# Ambiguities in Spatial Language Understanding in Situated Human Robot Dialogue *

**Changsong Liu**
Department of Computer Science and Engineering
Michigan State University
East Lansing, MI 48824
cliu@msu.edu

**Jacob Walker**
Department of Computer Science and Engineering
Michigan State University
East Lansing, MI 48824
walke434@msu.edu

**Joyce Y. Chai**
Department of Computer Science and Engineering
Michigan State University
East Lansing, MI 48824
jchai@cse.msu.edu

### Abstract

In human robot dialogue, identifying intended referents from human partners' spatial language is challenging. This is partly due to automated inference of potentially ambiguous underlying reference system (i.e., *frame of reference*). To improve spatial language understanding, we conducted an empirical study to investigate the prevalence of ambiguities of frame of reference. Our findings indicate that ambiguities do arise frequently during human robot dialogues. Although situational factors from the spatial arrangement are less indicative for the underlying reference system, linguistic cues and individual preferences may allow reliable disambiguation.

## Introduction

With the emergence of a new generation of social robots (Fong, Nourbakhsh, and Dautenhahn 2003), techniques to enable situated language understanding have become increasingly important. In situated human-robot interaction, it is ideal for human partners to use spatial expressions (e.g., *"the cup to the left of the computer"*) rather than complex and descriptive noun phrases (e.g., *"the green cup with a big S on it"*) to refer to objects in the environment (Skubic et al. 2004; Moratz 2006).

However, correctly interpreting spatial language and identifying intended referents can be challenging. For example, suppose a human and a robot are co-present in an environment where three cones surround a chair (as shown in Figure 1). The human issues a command *"go to the cone to the right of the chair"*. Given this expression, the robot could have different interpretations of the intended object. It could be Cone 1 if *"right of the chair"* is interpreted based on the speaker's perspective; or Cone 2 if it is based on the robot's perspective, or Cone 3 based on the perspective of a third entity, namely the chair's intrinsic right side. So which cone should the robot go to?

This example shows a typical problem with spatial language understanding, which arises from the ambiguities of the underlying reference system of spatial language. This reference system is often called *frame of reference* in the literature (Levinson 1996; Frank 1998) (more description in section 2). A spatial situation may allow a speaker to use multiple frames of reference. Sometimes, this poses no problem since different available frames of reference lead to an unique interpretation. However, very often, as in the example shown here, they do lead to ambiguous interpretations for a listener.


Figure 1: An example of ambiguity.

A large body of research on human-human communication (Petra Weiss and Miller 1996; Tenbrink 2005; Coventry et al. 2005) have indicated that the speaker's choice of frame of reference depends on a wide range of situational and discourse-related factors. Inspired by these previous studies, we conducted an empirical investigation to understand the prevalence of interpretation ambiguities caused by frames of reference in human-robot interaction and examine different contextual factors for disambiguation. Our findings indicate that ambiguities do arise frequently during human robot dialogues. Although situational factors from the spatial arrangement are less indicative for the underlying reference system, linguistic cues and individual preferences may allow reliable disambiguation.

## Frame of Reference

One most commonly used class of spatial language is projective terms (Eschenbach 2005), such as *front, back, right, left*. Projective terms describe the spatial relation between two entities based on an underlying reference system, which is often called *frame of reference*. In this reference system, one entity serves as *relatum*, and the other entity (the *referent*) is spatially included in a region that is derived from the *relatum* and a *principal axis* (*direction*). A particular

*This work was supported in part by IIS-0347548 and CNS-0957039 from the National Science Foundation.

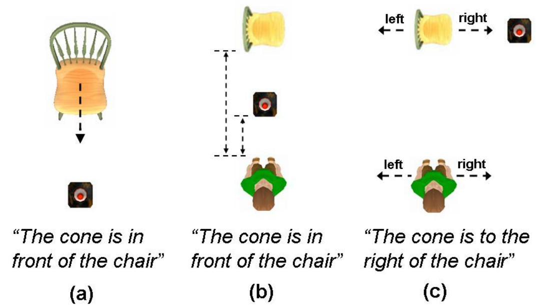



Figure 2: Conceptualization of frame of reference: (a) Intrinsic; (b) Relative use of *"front"*; (c) Relative use of *"right"*.

*perspective (point of view)* is then taken to specify the particular relation. Levinson (Levinson 2003) systematically categorized frame of reference into three kinds: *absolute*, *intrinsic* and *relative*. In an absolute frame of reference, fixed-bearing systems such as north and south serve as anchor directions (Levinson 2003). This option is ruled out in our study because it is usually not preferred in small scale (indoor) scenarios in Western countries (Tenbrink 2005).

In an intrinsic frame of reference, the perspective point (called *origin* by Levinson) is coincided with the relatum, and the position of the referent is described by referring to the relatum's intrinsic perspective. It thus requires the relatum to possess intrinsic sides which then serve as basis for reference. For example, in the expression *"the cone is in front of the chair"*, the position of the cone is specified by referring to the chair's front side (Figure 2(a)). The intrinsic sides/parts of an object is often determined based on human cognition and behavioral conventions with regard to the object's features and functionalities (Levinson 2003). For example, front and back are found by taking canonical direction of motion, canonical direction of use etc. into account. For the lateral sides (i.e., left and right), two types of objects have opposite assignments: An object that serves as body extension (e.g., a chair) inherits its right and left from the human beings who wear it or drive it or sit in it; If an object is characteristically confronted by human beings, such as a desk or a laptop computer, its right/left side is then transferred from the closest right/left side of human.

The relative frame of reference, on the other hand, uses the position of a *viewer* as the origin instead of referring to inbuilt features of the relatum. Relative uses of frontal terms (front/back) deal with distance and occlusion, thus the referent is *"in front of"* the relatum means the referent (partially) occludes the relatum or is closer to the viewer than the relatum is (Figure 2(b)). Relative uses of lateral terms (right/left) are interpreted as the viewer projects his own intrinsic right/left sides onto the relatum, so that the relative position of the referent to the relatum can be specified using the imposed sides (Figure 2(c)). Unlike the intrinsic frame of reference which is viewer-independent and relies on the inherent properties of the relatum, the relative frame of reference is viewer-dependent and irrelevant to the relatum's intrinsic directions.

Besides the general categorization, both the intrinsic and the relative frames of reference have three variations depending on whether the speaker, the addressee, or a third entity serves as the origin. These three variations are called *egocentric*, *addressee-centered* and *object-centered* in the literature (Schober 1996; Trafton et al. 2005). The categorization of frame of reference is summarized in Table 1.

| **Frame of Reference** | | **Origin** | **Example** |
|---|---|---|---|
| intrinsic | egocentric | speaker | "*the cone to the right of me*" |
| | addressee-centered | addressee | "*the cone to the right of you*" |
| | **object-centered** | **3rd entity** | ***"the cone to the right of chair"*** |
| relative | **egocentric** | **speaker** | ***"the cone to the right of the chair (from my viewpoint)"*** |
| | **addressee-centered** | **addressee** | ***"the cone to the right of the chair (from your viewpoint)"*** |
| | object-centered | 3rd entity | "*the cone to the right of the chair (viewed from the chair's position)*" |

Table 1: Categorization of frame of reference.

Among different types of frames of reference, the *egocentric intrinsic* and the *addressee-centered intrinsic* are easy to recognize since they tend to have *"me/my"* or *"you/your"* in the expressions, as in *"in front of me/you"* or *"to my/your right"*. The *object-centered relative* which is not quite applicable in two-party conversation. The remaining types: the *egocentric relative*, the *addressee-centered relative*, and the *object-centered intrinsic* are often simultaneously applicable to a single expression (e.g., *"the cone to the right of the chair"* as shown in the previous example). Previous study (Tenbrink 2005) has indicated that these three types of frames of reference are in general not distinguishable on the basis of the linguistic form alone. Therefore, this paper particularly focuses on the ambiguities between these three frames of reference (marked as bold in Table 1), which caused different interpretations in our example in Section 1.

## Related Work

Frame of reference and spatial language have been investigated from a number of perspectives. Given a situation, the speaker's choice of one particular frame of reference is shown to be flexible, as a function of several situational and contextual factors. For instance, the configuration of an environment and spatial cues can affect the choice of frame of reference (Taylor and Tversky 1996; Mintz et al. 2004), so can the features of objects and the functional relationship between objects (Carlson and Radvansky 1996), the communicative aspect of tasks (Petra Weiss and Miller 1996; Tenbrink 2005), and the discourse history and alignment (Schober 1993; Watson, Pickering, and Branigan 2004). Choices of frame of reference can also be affected by the speakers' individual preferences (Levelt 1982; Tenbrink, Maiseyenka, and Moratz 2007).

In human robot interaction, previous work (Moratz and Tenbrink 2006) found that human users overwhelmingly choose addressee-centered frame of reference to facilitate the robot's comprehension of spatial referents. Such *partner-adaption* (Tenbrink 2005) behaviors greatly reduced the potential ambiguity when the speaker's and addressee's perspectives were not aligned. However, these studies have not investigated potential ambiguities between the relative and the object-centered intrinsic frame of reference.

In a recent study based on a simulated 2D environment (Moratz 2006; Tenbrink, Maiseyenka, and Moratz 2007), ambiguities were found between relative and object-centered intrinsic frame of reference which have led to conflicting interpretations. However, they only used 2D pictures with abstract icons representing a robot and objects to elicit users' spatial expressions. Such 2D pictures forced a "survey" perspective (Tenbrink 2005) which is very different from the normal perspective in situated human robot interaction. Thus it is not clear to what extent their data reflect the truly situated environment. Nevertheless, these earlier works in human robot interaction have inspired our work in this paper. The difference between our work and previous work is that we focus on the study of ambiguities between the relative and the object-centered intrinsic frame of reference in a truly situated context and we are interested in identifying contextual factors that may lead to reliable disambiguation.

## Empirical Investigation

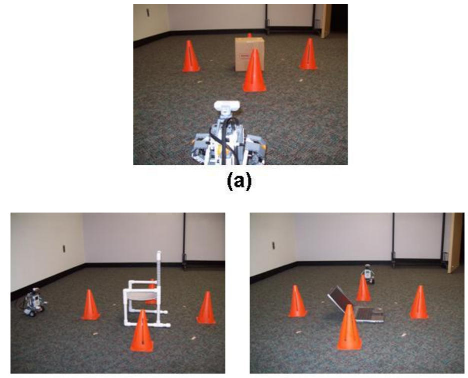
We conducted empirical studies in a Wizard-of-Oz setting, based on human interaction with a LEGO NXT robot. The users were made to believe that they were interacting with an autonomous robot, meanwhile the actual responses of the robot were controlled by a human. Microsoft Robotics Developer Studio was used to implement control routines and the wizard interface.

### Spatial Arrangement

The spatial arrangement in our studies contained a human user, a robot, four identical cones, and a center object which was either a box, a chair or a laptop computer. Each of the four cones was placed on each of the four sides of the center object. During the experiment, the user and the robot were situated at different positions. Figure 3 shows three examples of the spatial arrangement (from the human user's view) in our study. More specifically, our spatial arrangements were defined by the following three factors:

*Center object.* As mentioned in section 2, determining an object's intrinsic sides depends on its features and functionalities. To study whether the features of different objects may affect the choice of frame-of-reference, we used three different objects to serve as a potential relatum in our arrangements: the box has no intrinsic sides, the chair is a body extension object, and the laptop is a confronted object.

*Displacement between human and robot.* Previous studies have indicated that the difficulty of taking other's spatial perspective is a function of the displacement between the egocentric view and the to-be-imagined view (Presson and Montello 1994; Rieser 1989), and that the difficulty of taking a different perspective may influence the speaker's choice of frame of reference (Schober 1996). Inspired by these findings, we designed four settings of displacements in our studies: 0° (the user and the robot shared the same view point), 90° (the robot faced to the user's right or left), 180° (the user and the robot faced to each other), and 135° (the robot's facing direction was between the 90° and 180°).



(a)

(b) (c)

Figure 3: Examples of spatial arrangements from the human user's point of view: (a) The center object is a box, and the robot has a same view point from the human. (b) The center object is a chair, and the robot is at a 90 degrees position from the human. (c) The center object is a laptop, and the robot has an opposite view point (i.e., 180 degrees) from the human.

*Orientation of the center object.* According to the "principle of ease of spatial terms" (Tversky, Lee, and Mainwaring 1999), asymmetric spatial terms (front/back) is easier to produce and understand than symmetric terms (right/left). In other words, whether the referent is on the relatum's frontal or lateral side is related to the difficulty of using the intrinsic frame of reference, thus it may influence the speaker's choice of frame of reference. In our experiment, we addressed this issue by varying the orientation of the center object with four settings: either faced toward or away from the robot, or to the right or the left of the robot.

With these three factors, the total number of all possible spatial arrangements is $3 \times 4 \times 4 = 48$. Since in each arrangement, any one of the four cones could be the target (i.e., referent to be specified by a user), in total we have 192 (i.e., $48 \times 4$) experimental configurations. Obviously, it is not possible to conduct experiments based on all of these arrangements. The orthogonal array method (Sloane 2001) solves this problem by selecting a subset of configurations which guarantees testing the pair-wise combinations of all the factors with an even distribution. Thus we applied the orthogonal array method and selected 24 out of all the possible arrangements for our experiments.

### Experimental Procedures

We recruited 18 students (4 female and 14 male) with various majors at Michigan State University to participate in our experiment. Each user went through all 24 arrangements one by one, in a randomized order. In each arrangement, the moderator indicated the target cone using a laser pen pointer, and the user then described to the robot which one of the four cones the robot should move to. As the four cones are identical, only spatial relations can be used to distinguish the intended cone from the others.

To simulate the realistic situation in which the robot fails to understand human language, we also simulated communication failures in our experiments as follows. In 8 of the 24 arrangements the users' expressions would be correctly understood and the corresponding actions would be performed by the robot. In 12 configurations the robot would initially fail to understand the user's command (i.e., by replying *"I don't understand, please rephrase your command."*) and would succeed when the user made his second attempt. There were 4 arrangements in which the robot would never succeed to follow the commands, and we would stop the user and continue to the next arrangement after three attempts were made. Each of the 24 arrangements was assigned to one of the three cases in a random manner.

## Empirical Findings

Two of the 18 users did not use any descriptions based on frame of reference, but rather using path-based descriptions, such as *"drive a little forward, and turn 45 degrees to left, and then drive forward, and then turn right"*. We excluded the data from these two users for our analysis. From the remaining 16 users, we collected 672 utterances in total. The average number of utterances for one user is 42.[1] For each utterance, we manually annotated the underlying frame of reference based on its corresponding arrangement in the experiment. Next we report our findings based on this data.

### Overall Patterns

Among the 672 utterances, 304 utterances were issued after a simulated communicative failure. When simulated failure happened, 67.8% of the time (206 utterances) users just repeated the same utterance or made some trivial changes, such as changing the verb from *"go"* to *"move"* or changing the noun from *"laptop"* to *"computer"*. In the other 32.2% (98) utterances, users rephrased their expressions. They often did this by switching to a different strategy (66 (21.7%) utterances), such as switching from a frame of reference based expression to a path description or vice versa. It was less likely (only 32 (10.5%) utterances) that users would switch from one type of frame of reference to another. When this did happen, users most often (23 (7.6%) utterances) switched to the egocentric/addressee-centered intrinsic frame of reference. These observations implied that users were generally consistent with their choices of frame of reference in a given situation.

Out of 672 utterances, 193 did not involve any frame of reference, but rather were based on other strategies such as path or distance based descriptions. Thus they were excluded from the analysis of ambiguity of frame of reference. In another portion of 103 utterances, either the user or the robot was used as the relatum, namely either the egocentric or addressee-centered intrinsic frame of reference was used. These utterances were not ambiguous because of their unique linguistic forms (e.g. *"in front of you/me"*).

Since our investigation focused on utterances that were potentially ambiguous with respect to the underlying frame of reference, we only include the remaining 376 utterances for analysis. By removing 86 duplicated utterances (users repeated the previous utterance or made trivial changes) from the 376 utterances, we have 291 utterances without duplication. For those utterances, either the egocentric relative, the addressee-centered relative, or the object-centered intrinsic frame of reference can be applied. The interpretation could be different from one frame of reference to another.

Based on the spatial arrangement from which each utterance was collected, we annotated the 291 utterances in terms of its underlying frame of reference. For some utterances, multiple types of frame of reference might apply since they all lead to the same interpretation. In such cases, we considered these frames of reference as coincided, which did not present any ambiguity in interpretation.

### Situational Factors

Based on the 291 utterances and their annotations, we examined how situational factors may affect speakers' choices of frame of reference.

*Center Object.* Table 2 shows the number of utterances using different types of frame of reference with respect to different types of the center object. The utterances are separated into three groups (the three rows of the table) with respect to the center object in the spatial arrangement. Each of the first three columns shows the number of utterances that used the corresponding frame of reference. The last column shows the numbers of utterances with coincided frames of reference.

| | egocentric relative | addressee-centered relative | object-centered intrinsic | coincident |
|---|---|---|---|---|
| box | 2 | 60 | N/A | 18 |
| chair | 0 | **45** | **28** | 31 |
| laptop | 0 | **49** | **24** | 34 |

Table 2: Numbers of utterances using different frames of reference for different center objects.

As shown in the first row of the table, the box does not possess intrinsic sides, thus only relative frame of reference can be used. In this case users overwhelmingly adopted the robot's perspective by using the addressee-centered frame of reference. This implies that when the relatum has no intrinsic sides, addressee-center relative frame of reference is preferred and ambiguity is not likely to arise.

The second and third rows show the cases of the chair and the laptop. Similar to the case of the box, the egocentric frame of reference rarely happened. However, when objects with intrinsic sides are present, there is a high potential of ambiguity between the addressee-centered relative and object-centered intrinsic frames of reference as the speakers' choices between them can be flexible. We have 146 truly ambiguous utterances as highlighted in the second and third rows in Table 2. Among them, 52 (35.6%) utterances are truly with the object-centered intrinsic frame of reference and 94 (64.4%) utterances are truly with the addressee-centered relative frame of reference. Thus, our following analysis is focused on the ambiguity between these two types of frame of reference (we just call them *relative* and *intrinsic* respectively).

[1] A small portion (4.5%) of utterances was discarded due to users' incomplete utterances and recording errors.

*Displacement and Orientation.* As discussed in Section 4.1, the displacement between the user and the robot and the orientation of the center object were also involved in constructing the spatial arrangements. Each of the two factors has four settings. The numbers of utterances using the relative and the intrinsic frame of reference under each setting of the two factors are shown in Table 3. To investigate whether different settings of the two factors may affect speakers' choices of frame of reference (i.e. the proportion of using one frame of reference), for each factor we used chi-square test to compare the proportions of the relative and the intrinsic frames of reference under different settings. However, the results indicate that the proportions of using frames of reference are not significant different under different settings of the two factors (for the displacement, $\chi^2 = 9.08, p < 0.247$; for the orientation, $\chi^2 = 5.42, p < 0.609$).

| | **Displacement** | | | |
|---|---|---|---|---|
| | 0° | 90° | 135° | 180° |
| Intrinsic | 20 | 3 | 19 | 10 |
| Relative | 25 | 10 | 23 | 36 |
| | **Orientation** | | | |
| | away | left | right | toward |
| Intrinsic | 11 | 19 | 19 | 3 |
| Relative | 27 | 29 | 24 | 13 |

Table 3: Numbers of utterances using two frames of reference for different settings.

## Linguistic Cues

Although situational factors alone may not allow reliable disambiguation of the underlying frame of reference, the linguistic expressions specified by the users may provide important cues.

Users were usually (in 121 of the 146 utterances) explicit with the relatum in their utterances. In the remaining 25 utterances, the relatum was not explicitly specified, as in the utterance like *"go to the left cone"*. As shown in Figure 4, when the relatum was underspecified, it was most likely (96.2%) the relative frame of reference is used. It is significantly greater than the overall percentage (64.4%) of using relative frame of reference ($z = 3.30, p < 0.0001$). However, when the relatum is explicitly mentioned, both the intrinsic (46.3%) and the relative (53.7%) frames are likely.

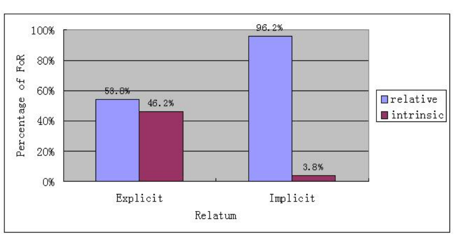



Figure 4: Proportion of the underlying intrinsic and relative frames of reference with explicit and implicit relatum.

The spatial terms used in the utterances also provide cues for the underlying frame of reference. Here we call *"front"* and *"behind/back"* as frontal term and *"left"* and *"right"* as lateral term. Figure 5 shows the proportion of the underlying relative and intrinsic frames of reference with respect to whether a frontal or a lateral term was used in the utterances. As shown in the figure, when a lateral term is used, it is more likely (84.2%) to be generated by the relative frame of reference. This is also significantly greater than the overall percentage of using relative frame of reference ($z = 3.61, p < 0.0001$). When a frontal term is used, both types of frame of reference are likely to happen.

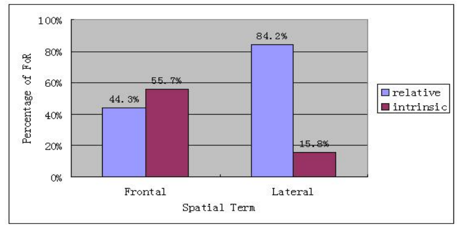



Figure 5: Proportion of the underlying intrinsic and relative frames of reference with frontal and lateral terms.

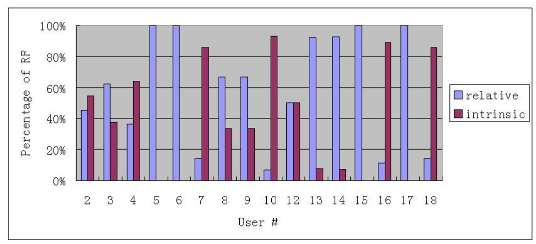



Figure 6: Individual preferences of using frames of reference.

## Individual Preference

Previous studies (Levelt 1982; Tenbrink, Maiseyenka, and Moratz 2007) found that there could be personal style differences between speakers in using spatial language, such as individual preferences on choices of linguistic forms, choices of relatums, or even choices of frames of reference. The results from our study have confirmed the strong individual preferences on the choices of frame of reference.

In the 146 utterances that are potentially ambiguous between the intrinsic and relative frame of reference, we observed that more than half (10) of the 16 users showed strong individual preferences on either the relative or the intrinsic frame of reference. Figure 6 shows the percentage of using the relative and intrinsic frames of reference from each user. As shown in the figure, User 7, 10, 16 and 18 consistently used the intrinsic frame of reference while User 5, 6, 13, 14, 15, and 17 consistently used the relative frame of reference. This observation suggests that uncovering the user's individual preference during interaction provide a potential direction to resolve the ambiguities in frame of reference.

# Discussion and Conclusion

Based on a small data set collected from human-robot interaction, we have investigated the ambiguities of spatial language understanding caused by the underlying frame of reference. Our studies have shown that, given a spatial arrangement of objects, it is often the case that multiple frames of reference can simultaneously apply (291 cases in our data). Although it is likely this situation does not cause any ambiguity in interpreting intended objects, about 50% of our cases (146 out of 291 cases) actually lead to ambiguous interpretation of intended objects. The ambiguity caused by frames of reference is quite prevalent in our data.

Our results indicate the property of the object that can be used as a relatum in the environment affects the ambiguity of the underlying frame of reference. Objects without intrinsic sides, when used as relatums, are not likely to lead to ambiguity. Object with intrinsic sides will likely lead to ambiguity between the addressee-centered relative and the object-centered intrinsic frame of reference. While spatial factors (e.g., displacement and orientation) may impose different levels of difficulties (and thus require more efforts) for the speaker to use a specific frame of reference, they are less indicative of the underlying frame of reference. Other contextual cues such as linguistic expressions and strong user preference may help reliable disambiguation.

In human human interaction, the ambiguity of language interpretation can be handled gracefully. Accordingly to the Gricean maxims, if the speaker could foresee the potential ambiguity, he will provide more information to avoid such ambiguity. Such prediction can be generally based on the feedback from the listener, such as back-channel or the gaze direction of the listener. However, in human robot interaction, automated generation of intelligent robot behaviors is still at its infancy. Without good simultaneous feedback from the robot (which is exactly the case in our study), it is difficult for human partners to anticipate potential ambiguities. This points to the future research direction that addresses collaborative behaviors to resolve ambiguities beyond the handling of frames of references.

## References


Carlson, L. A., and Radvansky, G. A. 1996. The influence of functional relations on spatial term selection. *Psychological Science* 56–60.

Coventry, K. R.; Cangelosi, A.; Rajapakse, R.; Bacon, A.; Newstead, S.; Joyce, D.; ; and Richards, L. V. 2005. Spatial prepositions and vague quantifiers: Implementing the functional geometric framework. *Spatial Cognition IV. Reasoning, Action, and Interaction* 98–110.

Eschenbach, C. 2005. *Contextual, functional, and geometric components in the semantics of projective terms*. Oxford University Press, Oxford. chapter 6, 71–91.

Fong, T.; Nourbakhsh, I.; and Dautenhahn, K. 2003. A survey of socially interactive robots. *Robotics and Autonomous Systems* 42(3-4):143 – 166.

Frank, A. U. 1998. Formal models for cognition-taxonomy of spatial location description and frames of reference. *Lecture Notes in Computer Science* 293–312.

Levelt, W. J. M. 1982. Cognitive styles in the use of spatial direction terms. *RJ Jarvella; W. Klein (eds.), Speech, place, and action: studies of deixis and related topics* 251–268.

Levinson, S. C. 1996. Frames of reference and molyneuxs question: Crosslinguistic evidence. *Language and space* 109–169.

Levinson, S. C. 2003. *Space in language and cognition: Explorations in cognitive diversity*. Cambridge Univ Pr.

Mintz, F. E.; Trafton, J. G.; Marsh, E.; and Perzanowski, D. 2004. Choosing frames of referenece: Perspective-taking in a 2d and 3d navigational task. In *Human Factors and Ergonomics Society Annual Meeting Proceedings*, volume 48, 1933–1937. Human Factors and Ergonomics Society.

Moratz, R., and Tenbrink, T. 2006. Spatial reference in linguistic human-robot interaction: Iterative, empirically supported development of a model of projective relations. *Spatial Cognition and Computation* 6(1):63–107.

Moratz, R. 2006. Intuitive linguistic joint object reference in human-robot interaction. In *Proceedings of the Twenty-first National Conference on Artificial Intelligence (AAAI), Boston, MA. AAAI, Menlo Park, CA*.

Petra Weiss, J. G., and Miller, G. A. 1996. Factors affecting spatial deictic communication: A comparison of german and american english. In *Proceedings of the Second Colloquium on Deixis "Time, Space and Identity", March 28-30 1996, Nancy*, 45–73.

Presson, C. C., and Montello, D. R. 1994. Updating after rotational and translational body movements: Coordinate structure of perspective space. *Perception* 23:1447–1447.

Rieser, J. J. 1989. Access to knowledge of spatial structure at novel points of observation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15(6):1157–1165.

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

Schober, M. F. 1996. Addressee-and object-centered frames of reference in spatial descriptions. In *American Association for Artificial Intelligence, Working Notes of the 1996 AAAI Spring Symposium on Cognitive and Computational Models of Spatial Representation*, volume 47, 92–100.

Skubic, M.; Perzanowski, D.; Blisard, S.; Schultz, A.; Adams, W.; Bugajska, M.; and Brock, D. 2004. Spatial language for human-robot dialogs. *IEEE transactions on systems, man and cybernetics* 34:154–167.

Sloane, N. J. A. 2001. A library of orthogonal arrays. *Information Sciences Research Center, AT&T Shannon Labs* 9.

Taylor, H. A., and Tversky, B. 1996. Perspective in spatial descriptions. *Journal of Memory and Language* 35:371–391.

Tenbrink, T.; Maiseyenka, V.; and Moratz, R. 2007. Spatial reference in simulated human-robot interaction involving intrinsically oriented objects. In *Symposium Spatial Reasoning and Communication at AISB*, volume 7.

Tenbrink, T. 2005. Identifying objects on the basis of spatial contrast: An empirical study. In *Spatial Cognition IV. Reasoning, Action, and Interaction*, volume 3343 of *Lecture Notes in Computer Science*, 124–146. Springer Berlin / Heidelberg.

Trafton, J.; Cassimatis, N.; Bugajska, M.; Brock, D.; Mintz, F.; and Schultz, A. 2005. Enabling effective human-robot interaction using perspective-taking in robots. *IEEE Transactions on Systems, Man and Cybernetics, Part A* 35(4):460–470.

Tversky, B.; Lee, P.; and Mainwaring, S. 1999. Why do speakers mix perspectives? *Spatial Cognition and Computation* 1(4):399–412.

Watson, M. E.; Pickering, M. J.; and Branigan, H. P. 2004. Alignment of reference frames in dialogue. In *Cogsci 2004: 26th Annual Meeting of the Cognitive Science Society*, 5–7.