.2%

Table 1: The mean reference frame index for each of the conditions in Experiment 1.

Discussion

The results show that participants were more likely to use an intrinsic reference frame after hearing the confederate use an intrinsic reference frame, than after hearing the confederate use an alternative reference frame. This is in line with other research showing that interlocutors align reference frames (Watson et al, 2004) and reflects a general tendency for interlocutors to align representations that underlie language production and comprehension (Pickering & Garrod, 2004). More importantly, they show that participants were more likely to use a relative/deictic reference frame after hearing an ego-referent description than an intrinsic reference frame. Assuming that this tendency reflects reference-frame alignment, it constitutes evidence that supports the Traditional taxonomy of reference frames and is inconsistent with Levinson's taxonomy.

Experiment 2

The evidence from Experiment 1 supports the Traditional taxonomy; however, there is a possible alternative explanation. In Experiment 1, in the intrinsic and the relative/deictic prime reference-frame conditions the reference object was in the *same* position on the target scene as on the match scene. However, in the ego-referent condition the reference object was in a *different* position on the match scene to the target scene. This is because the two people were the reference object for the prime and were always at the bottom of the screen, whereas the reference object on the target was the tri-axial object, positioned in the centre of the screen.

Previous research has shown that the more factors the prime and target have in common, the more interlocutors align (e.g. Branigan et al, 2000; Cleland & Pickering, 2003). Hence, participants might not have used an intrinsic reference frame, in the ego-referent condition, because of

the reduced common factors between match scene and target scene. They may instead have used a relative/deictic reference frame because this is preferred to the intrinsic reference frame.

Experiment 2 therefore tested whether or not a different position of the reference object on the match scene and target scene affects levels of alignment. This experiment added a fourth condition (*intrinsic-different*) where the confederate used an intrinsic reference frame, but the reference object was placed in the position in which two people appeared in the other conditions. If interlocutors align reference frames independently of position then participants should be more likely to use a reference frame when they had just heard the confederate use that reference frame compared to an alternative reference frame, regardless of whether the reference object was in the same position or a different position on the match scene and target scene

Method

Participants: 18 native English-speaking students of the University of Edinburgh who had not participated in Experiment 1 were paid to participate. The confederate was a male postgraduate at the University of Edinburgh.

Materials and Design: In this experiment each match and target scene contained three objects: a football, two people side-by-side viewed from overhead, and one of 10 tri-axial objects also viewed from overhead. The perspective was switched to overhead because of the difficulty in finding tri-axial objects that, when positioned at the bottom of the screen, appeared to be looking into it in an analogous fashion to the people in Experiment 1 (see Figure 2). There was one within-participants and within-items factor of Prime Reference frame, with four levels (*relative/deictic, intrinsic, intrinsic-different, ego-referent*). The first three conditions were analogous to those in Experiment 1, except from an overhead perspective. The *intrinsic-different* condition had the reference object, on the match scene, positioned at the centre and bottom of the screen, as shown in Figure 3. The confederate then described this as *the ball in front of the tank.* In this condition, the confederate used an intrinsic reference frame; however the reference object was in a different position on the target scene to the match scene (the reference object was always in the centre of the screen for the target scenes).

The target scenes always contained two people at the bottom of the screen (centrally placed), one of the objects in the centre of the screen and a ball to the left or right of this object (analogous to the layout in Figure 2).

The reference object used in the match scene was always different to the reference object used in the target scene. In each list all reference objects appeared once per condition in the match scene, and once per condition in the target scene. This meant that each object was seen in the match scenes

four times and seen in the target scenes four times. As there were 10 objects this made a total of 40 trials (10 per condition) in the experiment. All of the trials were experimental; there were no fillers.
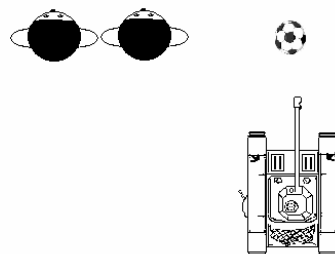

Figure 3: Stimulus from the *intrinsic-different* condition in Experiment 2.

Procedure: The procedure was the same as for Experiment 1.

Results

The responses were coded and converted to a reference frame index score as in Experiment 1. Table 2 shows the mean RI and the mean percentage intrinsic and relative/deictic scores for each of the four conditions. The mean RI scores were analysed using two one-way, within-participants (F1), and within-items (F2) ANOVAs, with Prime Reference frame (intrinsic, relative/deictic, intrinsic-different and ego-referent) as the factor and levels. There was a significant main effect of Prime Reference Frame by both participants and items ($F1(3,51) = 5.4$; $p < .01$, $F2(3,27) = 3.1$; $p < .05$). Planned comparisons showed that there was a significant difference between the *relative/deictic* and *intrinsic* Prime Reference Frame conditions, both by participants and items ($t1(17) = 4.4$; $p < .01$, $t2(9) = 2.6$; $p < .05$). Participants used an intrinsic reference frame more after hearing an *intrinsic* reference frame (0.77 RI) than after a *relative/deictic* reference frame (0.6 RI). There was also a significant difference between the *intrinsic* and the *intrinsic-different* Prime Reference Frame conditions both by participants and items ($t1(17) = 2.28$; $p < .05$, $t2(9) = 2.53$; $p < .05$). Participants used an intrinsic reference frame more often after hearing an *intrinsic* reference frame when the reference object position was held constant from the match scene to the target scene (0.77 RI) than when the reference object position changed from match scene to target scene (0.68 RI).

Condition	Relative/deictic	Intrinsic	Intrinsic-different	Ego-referent
Mean Reference frame Index	0.60	0.77	0.68	0.68
Mean percentage Intrinsic responses	50.6	62.8%	57.8%	55%
Mean percentage Relative responses	33.9	18.9%	27.8%	26.1%

Table 2: The mean reference frame index scores, percentage intrinsic and deictic responses for each of the four conditions in Experiment 2.

There was no significant difference between the *relative/deictic* and *ego-referent* Prime Reference Frame conditions (t1(17) = 1.47; p > .05, t2(9) = 0.92; p > .05). Participants used an intrinsic reference frame just as often after hearing the confederate use a *relative/deictic* reference frame as after hearing the confederate use an *ego-referent reference* frame.

General Discussion

The two experiments presented here set out to test whether reference frames should be categorized according to the Traditional taxonomy or Levinson's taxonomy. The results of Experiment 1 supported the Traditional taxonomy: Participants were more likely to use a relative/deictic reference frame after hearing the confederate use an ego-referent reference frame. We interpret this finding as showing that the two different utterances are implicitly treated as uses of the same reference frame.

Experiment 2, however, raises some questions about this interpretation. In this experiment the three conditions that were replicated from Experiment 1 showed a similar pattern of reference frame usage. However, when the confederate used an intrinsic reference frame and the reference object was in a different position on the match scene and the target scene, the participant was less likely to use an intrinsic reference frame than when the reference object's position was held constant. Despite this, participants were still more likely to use an intrinsic reference frame following an intrinsic prime than following a relative/deictic prime. This suggests that the reference object's position on the match and target scenes did affect the likelihood of alignment, but that this is not the reason for the patterns of reference frame use after the confederate used an ego-referent reference frame. Overall there were more intrinsic responses in Experiment 2 than Experiment 1. This is probably due to the greater difficulty of using a relative/deictic reference frame from the unusual, overhead viewpoint. Nevertheless the pattern of alignment across the two experiments is similar.

We therefore tentatively conclude that the results presented here support the Traditional taxonomy of reference frames. Hence it is the object setting the orientation and direction parameters of the reference frame that is important for classification of reference frames. Our results show, further, that it is these aspects of a reference frame on which interlocutors align in dialogue, and that these parameters may be the most salient features of a reference frame. Such a conclusion must be tentative because many other factors relevant to spatial language may have had an influence (e.g. features of the reference object, syntactic structure of the descriptions). Such factors and their influence on alignment in spatial language in dialogue are issues for future research.

References

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition,* 75, B13-B25.

Carlson, L. A., & van Deman, S. R. (2004). The space in spatial language. *Journal of Memory and Language,* 51,418-436.

Carlson-Radvansky, L. A., & Irwin, D. E. (1993). Frames of reference in vision and language: Where is above? *Cognition*, 46, 223-244.

Carlson-Radvansky, L. A., & Jiang, Y. (1998). Inhibition accompanies reference frame selection. *Psychological Science*, 9, 386-391.

Cleland, A. A., & Pickering, M. J. (2003). The use of lexical and syntactic information in language production: Evidence from the priming of noun-phrase structure. *Journal of Memory and Language*, 49, 214-230.

Levinson, S. C. (1996). Frames of reference and Molyneux's questions: Cross linguistic evidence. In P.Bloom, M. A. Peterson, L. Nadel, & M. Garrett (Eds.), *Language and space,* (pp. 109-169).Cambridge, MA: MIT Press.

Levinson, S. C. (2003). Space in language and cognition: Explorations in cognitive diversity. Cambridge: Cambridge University Press.

Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and perception*. Cambridge, MA: MIT Press.

Pickering, M. J., & Garrod, S. C. (2004). Towards a mechanistic theory of dialogue. *Behavioral and Brain Sciences, 27,* 169-226.

Watson, M. E., Pickering, M. J., & Branigan, H. P. (2004). Alignment of reference frames in dialogue. *Proceedings of the 26th Cognitive Science Meeting: Chicago, 2004.*