# Identifying Objects on the Basis of Spatial Contrast: An Empirical Study

Thora Tenbrink

Faculty of Linguistics and Literary Sciences, University of Bremen,
Bibliothekstr. 1, 28359 Bremen, Germany
tenbrink@sfbtr8.uni-bremen.de

**Abstract.** In contrast to most research on spatial reference, the scenario in our human-robot experiments focuses on identifying rather than localising objects using spatial language. The relevant question in such a task is "Which" rather than "Where". In order to gain insights about the kind of language to expect in such a scenario, we collected participants' linguistic choices in a web-based empirical study. Spatial scenarios were presented that varied with respect to number, shape, and location of elements, and with respect to possible perspectives. The linguistic analysis reveals that speakers adhere to underlying principles similar to those known for non-spatial object reference. If objects only differ in spatial position, a reference system and spatial axis is chosen that is suitable for contrasting the target object from competing ones. The exact spatial location is usually not specified if there are no competing objects closeby.

## 1 Introduction

Much research on spatial reference focusses on the ways in which human speakers describe the spatial relation between two entities in a given context, answering a question like "Where is the object?" (e.g., the MPIP research reported in Levinson 2003; contributions in Olivier & Gapp 1998; van der Zee & Slack 2003). Spatial reference, however, is not confined to this area. In a different scenario, one of several similar objects needs to be identified on the basis of spatial location, due to the absence of other cues such as distinguishing features or perceptual salience. The relevant question in such a task is "Which" rather than "Where".

Up to now, research in spatial cognition has largely neglected this area. It has, however, extraordinary relevance in a human-robot interaction context in which the artificial communicator is not capable of making fine perceptual distinctions or interpreting pointing gestures (Moratz et al. 2003). In such a context, it is highly advantageous to be able to identify present objects on the basis of their spatial location alone, neglecting material-related differences that might be referred to by humans with more developed – and, importantly, shared – perceptual abilities. Thus, in the scenario we are interested in, there are several similar objects present together with an instructor and the instructee, usually a robot; the latter is instructed to move to one of the objects, specified on the basis of spatial position using projective terms, e.g.: "Move to the box on your right."

C. Freksa et al. (Eds.): Spatial Cognition IV, LNAI 3343, pp. 124 – 146, 2005.

Identifying one of several similar objects in a spatial scenario differs in some basic respects from localising an already agreed upon object by specifying its position relative to another object. For instance, it should involve taking the overall configuration into account to a higher degree than in "Where" scenarios, in which other objects present can usually be ignored (or are simply absent in experimental settings). In contrast, "Which" questions presuppose (shared) knowledge of the situation to such a degree that the target object can be distinguished sufficiently from all other competing objects in the scenario, making the choice of spatial reference highly dependent on the number and arrangement of objects present.

Since the question of how objects are singled out of several candidates using reference to spatial position has largely been disregarded so far, there is a gap in our knowledge with regard to the linguistic variability in such a situation, and with regard to how different situational contexts and conditions influence speakers' choices. In order to fill this gap, we collected spatial instructions of the above-mentioned kind (answering a "Which" question on the basis of spatial location) in a web-based empirical study (presented in this paper) employing native speakers of English.

This study is part of the empirical work carried out in the SFB/TR 8[1] project I1-[OntoSpace] in order to investigate speakers' choices in different scenarios, supplementing various kinds of real-world experiments in human-robot interaction (HRI). On the basis of the empirical results, a flexible spatial ontology is being built in order to be employed in the HRI environment, which itself is continuously being improved by the work of other SFB/TR 8 projects. The study presented here was designed to provide insights about the range of linguistic variation (in English) to expect in a scenario that is comparable to our HRI settings as far as possible, and investigate underlying principles to which speakers may adhere.

### 1.1 Contrastive Spatial Reference

**Spatial Localisation.** A large body of research in the area of spatial cognition is devoted to the identification of spatial reference systems employed by speakers in order to localise an object in a spatial setting (e.g., Carlson 1999; Taylor et al. 1999; Bryant et al. 2000). Projective spatial expressions specifying the *reference (half-)axis* (e.g., *left, right, in front, behind*) together with a *relatum* and a *perspective* are the typical ingredients needed in order to specify a spatial relation. The perspective employed can be based on three kinds of *origin* (Herrmann 1990): speaker, listener, or a further entity (as in: "Coming out of the church, the bus station is on the right"). Likewise, speaker, listener, or a third entity (such as a landmark) available in the current context can serve as relatum. The conflation of point of view (origin) and relatum leads to the kind of reference system called *intrinsic* (e.g., by Levinson 2003), while in a *relative* reference system origin and relatum are distinct. Furthermore, *internal* vs. *external* relationships need to be distinguished (Herskovits 1986), the former specifying object parts where other objects can be located inside of other objects. Note that choice of perspective also plays a role here (cf. Retz-Schmidt 1988), since objects can be

[1] Transregional Collaborative Research Center SFB/TR 8 Spatial Cognition: Reasoning, Action, Interaction (Bremen / Freiburg, Germany), funded by the DFG.

viewed from the outside (such as a picture, where the *left side* depends on the observer's viewpoint) or from the inside (as is usually the case when talking about the internal parts of a person; but it can also be true for objects that can be occupied by a person, such as a car). This difference is often referred to as *handedness* in the literature (e.g., Levinson 2003).

If there are several objects of the same class present, it is furthermore possible to use the whole group of similar objects as a relatum. This option is called *group-based reference* by Moratz (Moratz & Fischer 2000). In contrast to situations where the relatum is an object of a different kind (here called *landmark-based relative reference system*), in a group-based relative reference system the relatum consists of one or several objects of the same kind as the target object.

Now, the question arises whether language directly reflects reference systems. For instance, it has been claimed (Eschenbach 2004) that German spatial adjectives like *link-* (*left*) can only be employed when there is at least one other object of the same class present. However, first results of our spatial human-robot interaction experiments (Tenbrink & Moratz 2003) show that more complex factors come into play. With regard to the external/internal distinction, Herskovits (1986:173f.) presents an overview of the range of linguistic options. Intuitively, for instance, *in the left/right* can only be used for the *interior* of objects, i.e., for internal object parts, while *to the left/right* is only used for external relationships. But in most cases, the language used for different reference systems is more or less the same: e.g., *on the left/right* can be used for external as well as internal relations, and *above* is used for gravity-based as well as intrinsic interpretations (Carlson 1999). In general, relative and intrinsic reference systems cannot be distinguished on the basis of the linguistic form alone. Speakers have to provide additional linguistic material if they wish to specify information about underlying reference systems, e.g., by explicitly mentioning the relatum or the point of view. However, the fact that the same linguistic form can correspond to a variety of underlying reference systems does not necessarily cause ambiguity in interpretation, as will be seen in the analysis below.

In addition to the reference systems in which projective expressions are employed, there are further options of spatial reference available: for instance, speakers can refer to *absolute* reference systems such as the earth's cardinal directions (e.g., *north* and *south*), using different linguistic material. Absolute systems, however, are usually dispreferred in small-scale (indoor) scenarios in Western countries (as opposed to some other cultures, see Levinson 2003). Furthermore, speakers can apply different (non-projective) kinds of spatial reference terms, for example, distance-related expressions such as *close* or *near*.

In order to specify the semantics of projective terms, applicability areas have been identified (e.g., Herskovits 1986; Franklin et al. 1995; Gapp 1995; Carlson-Radvansky & Logan 1997; Carlson et al. 2003). For instance, the expressions *left* and *behind* can be used straightforwardly for 90° and 180° angles, respectively. However, the more the angles between the target object and the relatum depart from these *focal axes*, the more linguistic modifications are used for specifying the spatial relation. Simple expressions are acceptable and applicable in a certain range; outside this range

compounds or modifiers such as *left front* or *a little bit to the left* are more typical (Zimmer et al. 1998). Vorwerg (2001) describes in detail the graded typicality structure of spatial expressions as categories on the basis of psycholinguistic experiments in which participants are asked to point out the "best fitting" expression for a given spatial relation, to place an object on the basis of a spatial description, etc.

**Task Dependency.** Psycholinguistic experiments such as those designed for highlighting the graded structure of projective terms usually do not test for choices speakers make in actual discourse, which naturally depend on the level of specificity needed in a specific discourse situation. If a vague characterization of the spatial relationship is sufficient, applicability areas for simple expressions might be much larger than in a different context where a precise description is vital (as in Zimmer et al. 1998, where the interlocutor needs to find a hidden element on a screen), or where the "best fitting" expression is to be identified (Vorwerg 2001). Thus, an important factor is the *motivation* why a spatial relation should be described at all (in other words, the *question under discussion*). Many different kinds of motivations occur in actual discourse, resulting in diverging usages of spatial expressions. In route descriptions, for example, typically a goal location is described via reference to streets and landmarks which can be easily identified in the real world (e.g., Tversky & Lee 1998). There, salience and dimension of buildings play a role in the choice of landmarks; spatial relations are often sufficiently outlined via simple and vague expressions. Such a scenario differs fundamentally from psycholinguistic spatial localisation experiments where participants are asked to specify an entity's location relative to another.

Descriptions of spatial relations can also be used as a means of identifying – or finding – a target object, such as when a "Where" question is posed in order to identify the correct target object out of several possibilities. Thus, describing a non-prototypical spatial relationship via a simple expression like *left* can be sufficient in a context where there are no competing objects in the left region of the relatum. An everyday example may illustrate this: "Where is the key?" can be answered sufficiently by a vague expression like "Left of the cup" no matter what the precise spatial relationship between the objects is, if there are no other keys closeby. Ultimately, the target object can be identified by establishing sufficient contrast to competing objects, similar to scenarios where "Which" questions are asked. The latter are the target of the present research since they are specifically suited for singling out a target object in contrast to other candidates.

The present research is motivated linguistically, addressing the pragmatics of spatial communication as well as assessing the range of linguistic options for the human-robot interaction scenarios targeted in our research group. In contrast to typical psychological interests and standards we wish to leave as much freedom to our participants as possible in order to learn about their intuitive linguistic strategies (cf. Fischer 2003), since in our project we aim at enabling robots to interpret spontaneous utterances correctly. However, some important basic hypotheses are inspired by previous psycholinguistic studies on (non-spatial) object reference. These findings will be examined next.

**Contrastive Object Reference.** In their seminal work on object reference, Herrmann & Deutsch (1976) formulated principles of *greatest distance, dimension preference,*

*redundant verbalisation,* and *partner-adapted verbalisation.* These principles capture that speakers, in choosing a reference strategy in an object identification task with several different objects, usually do at least the following:

First, they analyse the target object with respect to properties that can establish a (maximum) contrast to competing objects. Thus, if there are two black boxes of different size, the speaker chooses *size* for object reference. In case of several properties in which the objects differ, the speaker chooses the property where the distance to the competing object is most obvious. Thus, if there are two boxes, one of which is very small and dark blue, while the other is very big and black, *size* – rather than *colour* – will be chosen for reference (*greatest distance*). Individual preferences also play a role, especially if the distances are viewed as equal (*dimension preference*).

Second, the speaker encodes as many properties as needed for unambiguous object reference, but usually not more, being economic (cf. Grice's maxim of quantity, Grice 1975). But if the object reference task is complicated by the availability of multiple options, minimal differences in distance, and equal levels of dimension preference, speakers may encode more properties than needed (*redundant verbalisation*).

Third, speakers adapt to their interaction partner's view of the situation, taking into account cognitive and social distinctions, etc. (*partner-adapted verbalisation*).

Herrmann & Deutsch are exclusively concerned with object reference in non-spatial settings, designing their experiments purposively in a way that spatial reference is ruled out. In the present work, the opposite is the case. Target objects differ from competing objects only in spatial position; objects of a different class may also be present, serving as possible relata for reference. Nevertheless, the principles established by Herrmann & Deutsch can be applied to spatial scenarios, motivating hypotheses that can be approached by linguistic analyses of the data collected in the web study (though it cannot be expected, given the kind of open scenario adopted, that they will be verified conclusively). The main research question posed here is:

*What principles of object reference apply when only spatial reference is available?*

Herskovits (1986) noted that, although the graded structure of projective terms applies in most communicative contexts, there are situations in which an expression like *to the right*, without modification, is capable of denoting the full right side (i.e., a half-plane) with respect to the relatum (p. 182): "[T]he loosest interpretation of the preposition is adequate, provided that obvious contrasts in the context allow the expression to fulfill its function of identifying the place of the located object." This applies if there are no competing objects in the same spatial region. In case of the presence of further objects within the half-plane, unmodified projective terms can nevertheless be employed for contrastive reference. In that case, Herskovits' *shifting contrast near principle* applies (p. 81): "If two objects, A and B, are placed in a relation to a reference object in such a way that the ideal meaning of a preposition (…) is truer of A than of B, then one can use that preposition to discriminate A from B so that the locative phrase will be assumed true of A but not of B." For instance, if *to the left* is "truer" of A than of B, i.e., A is closer to the left reference axis than B is, A will be recognized as the target object even if *to the left* could also be applied to B.

Starting from these observations more concrete research questions can be formulated as follows.

*How do speakers choose a reference system and point of view?* If a partner is present it can be expected that participants will often choose their partner's perspective, especially if the partner is expected to act (Herrmann & Grabowski 1994:123), and that speakers will adapt their utterances to their interlocutor's in various respects (Clark 1996; Pickering & Garrod in press), for example with respect to the choice of reference systems (Watson et al. 2004). These findings further specify Herrmann and Deutsch's above principle of partner-adapted verbalisation. But the literature does not provide much evidence with respect to which reference systems are preferred if several options are available in an object identification scenario. This is so in part because such scenarios have not been in focus very often in spatial cognition research, and specifically, group-based reference has hardly been mentioned in the literature so far at all. Likewise, little can be said for the case of several options for perspective when there is no interaction partner. The present hypothesis is that the identification task – requiring reference on the basis of spatial contrast – plays a role in the decision, since situations may arise in which one kind of reference system or perspective enables a clearer contrast than the other ones available. On the grounds of Herrmann & Deutsch's principles, this means that, *in addition to adapting to their interlocutor, speakers choose a reference system and perspective that is suitable* (just as a unique object property is) *for distinguishing the target object from competing ones*.

*How do speakers choose a reference axis?* Within a reference system, the frontal (*front/back*), the lateral (*left/right*), and the vertical (*above/below*) axes are available for reference. With "Where" questions, the reference axis is chosen that the target object is closest to. But in contrastive reference, this may not yield unambiguous reference. It is hypothesized here that competing objects play a role in deciding about a reference axis: *A reference axis is chosen that is best suitable for distinguishing the target object from competing ones, considering the principle of greatest distance*.

*How explicit are speakers about underlying reference systems and origins?* The situation may offer various reference systems yielding similar results (e.g., Carlson 1999): in some situations, *to the left* could equally well be used for an intrinsic reference system conflating origin and relatum, for a group-based reference system using the other objects as relatum, and for a landmark-based relative reference system. In such situations, *to the left* can simply be used without further specification since no conflicts arise. Similar observations apply with respect to the chosen point of view (origin); if several options are available yielding no difference people do not need to provide an expression like "from my point of view". In intrinsic reference systems, the origin is often specified because it coincides with the relatum, as in "in front of me". The interesting case is when different reference systems and points of view yield different results, so that *to the left* can be interpreted in different ways. Herrmann & Grabowski (1994:132) state that speakers are usually *not* explicit with regard to the perspective used. This may be because they tacitly assume that the interlocutor will understand the intended meaning even without the additional effort, since perspective needs to be expressed by additional linguistic material and since the partner's perspective is (in many situations) conventionally preferred. But with respect to reference systems, no such conventions are known so far. Thus, it is hy-

pothesized here that *relata – but not necessarily origins – will be made explicit in case of potential conflict*.

*Under what circumstances do speakers modify and combine projective terms?* With "Where" questions, speakers increasingly use modifiers and compounds as distance to the reference axis increases. According to Herskovits (1986) this is not expected in identification tasks (see above). Likewise, Herrmann & Deutsch's principles predict that speakers will not provide more information than needed, unless several options with equal properties compete. Thus, it is hypothesized that *speakers use a projective term without modification or combination with another projective term in case there are no competing objects for which the same description applies to the same degree*. If the target object is placed where it could equally well, and equally unambiguously, be referred to by two terms, such as *to the left* and *in front*, both may be combined.

## 2 Empirical Study

In order to address the above research questions along with the more basic aim of assessing the linguistic range of variety in settings involving contrastive spatial object reference, we collected linguistic data in a web-based empirical study. A major advantage of this approach is that large amounts of native speaker data can be collected with very little effort. On the other hand, since no clarification questions can be asked, there is a higher potential for misunderstandings. Furthermore, participants can vary factors about their participation, such as distraction, pauses between tasks, advice by other people, etc., which are controllable only in a setting involving the co-presence of experimenter and participant. They may also answer untruthfully to questions about their person regarding age, gender, language skills, etc, and they may re-start and participate several times. Further evaluation of advantages and limitations of web experiments can be found in Reips (2002). At the present stage, the advantages outweigh the disadvantages especially in light of the fact that most of the uncontrollable aspects are not considered to be central influencing factors. Nevertheless, in subsequent studies a higher degree of control is desirable, which presumably entails the participation of a smaller number of speakers.

Since the web-based study was carried out in order to establish a corpus of natural language data, in the analysis conventional methods of corpus linguistics are employed. The main focus is on the identification and description of qualitative structures in data collected in an open setting, supported by relative frequencies of usage. No statistical measures are computed because of the open setting and the large amounts of variation and potential interdependencies that render pairwise comparisons statistically less reliable. The present aim is to point to a number of systematic patterns that can be subsequently validated by more controlled experimentation.

### 2.1 Method

Speakers of English were asked via mailing lists and personal communication to participate in a web-based empirical study accessible at www.language-experiments.org

between September 23rd and December 31st, 2003. Participation was voluntary and not paid for. Altogether, approximately 200 self-assessed native speakers of English[2] participated. Their contributions show that there were very few problems in interpreting the instructions. Since the qualitative analysis was non-automatic and therefore time-consuming, not all of the data could be analysed exhaustively. Therefore, a target number[3] (60) of utterances for each single analysis was chosen that was considered sufficient for a fairly broad and informative exploration of the range of variety and underlying principles of speakers' choices[4]. The contributions of native English speakers were extracted at random out of the collected pool of data, annotated using the text markup tool *Systemic Coder*[5] version 4.5, and analysed linguistically. For each situation, the preferred linguistic options were identified and analysed with respect to the above research questions. Furthermore, differences between situations were examined by comparing frequencies of linguistic categories.

Each of the participants answered 15 different randomly assigned questions[6] in randomized order[7] out of a pool of 21 possible tasks which cover a range of different scenarios. A selection of the 21 tasks is analysed in this paper in order to examine the impact of specific changes in the scenario. They belong to four conditions that differ with respect to the possible perspectives on the configurations. Randomizations of tasks are only inside conditions, treating the conditions 3 and 4 as one category. In each condition, the same situations are shown (see Fig. 1 below). In the simplest scenario, S (situation) 1, three identical squares are located in a row, enabling unproblematic group-based reference. S2 shows the same scene except that the middle element is not a square but a circular element[8], providing a further option for the choice of a relatum, i.e., an element of a different kind. S3 presents only two elements in a spatial relationship that does not correspond to any of the focal axes in any kind of reference system. Thus, linguistic modifications can be expected here. Furthermore, especially in the conditions that include a viewer, it is hypothesized that the option of a group-based reference system is not available. Thus, the three configurations offer a range of spatial relationships that can be conceptualised and referred to in various

[2] Since also non-native speakers were asked to participate the overall number of participants was much higher. Age was asked for in a questionnaire but not provided by many participants; those who stated their age were predominantly between 15 and 50 years old, which corresponds to the target of the experimental design. Since there are no hypotheses regarding the impact of age in adult speakers, age effects and distribution are not considered further.

[3] As will be seen below, this number was deviated from in the analyses of S2(C4) and S2(C3).

[4] Since not all participants completed all tasks, the tasks were partly completed by the same persons and partly by different ones. In treating the data as a corpus this unfortunate circumstance is regarded as marginal for the present kind of analysis.

[5] Available freely at http://www.wagsoft.com/Coder/index.html.

[6] The decision to limit the number of tasks for each participant to 15 was taken in order to minimise the time and effort required for participation.

[7] Effects of order were also not computed in the present analysis, since they are regarded as a non-trivial additional factor that needs to be treated with specific care. Randomization guarantees a fairly even distribution of task positions.

[8] This term is here used by way of contrast to the marking of the goal element by a circular line. The participants naturally used *circle* to refer to the circular element.

ways. This variability is further enhanced by the options of perspective, which are varied by the four conditions as follows.

**Outside View.** In the first condition, participants are presented with pictures that only contain squares and circular elements in the three configurations as depicted in Fig. 1, but without the X. In each picture, one of the elements is marked by a circle. The question to be answered by the participants is simply, "Which element of the picture is marked with a circle?" In this condition, the only available view direction is provided by the fact that the participants look at a picture on a screen (outside perspective). Concerning reference frames, one option is to use the picture itself as a relatum, applying the projective terms for the internal parts (regions) of the picture. Another option is to use a relative reference system employing either (some of) the other objects as a group relatum, or the circular element (if present) as a landmark relatum.

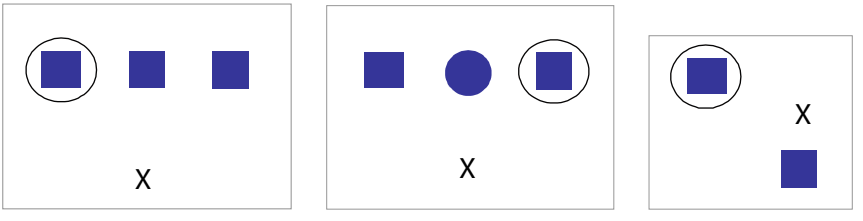


**Fig. 1.** Three situations S1, S2, and S3 (Condition 2: Inside view)

**Inside View.** In the second condition, an X appears in the picture in addition to the elements (Fig. 1). The instruction is, "Now imagine that you are looking at the figures from the position marked X. How do you describe now which element is marked with a circle?" This condition departs from the previous situation in that the participants are asked to imagine themselves *inside* the picture. Thus, the origin in their reference frames is supposed to be X directed towards the objects, but the outside perspective is still available. The reference frames available are those known from condition 1 plus intrinsic or relative reference systems specifying the target object's position relative to X, either from the imagined view direction of X or from the outside perspective.

**Partner View.** The third and fourth condition were designed to simulate a real world setting as much as possible, in order to represent the human-robot interaction settings that we use in our project. Now, the position of an interaction partner, Y, is added in the pictures. Additionally, both X and Y are assigned a clear view direction. In condition 3, X has the same position as in the inside view condition, and Y is positioned at a 90° angle with respect to view direction. In condition 4, the positions of X and Y are reversed. In each case, the participants read "Finally, please imagine that the figures are real world objects. You are located at X, and now your task is to instruct person Y to go to the object marked with a circle. A star ✸ shows the direction each of you is facing in." Figures 2 and 3 show the two pictures of these two conditions that are analysed in the present paper.

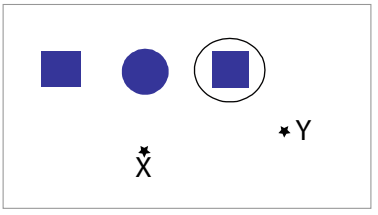


**Fig. 2.** S2 in cond. 3, partner view

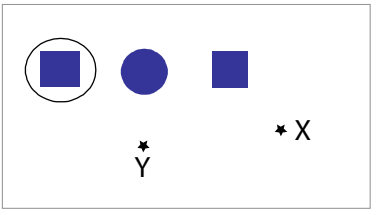


**Fig. 3.** S2 in cond. 4, reversed partner view

Thus, in these conditions view directions are given explicitly, and the participants are asked to imagine a dialogue situation. However, since there is no real interaction and no feedback from the interaction partner, grounding and alignment processes such as those described in Clark (1996) and Pickering & Garrod (in press) are ruled out, similar to imagined-partner experimentation as reported in Herrmann & Grabowski (1994). The instructional task differs slightly from the previous ones. This situation requires a lot of imagination by the participants; therefore, tasks in condition 3 and 4 are presented only after the first two conditions. Now, in addition to the origins and reference frames available in the previous conditions, there is the option of using Y as origin or relatum, or both.

### 2.2 Results

In a first examination of the data, 12 main linguistic categories were identified that occur frequently across contexts, covering the variability in utterances that refer to the goal object. These categories were used for a detailed analysis in order to examine the question how speakers choose reference terms in a variety of situations. The following list provides an overview of the categories together with the abbreviated reference term used in later sections. Note that the categories 1-11 always presuppose the presence of projective terms, which are the main focus of the present analysis, while 12 and 13 do not.

1. The projective term is a noun in a prepositional phrase and is neither modified nor further specified by providing information on the relatum, as in "The square on the right": Unmodified noun-in-pp
2. As before except that now the relatum is specified, either by another prepositional phrase, as in "to the right of the circle", or by a possessive, as in "to my right": Unmodified noun-in-pp plus relatum
3. The projective term is an unmodified adjective, stand-alone expression, preposition, or adverb (excepting handedness adjectives), e.g., "right (square)" or "the one above": Unmodified adjective / prep / adv
4. The projective term is an unmodified handedness adjective, as in "right-hand square": Handedness adjective
5. The projective term is a superlative (rightmost square): Projective superlative
6. Two projective terms of any kind are combined, as in "top left", "upper left", or "at the top left": Combination

7. The projective term is modified by a term denoting distance, as in "The one furthest to the right" or "The square on the far right": Modification-distance
8. The projective term is a noun in a prepositional phrase and is modified by counting[9] ("The third square on the right"): Modification-counting
9. The projective term is a noun in a prepositional phrase and is modified by a precisifying adverbial ("the square just to my right"): Modification-precisification
10. The projective term is part of a complex description combining several kinds of modification, as in "Go to the cube just in front of you to the right, before the sphere": Complex-description
11. A comparative relating to height is used, such as "the higher square": Comparative
12. A non-projective term is used that relates to distance, either as comparative ("the farther square") or as superlative ("farthest"): Distance-related
13. Other (mostly other kinds of non-projective expressions).

The distinctions are made on the basis of grammatical differences, principally induced by various kinds of modifications of the projective terms. Category 3 subsumes some forms because they occur only infrequently and in similar situations. This finding corresponds to the assumption that the different syntactic forms of projective terms do not necessarily reflect semantic differences (cf. Miller & Johnson-Laird 1976, Coventry & Garrod 2004). Nouns in prepositional phrases (cat. 1) are treated separately for three reasons: First, in the present data they occur frequently and in different situations than the occurrences subsumed in 3. Second, they involve a more complex linguistic construction than those in 3, containing a preposition that may itself contribute semantic content. Third, they can take an additional prepositional phrase or a possessive that specifies the relatum. This option is distinctly categorised as 2. Prepositions and adverbs such as *above* can also occur in constructions that inform about the relatum, but because they occur only very infrequently in the data this option is not assigned a distinct category. Handedness adjectives are treated separately because they appear fairly frequently in some configurations, and they may induce a semantic difference by implying a human-like origin.

The utterances are also analysed with respect to the choice of axes, perspectives and relata. However, as pointed out above it is assumed here that there is no one-to-one correspondence between linguistic surface forms and reference systems. Since utterances are in many cases underdetermined with respect to the underlying perspective and the intended relatum (as will be exemplified in some detail below), these two categories are coded with respect to all possibilities that are recognized to be compatible with the task. Since, as the present data show, participants are often creative in their conceptualisations of the spatial scenes, it is even possible that additional interpretations would yield the same results, so that the percentages listed should be regarded as tentative, only reflecting a lower limit. Note that this kind of underspecification usually does not result in ambiguity in referential identification of the target element, since this involves different interpretation processes than identification of the reference system or perspective. - Choice of axis and perspective do not

[9] This is exceptional since a short general introduction given to the participants before starting with the pictures explicitly stated that counting should be avoided.

apply for distance-based descriptions, but also there, a relatum may or may not be given explicitly.

The following situations are analysed: Condition 1 (C1): S1, S2, S3; Condition 2: S2 and S3; Condition 3: S2; Condition 4: S2. While the pool of collected data contains many more situations, this selection enables the systematic comparison of linguistic choices in situations that differ with respect to only one factor.

**Overview of Results.** Table 1 below summarizes the results of the linguistic analyses, providing the percentage of frequency of the various linguistic categories occurring in each situation under analysis. N is the analysed number of utterances containing projective terms in each situation. In cond. 1 and 2, 60 utterances were examined in each task under analysis, very few of which do not specify the target object directly by projective terms. In S2(C4) and S2(C3), in contrast, a high percentage of contributions do not refer directly to the goal element. These are not represented in the table but treated separately below.

**Table 1.** Variability in goal-related utterances (% in each task)

| Task | S1(C1) (N=60) | S2(C1) (N=60) | S3(C1) (N=60) | S2(C2) (N=60) | S3(C2) (N=59) | S2(C4) (N=51) | S2(C3) (N=57) |
|---|---|---|---|---|---|---|---|
| **Linguistic category** | | | | | | | |
| **unmodified noun-in-pp** | 20.0 | 28.3 | 10.0 | 28.3 | 13.6 | 11.8 | - |
| **unm. noun-in-pp + relat** | - | 8.3 | - | 16.7 | 22.0 | 56.9 | 5.3 |
| **unmod. adj., prep./ adv.** | 11.7 | 13.3 | 35.0 | 16.6 | 22.0 | 2.0 | - |
| **handedness adj.** | 6.7 | 10.0 | - | 6.7 | - | - | - |
| **proj. superl.** | 35.0 | 23.3 | 6.7 | 15.0 | 1.7 | 9.8 | 3.5 |
| **combination** | - | - | 40.0 | 13.3 | 20.3 | 2.0 | - |
| **mod.-dist.** | 21.6 | 15.0 | 1.7 | - | - | 11.8 | 3.5 |
| **mod.-counting** | 1.7 | - | - | - | - | - | 7.0 |
| **mod.- precisification** | - | - | - | - | 1.7 | - | 8.8 |
| **complex descr.** | - | - | - | - | - | 2.0 | 14.0 |
| **comparative** | - | - | 5.0 | - | - | - | - |
| **distance** | - | - | - | - | 13.6 | 3.9 | 57.9 |
| **other** | 3.3 | 1.7 | 1.7 | 3.4 | 5.1 | 2.0 | - |
| **Perspective** | | | | | | | |
| **outside view** | 96.7 | 100 | 100 | 76.7 | 18.6 | 33.3 | - |
| **X's persp** | - | - | - | 100 | 83.0 | - | - |
| **Y's persp** | - | - | - | - | - | 96.0 | 42.1 |
| **explicit** | - | 1.7 | - | 23.3 | 40.7 | 62.7 | 36.8 |

**Table 1.** (*Continued…*)

| Task | S1(C1) (N=60) | S2(C1) (N=60) | S3(C1) (N=60) | S2(C2) (N=60) | S3(C2) (N=59) | S2(C4) (N=51) | S2(C3) (N=57) |
|---|---|---|---|---|---|---|---|
| **Spatial Axis** | | | | | | | |
| **only lateral** | 96.7 | 100 | 11.7 | 83.3 | 54.2 | 90.2 | 35.1 |
| **only frontal** | - | - | - | - | 1.7 | - | 1.8 |
| **only vertical** | - | - | 46.7 | 1.7 | 6.8 | - | - |
| **lat. & front.** | - | - | - | 10.0 | 6.8 | 2.0 | 5.3 |
| **vert. & lat.** | - | - | 38.3 | 3.3 | 10.2 | 2.0 | - |
| **compass dir.** | - | - | 1.7 | 1.7 | 3.4 | 2.0 | - |
| **Relatum** | | | | | | | |
| **int. regions** | 96.7 | 88.3 | 98.3 | 66.7 | 13.6 | 25.5 | 1.8 |
| **group** | 96.7 | 88.3 | 88.3 | 66.7 | 13.6 | 27.5 | 1.8 |
| **landmark**[10] | - | 66.7 | - | 46.7 | - | 15.7 | 1.8 |
| **X** | - | - | - | 98.3 | 96.6 | 5.9 | 5.3 |
| **Y** | - | - | - | - | - | 86.3 | 94.7 |
| **explicit** | - | 11.7 | 6.7 | 25.0 | 49.2 | 72.5 | 83.9 |

In conditions 3 and 4, the perspective of Y (the "interlocutor") is consistently used, never that of X. This results in S2(C4) being more similar to S2(C2) than S2(C3) is, with regard to the origin's position. Therefore their order is reversed in the analysis. In the following, each of the above research questions is addressed separately.

**Choice and Explicitness of Point of View.** In condition 1, all projective expressions are based on the participant's outside view on the picture (the only available option), which is made explicit only once through all three situations examined. Note that the linguistic construction used for making perspective explicit is fairly complex, which may be a reason why it is the only such specification: "If I am viewing the picture as if it were a picture hanging on a wall, the answer is the left most square".

In S2(C2), all utterances are consistent with the perspective of X (as requested in the instruction to the participants), although the outside perspective is also available. 21.7% explicitly use this perspective by specifying X as the relatum, using an intrinsic reference system (e.g., "the one to my right"), where the relatum is identical to the origin. The other utterances are consistent both with the perspective of X and with the outside view, so that it cannot be decided which perspective is actually used. This does not, however, impede interpretation because none of the other elements could be referred to in this way in the present situation.

The second situation in condition 2, S3(C2), poses a challenge for the participants because the view direction is less clear than in the other scenarios, where all objects are placed together at one side of X. Here, in contrast, "looking at the elements" (as in the instruction) leaves some freedom for interpretation. The participants are clearly aware of this problem and therefore develop interesting strategies. One explicitly states that it is impossible to give an answer since the view direction of X is unknown;

[10] Utterances like "to the left" are coded as consistent with a landmark-based reference system because they could be expanded to "… of the circle", while "lefthand" and "leftmost" are not.

therefore, only 59 utterances are analysed. In contrast to the other tasks in conditions 1 and 2, in S3(C2) as many as 10 utterances (16.9%) are not projective; 8 of these rely solely on distance, using expressions such as "farther" or "furthest from me", which are independent of the view direction. This strategy circumvents the perspective problem. Another strategy is to avoid ambiguity via choice of axis (see below).

In conditions 3 and 4, all utterances containing projective terms are consistent with the perspective of Y, though in S2(C4) this view direction coincides with the outside view as in S2(C2) above. Therefore, in one-third of these cases perspective cannot be determined although the descriptions are not ambiguous. In the other cases, perspective is specified by mention of the relatum, as in "Go to the square on your left".

**Choice and Explicitness of Reference Systems.** A frequent linguistic option is the usage of a projective term as an unmodified noun in a prepositional phrase, such as "the one on the right". Here, the underlying reference system is underdetermined if the relatum is not made explicit, which is mostly the case in tasks S1(C1), S2(C1), and S2(C2). It could then be either group-based relative (in case the utterance was continued as "… of the other elements"), landmark-based relative ("… of the circle"), picture-based ("… of the picture") or intrinsic ("… of X (or Y)"). If the underlying reference system does not yield any difference in interpretation, further explication is not necessary, and accordingly does not occur at all in S1(C1). In the similar situations S2(C1,C2,C4) the relatum is mentioned increasingly often with increasing alternative possibilities. In S2(C4), as many as 72.5% explicitly use Y as relatum, as compared to 11.7% in S2(C1) and 25.0% in S2(C2). This difference is explained through the fact that there is a true alternative interpretation in S2(C4), namely X as relatum and origin, while in S2(C1,C2) all other options do not yield a different interpretation.

The circular element (landmark) is explicitly used for reference in 11.7% of utterances in S2(C1), but only in 1.7% of all utterances in S2(C2), 5.9% in S2(C4), and 1.8% in S2(C3). This leads to the hypothesis that objects of a different kind are not necessarily favoured as relata, depending on the situation.

In S3(C1), as in S1(C1), differentiating between picture-internal and group-based reference systems is not necessary, since both interpretations yield the same results. Accordingly, they are not specified except for a few utterances that refer to the picture's corner. In S3(C2), in contrast to the other tasks analysed so far, the different available reference systems yield different spatial regions. Here, the relatum is specified in half of the utterances, assisting interpretation.

**Choice of Reference Axis.** In S1/S2(C1), all utterances rely on the lateral axis. In S2(C2), 83.3%, and in S2(C4), 90.2% of all utterances use the lateral axis as the single axis for reference. This clear result is not surprising since the second axis is only available when using the X (in C2) or Y (in C4) as a relatum, while in group-based reference or using the picture's internal regions only one axis is available. This axis coincides with the axis that is most suited to distinguish between the objects also when X (or Y) is used as relatum, since in that case, all three elements are located *above* or *in front of* the relatum, but only one (the target object) is located *on the right*.

Thus, it is surprising that one participant wrote "the element above", since this is not informative (i.e., unambiguous) in this situation.

In S3(C1) where the target object is located diagonally between the focal axes, interestingly more than half (59.3%) of the projective utterances rely on one single axis for reference. In these cases, the axis is obviously chosen for establishing a contrast to the other element in the picture, which is not located on the same half-plane, instead of describing the exact spatial location. Whenever only one axis is used there is a clear preference for the vertical axis (80% of utterances using only one axis, as compared to 20% that use only the lateral one). This can be explained by way of Herrmann & Deutsch's principle of greatest distance: The target object is located just a trifle more clearly in the upper part of the picture, and above the other object, than it is on the left. Another explanation is that the vertical axis is more salient than the lateral, as has been claimed time and again in the literature (e.g., Bryant et al. 2000).

In S3(C2), there is a high heterogeneity in participants' choices even though the situation contains only two elements. Nevertheless, some systematic preferences can be identified: if only one axis is used, the target element is preferably referred to as "right" (44.1% as opposed to 1.7% referring to the "front") even though it is not situated close to the right focal axis with respect to X (and not at all when using the outside view). Rather, it is situated between the right and front axes of X, even closer to the front axis than to the right one. But, unlike "to the right", "in front" could, in a different interpretation, also refer to the competing object. Furthermore, since the competing object is located clearly *on the left* with respect to the lateral axis, but only *in the middle* with respect to the frontal one, the lateral axis provides a clearer contrast. Thus, as hypothesized, reference systems and spatial axes are preferred that allow for unambiguous reference in the given situation.

**Modifications and Combinations of Projective Terms.** Projective superlatives, e.g., "the leftmost element", can be regarded as explicit linguistic constructions with regard to contrast, since they indicate that one specific element can be singled out that is located farther in the indicated direction than any other elements. Not surprisingly, this is the most preferred option in S1(C1). In the other tasks where the elements are lined up in a row as in S1(C1), namely, S2 in all four conditions, projective superlatives are still present but the frequency decreases with each additional factor in favour of other options. In S3(C1) where there are only two elements the projective superlative is very infrequent. One utterance here states "the leftmost (leftmore?) square is", thus providing the analyst with a reason why the superlative is not used here more often: it does not seem to be appropriate in a context where there is only one competing object. Since, unlike other superlatives, there is no expression like *leftmore* in English, the participants prefer other linguistic constructions in this scenario. Note that projective superlatives do not occur at all in the literature on other kinds of elicitation of spatial expressions in which applicability areas are to be identified.

Another frequent method of establishing spatial contrast is the usage of linguistic modifiers such as "furthest (to the right)" (mod.-distance). This is especially frequent in S1(C1). The presence of a distance modifier linguistically enforces a contrast to other objects present, allowing, for example, for other elements to be "not-far to the right" (e.g., to the right but not as far as another element).

Complex descriptions combining several kinds of linguistic modifications and combinations, as in "Walk to the square object closest to you, slightly on your right", providing a fairly precise spatial description, occur only in the partner conditions, most frequently in S2(C3). Similarly, modifiers precisifying the projective term as in "Go to the square immediately in front of you" occur only in S2(C3), although also in many other cases the elements are not situated near focal axes, which could lead to the expectation that precisifying adverbials and complex descriptions should occur with much higher frequency. However, only in S2(C3) is there a need to specify the position since there are competing elements located in a similar direction. Here, the goal object is situated close to Y, enabling simple description in terms of distance, as is done by almost one third of participants. However, those that do not choose this non-projective kind of description obviously run into problems in finding a suitable description. A simple projective description that establishes sufficient contrast to competing objects does not seem to be available here.

In the other cases where the elements appear in a row, the target element is located directly on the lateral axis only in the group-based and picture-internal cases but not in the case of using X or Y for intrinsic reference in S2(C2) and S2(C4). But intrinsic reference is indeed very frequent at least in S2(C4), as indicated by the fact that in 60.8% of all goal-related utterances the relatum Y is explicitly mentioned, as in "to your left". Thus, a vague spatial description is deemed sufficient by speakers if there are no competing objects closeby. Furthermore, since modifications and combinations of projective terms are rare even in the highly problematic situation S3(C2), it can be hypothesized that the strategy of choosing a reference axis that enables unambiguous identification may in some situations have stronger consequences than the graded semantic structure of projective terms.

In the following, some contrasts between linguistic choices are highlighted by comparing situations that differ with regard to one factor.

**Comparison of S1(C1) and S2(C1).** These two tasks differ only in the presence of an object of a different kind in S2(C1). Interestingly, the usage of modifications including superlatives seems to be lower when a landmark is present (38.3% in S2(C1) vs. 58.3% in S1(C1)). This can be interpreted as indicating a lesser need for explicitly establishing contrast. Since the middle element in this case is of a different kind, it does not appear as a competing candidate. The only other element of the same kind is positioned at some distance, so that unmodified projective terms are more often deemed sufficient. Furthermore, participants probably refrain from employing superlatives because there is only one other object of the same class present, as in S3(C1).

**Comparison of S1(C1) and S3(C1).** In S1(C1), there is one clear reference axis available which is used throughout. Modifications concern position on that axis, using superlatives and modifiers denoting distance. In S3(C1), in contrast, two reference axes are equally suitable. Here, modifications concern combinations of projective terms as well as comparatives. However, unmodified adjectives are used much more frequently than in S1(C1). In S1(C1), the preferred option is the use of projective superlatives which, in turn, are clearly disregarded in S3(C1). Furthermore, in S3(C1) only one participant refers to distance, as opposed to 21.6% in S1(C1).

Figure 4 contrasts the linguistic options used in the three situations in condition 1.

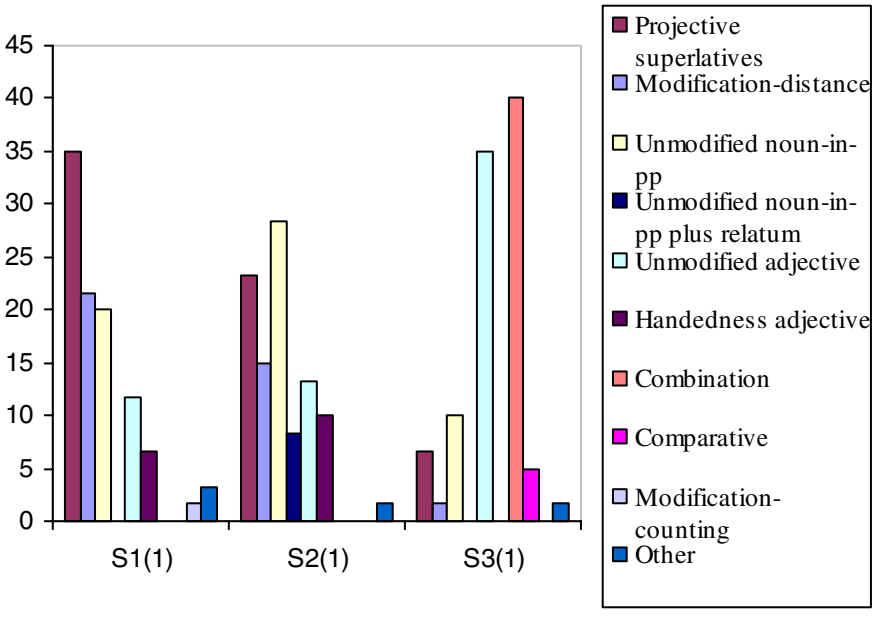


**Fig. 4.** The impact of the spatial configuration (without variation of perspective)

**Comparison of S3 in Condition 1 and 2.** The linguistic choices speakers make in S3(C2) differ considerably from those in S3(C1) although the pictures differ only with respect to the available origins and relata. Interesting differences occur, for example, with respect to the employment of a second projective term (40.0% in S3(C1) vs. 20.3% in S3(C2)), and the usage of distance-related terms (none in S3(C1) vs. 13.6% in S3(C2)). These differences point to the fact that the presence of X has a considerable impact on the conceptualisation of the situation, confirming that the participants indeed view the situations as different in spite of the fact that the outside view on the picture is still available in S3(C2), though participants are encouraged to use the perspective of X.

**Condition 3 and 4.** In these conditions, in addition to imagining themselves inside the picture at position X, participants are asked to instruct a person located at Y to go to one of the objects. To achieve this task, many participants do not refer directly to the goal object. The design of the study allows for free choice of instructional strategy, because the instruction to the participants only informs about the imagined persons' position and view direction, but not, for example, about the absolute scale, distances, and perceptibility of goal objects, or any other clues that could influence the choice of strategy. Therefore, I do not view utterances that are not goal-based as misinterpretations of the task, but rather as an interesting result with respect to how participants conceptualise the situation. In the present analysis (cf. Table 1), primarily

goal-based utterances are targeted. However, some interesting observations can be noted with respect to the other kinds of strategies employed in conditions 3 and 4.

In S2(C4), of a total of 114 collected utterances instructing an imagined person Y to move to a goal object, only 43.0% use projective expressions referring directly to the goal, and 1.8% use distance expressions. These are the utterances that appear in Table 1 above. Another 30.7% describe path directions like left/right or angles, and 23.7% first describe the path and then refer to the goal object. In S2(C3), of a total of 125 instructions, 19.2% use projective terms pointing directly to the goal object, 25.6% rely on distance, 40.0% describe path directions like left/right or angles, and 12.0% describe the path and then refer to the goal object. This variability of instructional strategies is similar to that found in our human-robot interaction scenarios (cf. Fischer & Moratz 2001). It is depicted in Figure 5. Note that, in comparison, the utterances in the other two conditions rely almost exclusively on projective expressions, with occasional usages of distance expressions in certain situations.

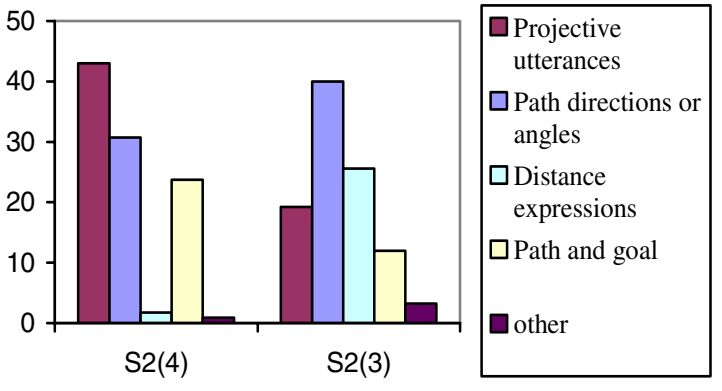


**Fig. 5.** Strategies in two tasks in conditions 3 and 4 differing only in perspective

The clear difference in usage of distance terms is explained through the proximity between Y and the goal object in S2(C3), especially because two competing objects are situated in the same spatial area as the target object. This option is not equally available in S2(C4) from Y's point of view, since another object is at equal distance.

## 3 Discussion

One of the primary aims of the present study was to investigate the variability with regard to linguistic choices in scenarios involving contrastive spatial object reference. Clearly, participants use a broad spectrum of variability, resulting in at least twelve distinct linguistic categories which could still be differentiated further (see Table 1 above). The data collected in the present study show that linguistic choices depend heavily on the spatial situation, i.e., the presence of other objects and the available perspective. Therefore, generalised predictions are difficult to formulate on a linguis-

tic surface level. A general result, however, is the finding that *instructional strategy* changes with condition: People evidently approach a task differently when asked to instruct someone to move somewhere, than when asked to single out one target object of several competing ones. But even here, clear differences appear between spatial scenarios depending on the observer's position.

Not all of the available linguistic options are used in all situations. In spite of the high variability, however, regular patterns of usage can be identified that can be analysed in relation to the hypotheses posed in section 0 above. The systematic variations found are repeatedly explained by much the same idea, reflecting the participants' motivation to fulfill the task of providing an unambiguous – contrastive – description of the target object in a given situation. In other words, speakers intuitively (but not stringently) adhere to systematic underlying principles that could, in a first approximation, be formulated as follows.

1. A reference system is chosen that allows for unambiguous reference, i.e., that produces at least one spatial region in which an unmodified projective term can be used unambiguously (see 2), if possible at all (otherwise see 5).
2. Unmodified projective terms can be used unambiguously in the following cases, and are therefore employed there more often than in other cases:
   - if the goal object is the only one on a half-plane with respect to the reference system used (regardless of whether it is located near the prototypical axis corresponding to the projective term used or not);
   - if competing objects also situated on the same half-plane are clearly farther away from the axis than the goal object is.
3. If 2 applies for more than one spatial region, the axis is chosen that the goal object is closest to (unless 4 applies), and an unmodified projective term is used.
4. If 2 applies for more than one spatial region and the goal object is located at equal distance from two axes (e.g., left and front) then either both projective terms are combined, or one is chosen at random or via individual preferences.
5. If all regions are occupied by more than one object, modifications, combinations of projective terms, counting and/or distance expressions are used. In such a case, distance expressions are specifically likely if the goal object is clearly either the closest one or the one that is farthest away from a suitable relatum.
6. Relata are preferably mentioned explicitly in case of conflict in interpretation; otherwise they usually remain implicit. Origins are seldom given except when conflated with relata in intrinsic reference systems.

Largely, these findings are in accord with previous results of research on object reference, where it is known that speakers analyse contrast on any dimension available, with respect to the objects present in a given scenario, and use it in order to achieve unambiguous reference. In psycholinguistic studies, as well as in the present web study, the available contrastive dimensions as well as the reference area are usually clearly delimited by the experimental setting. Real world scenarios involve far greater complexity. For example, objects often differ with respect to both spatial position and non-spatial kinds of features, where under certain circumstances reference on the basis of spatial position seems to be preferred (Pobel et al. 1988). However, the

factors influencing such choices and the variability with respect to related options seem to be largely unexplored. For example, in what kinds of situations do speakers deem a simple class name as sufficient for reference? Conceivably, this may be the case if a given configuration allows for unambiguous object reference via the class name. But this hypothesis does not only presuppose simplistic referability on the basis of a distinctive object name, but also a clear delimitation of the reference area (among further factors, see Freksa 1980). In real world scenarios, this may be a crucial factor influencing speakers' choices, especially in an open setting where the referential domain may be unclear. In a human-robot interaction scenario, for example, the speaker may not be informed about the limits of the robot's perception (Fischer & Moratz 2001). Brown-Schmidt & Tanenhaus (2003) show that both the form of referring expressions and their interpretation are constrained by previous specification of referential domains. Furthermore, attention focus plays a crucial role (e.g., Kessler et al. 1999). In ambiguous situations, reference is resolved with respect to a subset in focus. Attention can be directed by the speaker through focus and foregrounding on a linguistic level, influencing the listener's interpretation of spatial descriptions. The applicability of spatial terms is further influenced by functional features of the objects involved, and the relationship between them (e.g., Coventry & Garrod 2004). In natural communication, interactive processes facilitate the achievement of joint reference (Clark 1996, Pickering & Garrod in press), and previous experience (i.e., the discourse history) may influence later choices, for example in the employment of reference systems (Tenbrink & Moratz 2003). All of these major influencing factors need to be accounted for when dealing with natural conversation, in contrast to the simple scenario presented here.

Due to the open design of the study a considerably broad range of variety, and systematic patterns of choices, could nevertheless be identified, since speakers' linguistic behaviour was not governed by instructions to a high degree. The procedure adopted in the present work is in accord with the methodology adopted in our research project (cf. Fischer 2003), which ensures the production of intuitive language to a higher degree than more restricted settings would allow. The identification of intuitive strategies is vital for our research aim of enabling natural and effective human-robot interaction in spatial settings, specifically in light of the fact that current natural language interfaces are often not evaluated with respect to their effectiveness when confronted with users who are not informed about the robot's vocabulary – and if they are, this may lead to devastating results (Thrun 2004). Clearly, the results of this kind of empirical study and analysis differ in content and generalisability from results gained in more controlled psycholinguistic experiments. In the long run, it is desirable to work towards establishing methodologies combining the advantages of both approaches, as described, for instance, in von Stutterheim et al. (1993). Interdisciplinary collaboration is specifically targeted in the growing field of human-robot interaction research (Burke et al. 2004), where various research directions are combined out of necessity.

The present study can be regarded as an exploration of natural language produced in an open (though artificial) setting by unbiased speakers, which has led, on the one hand,

to an assessment of the diversity in linguistic choices in a range of situations allowing for different interpretations and viewpoints, and on the other hand, to the identification of systematic principles underlying speakers' choices, which need to be confirmed by more controlled and at the same time more restricted experimentation. In addition to that, in order to gain insights with respect to a broader range of settings the validity of the hypotheses also need to be tested in other configurations, and considering further influencing factors. Furthermore, the specific strategies users develop in a human-robot interaction setting, even if the spatial situation resembles the depictions in the present study, can only be addressed in real-world experimentation with users who are not informed about the robot's capabilities.

## 4 Conclusion

In this paper, results of a web-based empirical study designed to collect spatial localisation utterances and test for systematic patterns of usage were presented. The linguistic analysis was carried out qualitatively and quantitatively, providing relative frequencies of participants' linguistic and spatial choices. Insights concerning linguistic variability in interplay with conditions and configurations were gained. It was shown that, out of the range of linguistic options, some were clearly preferred under specific conditions while others did not occur there at all, though favored in other situations. The qualitative analysis revealed that the systematic patterns of usage point to underlying principles speakers adhere to when establishing contrastive reference in a spatial scenario, similar to those known for non-spatial object reference and in line with previous results in spatial language research. For instance, speakers choose a reference system and spatial axis that is suitable for contrasting the target object from competing ones. The exact spatial location is usually not specified if there are no competing objects closeby. This results in the frequent usage of unmodified projective terms even if the target object is located at considerable distance from the reference axis, which contrasts from the usage of spatial terms in other kinds of tasks.

## Acknowledgements

I thank the numerous anonymous participants for contributing their time and effort without reward, and the organizers of the web portal at www.language-experiments.org for hosting the study. T. Reinhardt's technical assistance in preparing the web study and data retrieval is highly appreciated, as well as K.-H. Wagner's support. Many thanks to three anonymous reviewers, M. Knauff, J. Restat, A. Stefanowitsch, and C. Vorwerg for very helpful comments on earlier versions of this paper, and the members of the SFB/TR 8 project I1-[OntoSpace] (J. Bateman, S. Farrar, K. Fischer, and R. Moratz) and other projects of the SFB/TR 8 Spatial Cognition for teamwork, cooperation, and invaluable discussions in relation to the work presented here. Funding by the Deutsche Forschungsgemeinschaft (DFG) is gratefully acknowledged.

## References


Brown-Schmidt, S. & M.K. Tanenhaus. 2003. Referential domains and the interpretation of referring expressions in interactive conversation. *Proc. DiaBruck, 7th Workshop on the Semantics and Pragmatics of Dialogue, Sept. 4th-6th 2003, Wallerfangen*, pp 15-19.

Bryant, D.J., B. Tversky, & M. Lanca. 2000. Retrieving spatial relations from observation and memory. In E. van der Zee & U. Nikanne (eds.), *Cognitive interfaces: Constraints on linking cognitive information.* Oxford: Oxford University Press, 94-115.

Burke, J.L., Murphy, R.R., Rogers, E., Lumelsky, V.J., & Scholtz, J. 2004. Final report for the DARPA/NSF interdisciplinary study on human-robot interaction. *IEEE Transactions on Systems, Man and Cybernetics, Part C*, Vol. 34 No.2, pp. 103-112.

Carlson, L.A. 1999. Selecting a reference frame. *Spatial Cognition and Computation,* 1 (4), 365-379.

Carlson, L. A., T. Regier, & E. Covey. 2003. Defining Spatial Relations: Reconciling Axis and Vector Representations. In E. van der Zee & J. Slack (eds.), *Representing Direction in Language and Space*. Oxford: Oxford University Press.

Carlson-Radvansky, L.A. & G.D. Logan. 1997. The Influence of Reference Frame Selection on Spatial Template Construction. In *Journal of memory and language* 37, 411-437.

Clark, H.H. 1996. *Using Language*. Cambridge, UK: Cambridge University Press.

Coventry, K. R. & Garrod, S. C. 2004. *Saying, seeing and acting: The psychological semantics of spatial prepositions*. Psychology Press: Essays in Cognitive Psychology series.

Eschenbach, C. 2004. Contextual, functional, and geometric features and projective terms. In L. Carlson and E. van der Zee (eds.), *Functional features in language and space: Insights from perception, categorization and development*. Oxford: Oxford University Press.

Fischer, K. 2003. Linguistic Methods for Investigating Concepts in Use. In: Stolz, Th. and K. Kolbe (eds.): *Methodologie in der Linguistik*. Frankfurt a.M.: Lang.

Fischer, K. & R. Moratz. 2001. From Communicative Strategies to Cognitive Modelling. Workshop Epigenetic Robotics, Lund.

Franklin, N., Henkel, L.A. & Zangas, T. 1995. Parsing surrounding space into regions. *Memory and Cognition, 23,* 397-407.

Freksa, C. 1980. Communication about visual patterns by means of fuzzy characterizations. *XXIInd Intern. Congress of Psychology*, Leipzig, July 1980.

Gapp, K.-P. 1995. An empirically validated model for computing spatial relations. In I. Wachsmuth, C.R. Rollinger & W. Brauer (eds.): *KI-95: Advances in Artificial Intelligence. 19th Annual German Conference on Artificial Intelligence* (pp 245-256). Springer: Berlin.

Grice, H.P. 1975. Logic and conversation. In: Cole, P. and J. Morgan (eds.), *Syntax and semantics*. New York, San Francisco, London (Vol 3: 41-58).

Herrmann, T. 1990. Vor, hinter, rechts und links: das 6H-Modell. Psychologische Studien zum sprachlichen Lokalisieren. *Zeitschrift für Literaturwissenschaft und Linguistik 78.* 117-140.

Herrmann, T. & W. Deutsch. 1976. Psychologie der Objektbenennung. Bern: Huber Verlag.

Herrmann, T. & J. Grabowski. 1994. *Sprechen: Psychologie der Sprachproduktion.* Heidelberg: Spektrum.

Herskovits, A. 1986. *Language and spatial cognition.* Cambridge University Press.

Kessler, K., I. Duwe & H. Strohner. 1999. Grounding Mental Models: Subconceptual Dynamics in the Resolution of Reference in Discourse. In: G. Rickheit & C. Habel (eds.), *Mental Models in Discourse Processing and Reasoning*. Amsterdam: Elsevier.

Levinson, S.C. 2003. *Space in Language and Cognition*. Cambridge University Press.

Moratz, R. & K. Fischer. 2000. Cognitively Adequate Modelling of Spatial Reference in Human-Robot Interaction. *Proceedings of the 12th IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2000*, November 13-15 2000, Vancouver, British Columbia, Canada, pp. 222-228.

Moratz, R., T. Tenbrink, J. Bateman, & K. Fischer. 2003. Spatial Knowledge Representation for Human-Robot Interaction. In Freksa, C., W. Brauer, C. Habel & K.F. Wender (Eds.), *Spatial Cognition III* (pp 263-286). Berlin, Heidelberg: Springer.

Olivier, P. & K.-P. Gapp (eds.), 1998. *Representation and Processing of Spatial Expressions*. Mahwah, New Jersey: Lawrence Erlbaum

Pickering, M. & Garrod, S. (in press). Toward a mechanistic psychology of dialogue. To be published in *Behavioral and Brain Sciences*.

Pobel, R., Grosser, Ch., Mangold-Allwinn, R. & Herrmann, Th. 1988. Zum Einfluß hörerseitiger Wahrnehmungsbedingungen auf die Überspezifikation von Objektbenennungen. Arbeiten der Forschergruppe "Sprechen und Sprachverstehen im sozialen Kontext" Heidelberg/Mannheim, Bericht Nr. 17.

Reips, U.-D. 2002. Theory and techniques of Web experimenting. In B. Batinic, U.-D. Reips, & M. Bosnjak (Eds.), *Online Social Sciences*. Seattle: Hogrefe & Huber.

Retz-Schmidt, G. 1988. Various views on spatial prepositions. *AI Magazine 9*: 2. 95-105.

von Stutterheim, C., R. Mangold-Allwinn, S. Barattelli, U. Kohlmann, & H.-G. Kölbing. 1993. Reference to objects in text production. In: J. Nuyts & E. Pederson (eds.), *Perspectives on Language and Conceptualization*. Belgian Journal of Linguistics 8., pp. 99-125.

Taylor, H.A., S.J. Naylor, R.R. Faust, & P.J. Holcomb. 1999. "Could you hand me those keys on the right?" Disentangling spatial reference frames using different methodologies. *Spatial Cognition and Computation*, 1 (4), 381-397.

Tenbrink, T. & R. Moratz. 2003. Group-based Spatial Reference in Linguistic Human-Robot Interaction. *Proceedings of EuroCogSci 2003: The European Cognitive Science Conference*, September 10-13, Osnabrück, Germany, pp 325-330.

Thrun, S. 2004. Toward a Framework for Human-Robot Interaction. *Human-Computer Interaction, Volume 19 (2004), Numbers 1 & 2*, pp. 9-24.

Tversky, B. & P.U. Lee. 1998. How Space Structures Language. In: C. Freksa, C. Habel & K.F. Wender (eds.), *Spatial Cognition. An Interdisciplinary Approach to Representing and Processing Spatial Knowledge* (pp. 157-175).

Vorwerg, C. 2001. *Raumrelationen in Wahrnehmung und Sprache: Kategorisierungsprozesse bei der Benennung visueller Richtungsrelationen*. Wiesbaden: DUV.

Watson, M.E., M.J. Pickering, & H.P. Branigan. 2004. Alignment of Reference Frames in Dialogue. *Cogsci 2004: 26th Annual Meeting of the Cognitive Science Society*, August 5-7, 2004, Chicago.

van der Zee, E. & J. Slack (eds.), 2003. *Representing Direction in Language and Space*. Oxford: Oxford University Press.

Zimmer, H.D., H.R. Speiser, J. Baus, A. Blocher, & E. Stopp. 1998. The Use of Locative Expressions in Dependence of the Spatial Relation between Target and Reference Object in Two-Dimensional Layouts. In C. Freksa, C. Habel & K.F. Wender (eds.), *Spatial Cognition* (pp. 223-240). Berlin: Springer-Verlag.