Reference Frames for Spatial Inference in Text Understanding

Berry Claus, Klaus Eyferth, Carsten Gips, Robin Hörnig, Ute Schmid, Sylvia Wiebrock and Fritz Wysotzki*

Methods of Artificial Intelligence
Department of Computer Science, Technical University Berlin
email: sppraum@cs.tu-berlin.de

Abstract. We present an approach to spatial reasoning that is based on homogenous coordinate systems and their transformations. In contrast to qualitative approaches, spatial relations are not represented by symbolic expressions only but additionally by parameters with constraints, which are subsets of real numbers. Our work is based on the notion of mental models in text understanding introduced in cognitive science. That is, we model the understanding of descriptions of spatial configurations by constructing a representation of the class of situations that are compatible with the description. Within our approach a spatial relation between two objects is represented by constraining the position of one of the objects with respect to the reference frame of the other one. That is, inferences of spatial relations not given explicitly in a text depend on the frame of reference which is presumed by the (human or computer) system. In this paper we describe a general framework for spatial reasoning based on the notion of mental models and we present some empirical results concerning the influence of the selected reference frame on the inference process.

1 Introduction

To reason about space is a fundamental human ability pervading our everyday life. In this paper we focus on a special aspect of spatial reasoning: the understanding of natural language descriptions of spatial scenes. Without this ability we could not make use of route descriptions in navigation ("to reach the post office, you have to turn left at the next crossing"), of instructions for installing or using technical equipment ("insert the disc in the slot under the console and press the right button") or for locating things ("the book is in the left part of the shelf in my room"). As the examples illustrate, understanding information of this kind can be crucial for performing purposeful actions. This is not only true for humans but also for computer systems and robots: It is a long standing goal

* Fritz Wysotzki is the coordinator of the project. Authors are listed in alphabetical order. This research was supported by the Deutsche Forschungsgemeinschaft (DFG) in the project "Modellierung von Inferenzen in Mentalen Modellen" (Wy 20/2–1) within the priority program on spatial cognition ("Raumkognition").

in computer science to control graphic user-interfaces or robot actions by means of a (quasi) natural language input. For this purpose there is need for a higher programming language making use of the operational semantics of spatial relations (like "left" or "in front") and taking into account physical constraints such as occupied space (cf. Ambler and Popplestone 1975; Lenzmann and Wachsmuth 1996).

Our aim is to construct an algorithm which is able "to understand" descriptions of spatial scenes. A common view in cognitive science is that human text understanding is to be construed as building up a so called *mental model* (cf. Garnham 1985, p. 6). That is, to understand a given text, the understanding (human or computer) system has to construct an internal representation of the situation described in the text. This idea guides our work on the understanding of spatial descriptions. A mental model of a description of a spatial configuration has to represent the objects introduced in the text and the spatial relations between them in a way which is "structurally identical" (cf. Johnson-Laird 1983, p. 419) to the described situation. Accordingly, understanding does not end up in determining the propositions (qualitative representation) capturing the "meaning" of the text (i.e. the intensional semantics). Instead, a representation formalism is needed that captures the class of all possible spatial configurations which are consistent with the description (i.e. a kind of model-theoretic semantics). Furthermore, this "spatial representation" has to be constructed incrementally from a sequence of propositions given in a text. The representation has to be updated with respect to new incoming information, and relations between objects not given explicitly should be inferrable if needed.

As a representation formalism, realizing the characteristics of mental models given above, we are using a directed graph with objects as nodes and spatial relations between pairs of objects as arcs. Each object node is associated with a 3-dimensional coordinate system with the origin determining the object position. Orientation of objects is represented by the axes of this coordinate system. A spatial relation between objects A and B is represented by a transformation matrix specifying the translation and rotation operations that map the coordinate system of object A onto that of object B. The idea that information about object positions is represented by homogeneous transformation matrices has biological plausibility, as was shown by O'Keefe (1989). We use parameters associated with constraints to represent the class of possible situations compatible with the given description. That is, the parameters represent subsets of real numbers. Our system is provided with an (extendable) set of predefined spatial relations (like "right of", "in front") associated with constraints on the transformation matrices representing the operational semantics of these spatial prepositions. To infer spatial relations not explicitly given in a text transformation matrices are multiplied. Inference proceeds along the arcs of the labelled graph representing a described configuration. If, for example, the information "B is *right-of* A" and "C is *on* B" is given, the relation "C is *right-of* A" can be inferred: The matrices annotated at the arcs $A - B$ and $B - C$ are multiplied and a new arc representing the relation between A and C is introduced. Note that inference is done not on the level of linguistic variables (qualitative expression), but on the level of possible positions and orientations of objects (as subsets of real numbers). Therefore, it is always possible to infer a spatial relation regardless of whether there exists a corresponding linguistic label or not.

The crucial characteristic of our approach is that it relies on "spatial representations" (cf. Glasgow and Papadias 1992, p. 356). It can be considered as an intermediating alternative to "language oriented", logic-based models and to "perception-oriented" depictional models. Logic-based models (e.g. Guesgen 1989, Hernández 1991, Mittal and Mukerjee 1995, Renz and Nebel 1997) operate on qualitative information, i.e. symbolic expressions (like *left-of(x,y)*). Inference of relations between objects ("qualitative reasoning") is performed by syntactical methods. A variant of "purely" logic approaches is to represent conceptual knowledge about spatial configurations by semantical networks (cf. Anderson and Bower 1973). As in our graph-based model, objects are represented by unique nodes. But objects and relations are represented by linguistic labels (symbolic expressions) only. In our approach objects and geometric relations are represented by coordinate systems and real valued parameters which can additionally be labelled by spatial concepts. We believe that representing quantitative aspects of spatial relations can be a fruitful approach to spatial (or other domains of) reasoning (cf. Guesgen 1997). Working on the model of a described configuration instead of a syntactical representation can have methodological advantages for constructing inference algorithms. The usefulness of constructive approaches to reasoning has already been demonstrated in the research on type theory in computer science (cf. Thompson 1991) and is also discussed in the area of spatial reasoning (see Schlieder 1998). Furthermore, model construction can be regarded as "semantic reasoning" which mimics human information processing (Johnson-Laird and Byrne 1991).

In depictional models (e.g. Waltz and Boggess 1979, Paivio 1971, Pribbenow 1993) object positions are represented by absolute values with respect to a global coordinate system. There is no need for inference of spatial relations. Instead information can be retrieved from the picture by a scanning process. In our approach, object relations are represented not absolutely but relatively. Because the information about a spatial configuration given in a text is underspecified with respect to object positions and orientations, our representation denotes the class of all possible depictions of the described situation (Johnson-Laird 1983). In contrast to qualitative approaches, where underspecification of symbolic descriptions is an inherent feature, the class of possible depictions is represented explicitly by the parameters of the transformation matrices (and their constraints). That is, our model could be extended to verify whether a picture corresponds to a verbal description, and to generate prototypical depictions.

We believe that this approach to spatial reasoning can be fruitful for different applications in the area of natural language interaction with computer systems and robots. The introduction of spatial inference by means of the manipulation of coordinate systems offers a more "natural" framework for spatial reasoning than the more prominent qualitative syntactic approaches. Additionally, our model corresponds to the approach of mental models in text understanding and reasoning. This enables us to mutually profit from work in cognitive psychology and artificial intelligence. On the one hand, we profit from insights gained in experimental cognitive psychology, and on the other hand, the model can be seen as a formal and more precise version of the cognitive approach of mental models in text understanding. In the following, we first give some general ideas about reference frames, arguing that the choice of reference frames is crucial for inferring spatial relations. Afterwards our model is described in more detail. In Sect. 4 psychological experiments regarding the problem of online inferences are presented. Finally we will give an outlook on further work to be done.

2 Frames of Reference

We investigate texts that contain sentences of the form "B is standing in spatial relation r to A.", such as (1):

(1) Der Kühlschrank steht links von Torsten.
The refrigerator stands on the left of Torsten.

The expression "links von" in (1) is itself ambiguous in that it allows for an intrinsic interpretation as well as a deictic one. The difference between the two readings relies crucially on the reference frame that is critical in linguistically localizing object B (the refrigerator in (1), we call it the *referent*) with respect to A (Torsten in (1), called the *relatum*). The intrinsic reading calls for a reference frame in terms of the intrinsic properties of the relatum itself, with respect to which the object has to be localized. Given an intrinsic reading of "left of" in (1), the refrigerator is located with respect to the left-hand side of Torsten. A deictic reading demands a reference frame not explicitly mentioned in the sentence, which is distinct from the one of the relatum. The most prominent candidate for the reference frame in question is the intrinsic one of the speaker as oriented towards the relatum (see Miller and Johnson-Laird 1976). Now, with respect to semantics, to know which of the two readings is the intended one is to know which of the candidate reference frames is the relevant one in order to interpret the sentence. But with respect to a mental model this is not the whole story: taking the intrinsic reading of the expression as the intended one in (1), the reader not only has to represent the objects mentioned in the sentence, but moreover, the represented relatum actually has to provide a spatial reference frame within which the referent may then be localized.

One way to think about the reference frame of the relatum is to attach a three-dimensional coordinate system to the represented object, derived from its intrinsic properties. We are then able to define the x-axis as the left-right axis of Torsten's body and take the negative region of the x-axis as the area on Torsten's left, i.e. the region where to localize the refrigerator in the mental model. This is exactly the way in which the formalism we chose in our project serves our needs. The relational expression is used to constrain the possible positions of the referent relative to the reference frame of the relatum. With such a conception in mind, we can speak of a *minimal mental model* just in case that all spatial relations explicitly mentioned in the text, and only these, lead to the addition of further edges (and eventually to the addition of a further node, if a mentioned object lacks a representing node). A minimal mental model does not include any inferred spatial relation between objects. As a valid inference we consider the following one: If sentence (1) is followed by sentence (2), it would be valid to infer that the bowl is on the left of Torsten.

(2) Die Schüssel steht auf dem Kühlschrank.
The bowl is standing on the refrigerator.

Such an inference would yield the positions of the referents of both sentences (refrigerator and bowl) relative to one and the same reference frame (the one of Torsten). We consider such a mental model in a sense as a more integrated spatial representation.

2.1 Is there any overall organizing principle?

A text describing an object arrangement might include a number of spatial relational expressions calling for an intrinsic reading involving a whole range of different relata (e.g., a text comprised of sentences (1) and (2)). The corresponding minimal mental model lacks an overall organizing principle, in that every object position is specified only with respect to the reference frame of the object to which it was related in the description, i.e. its relatum. Since a minimal mental model might thus be far from being an integrated spatial representation, one may reasonably doubt that a reader ends up in constructing a minimal mental model.

Within our formal approach we may restate the problem as follows: Is there any global or most dominating reference frame within which every object in a given described situation has to be localized at the time of encoding? If this is the case we might postulate a need for an inference in case of an intrinsic relational expression whenever the reference frame of the object that serves as a relatum in the text, differs from the global one. If we were to take the coordinate system of Torsten as the global one with respect to sentences (1) and (2), indeed we would have to infer the position of the bowl within the reference frame of Torsten at encoding time, as a consequence of the overall organizing principle. In terms of the psychology of text comprehension, an overall organizing principle may be taken as a coherence principle, defined under the auspices of a spatial mental model. Thus to infer a spatial relation not explicitly stated in the text as a consequence of the overall organizing principle may be characterized as a *bridging inference*, i.e. an inference which is drawn at encoding time in order to establish coherence (Garnham 1985, p. 157f).

In the example above, to qualify the reference frame of Torsten as global may seem rather arbitrary. Beyond the scope of text comprehension two structural principles for spatial information are of primary interest: the difference as well as the interplay of an egocentric reference frame and an allocentric reference frame. Both are discussed as organizing principles in structuring space within given contexts, especially in navigating in space using route knowledge as well as survey knowledge, often tacitly connecting route knowledge with an egocentric reference frame and survey knowledge with an allocentric reference frame (see Werner et al. 1997).

2.2 Egocentric reference frame

In analysing deictic expressions Bühler (1982) introduced the term "origo" as the name for the origin of the egocentric reference frame. But the origo does not necessarily coincide with the actual perceptible 'Here', as Bühler discusses under the heading of "deixis at phantasma". As his second major case of transposition Bühler (1982, p. 34) considers the phenomenon when "one is transported in the imagination to the geographical place of what is being imagined, one has this present before the mind's eye from a determinate point of view, which may be given and at which is found the self's own position in the imagined scene." In the spirit of Bühler's note on transposition we claim that a person reading a text about objects surrounding a protagonist has the opportunity to transpose herself into the described situation, the determinate point of view being the one of the protagonist. In this case the reader takes the protagonist's perspective[1], thus constructing an egocentric mental model. That is to say: Given the protagonist's perspective, in comprehending (1), the reader may localize the refrigerator in the egocentric mental model with respect to her own reference frame, which is - by definition of the protagonist's perspective - in congruence with the one of the protagonist (Torsten).

If we take the egocentric reference frame as an overall organizing principle as introduced above, it follows that a reader taking the protagonist's perspective localizes every object mentioned in the description within her own egocentric reference frame. As a consequence, for every object not directly linguistically localized within the protagonist's reference frame - which would give the position within the egocentric reference frame at the same time because of the protagonist's perspective - the position within the egocentric reference frame would have to be inferred at the time of encoding. This is the case whenever the relatum differs from the protagonist, e.g., for the localization of the bowl in (2), which is linguistically localized with respect to the reference frame of the refrigerator. Such an egocentrically organized mental model exhibits the interesting property that the representation of the object arrangement changes as a result of every movement (e.g., rotation) of the protagonist. Given (1), if the protagonist Torsten now turns to his left, the refrigerator will afterwards be in front of him. The same holds for the position of the refrigerator within the egocentric mental model, given a protagonist's perspective.

[1] We take the protagonist's perspective as a special case of what we call "internal perspective". With "internal perspective" we indicate any case in which the reader transposes herself into the described scene regardless of the imagined viewpoint.

2.3 Allocentric reference frame

There is an intuitive understanding of an object arrangement according to which the arrangement remains the same when the protagonist turns to his left, keeping his position constant. This intuitive understanding is reflected by the deictic reading of (1), which is not sensitive to the orientation of the relatum Torsten[2], but instead would only be sensitive to the orientation of a reference frame whose origin is marked as origo; let it be that of the speaker. Given a deictic reading, the refrigerator remains on the left of Torsten regardless of his orientation, as long as the speaker's reference frame remains constant. But in this special case of a deictic reading it seems obvious that it is not the object arrangement that changes if the speaker's reference frame changes with respect to the object arrangement. Instead, this indicates a change in the perspective of the speaker, thus altering the conditions for description (see Hörnig, Claus, and Eyferth 1997 on the identity of object positions in dependence on the reference frame and further arguments in favor of an allocentric reference frame.)

In constructing a mental model that captures the identity of the object arrangement when the orientation of the protagonist changes, we may localize every object within an allocentric reference frame. An allocentric reference frame as an overall organizing principle has to be external to the object arrangement and stipulated as constant. We propose a room-reference system for this purpose. Sentences (1) and (2) are part of a text which begins by introducing Torsten as standing in the middle of his kitchen. Now, what remains constant in the described situation, despite any re-orientation of Torsten, are the positions of all of the objects with respect to the kitchen (including that of the protagonist, since our texts did not mention any translation of the protagonist, thus keeping his position constant). The question, which reference frame to select as the allocentric one, can only be answered in dependence on the identity conditions of the represented object arrangement: If one would want to represent a re-orientation of the kitchen, one surely would need an allocentric reference frame external to the one of the kitchen (e.g., the object being originally in the *south* of the kitchen now being in the *west*, if the object position with respect to the kitchen remains constant.)

In order to construct a mental model with an allocentric room-reference frame, e.g., the reference frame of the kitchen, the room-node has to provide a coordinate system. It does not seem in itself problematic to provide an abstract coordinate system for the kitchen. But since the kitchen does not have intrinsic properties, linguistic expressions like "left of" would not allow for a straightforward interpretation based on the reference frame of the kitchen. For

[2] In terms of our formal approach, identity of two object arrangements with respect to a single global reference frame is dependent only on those parts of the transformation matrices that encode translation, but independent of rotation. For the identity of object arrangements it does not matter if the localized objects change orientation. In contrast, a re-orientation of the global reference frame would affect those parts of the matrices that encode translation as well, as is the case with the egocentric mental model.

this reason we suggest that the x-axis of the kitchen does not by itself admit an interpretation as the left-right-axis. If such a convention is needed, the allocentric reference frame must be supplemented by an extrinsic orientation as can be given by imposing an overview perspective on the reference frame of the kitchen (occasionally it has been suggested that in case of an overview perspective front/behind has to be mapped onto below/above, respectively).

Even if the allocentric reference frame is tied to an overview perspective, it remains essentially different from an egocentric reference frame as discussed above in several respects. While both cases clearly operate with an internal perspective, i.e. with a determinate point of view, the overview perspective is not egocentric. *Ego* is not surrounded by objects (*internal spatial viewpoint* in Bryant et al. 1992) but the whole object arrangement is in front of *ego* (*external spatial viewpoint* in Bryant et al. 1992). Especially, *ego* is definitely not part of the object arrangement. Finally, at least when tied to an allocentric reference frame, the overview perspective is required to be always in accordance with the allocentric reference frame, which itself is presupposed to be inherently constant.

The question of an overall organizing principle leads directly to the topic of inference. In the next section we will describe *how* spatial relations can be inferred within our formal approach to mental models. In Sect. 4 we will take a closer look at the issue of *when* and *what* inferences humans draw, and propose a possible interplay of the egocentric reference frame with an allocentric one.

3 Modelling Spatial Inferences by Transformation Matrices and Constraints

As discussed above, the understanding of texts about spatial scenes can be described as the incremental construction of a *mental model*. Mental models are characterized as internal representations of the described situation which is structurally identical to it in the sense that not only the relations explicitly given in the text are represented, but also inferences of relations holding in the situation can be made. To capture the characteristics we construct a directed labelled graph. Nodes in this graph represent the objects mentioned in the text, whereas the relations are represented by labelled arcs. We pursue an approach which was developed for robotics applications (cf. Ambler and Popplestone 1975). In this setting objects are associated with a coordinate system with distinguished axes for the left/right, front/back, and up/down directions. The spatial relation between two objects A and B can then be described by a transformation matrix T that maps the coordinate system of A onto the coordinate system of B. T consists of a matrix R for the rotation of the coordinate system, and a translation vector Δ. Transformation matrices have been used for qualitative reasoning as well (Mukerjee and Mittal 1995), but we use symbolic variables that can be instantiated by real numbers. For the simple case where only rotations about

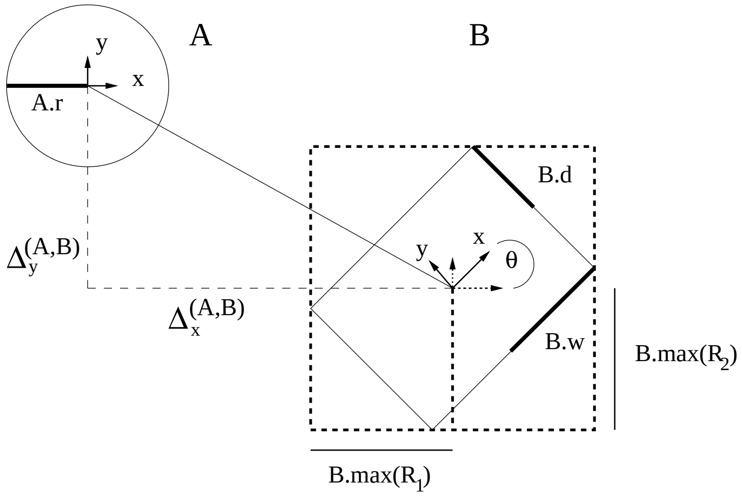

Fig. 1. Relation with relatum A and localized object B

the vertical z-axis are considered, the matrices have the form

$$T = \begin{pmatrix} & \mathrm{R} & & \Delta_x^{(A,B)} \\ & & & \Delta_y^{(A,B)} \\ & & & \Delta_z^{(A,B)} \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad R = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

In (Mukerjee and Mittal 1995), all matrix entries are elements of $\{-1, 0, 1\}$. This means that distances cannot be represented. Furthermore, there are only nine possible sectors for spatial relations in two dimensions, and the extension of objects (non-overlapping) cannot even be formulated. Therefore, we think metric information is necessary. The transformation is shown in Fig. 1. R_1 and R_2 are the first and second row vector of R, respectively, and are used to compute the size of the smallest bounding box containing B that is parallel to A. This bounding box is used to simplify computations and is shown with dotted lines. The thickest lines are used for the size parameters $A.r$, $B.w$, and $B.d$. A relation is then defined by constraints on the transformation matrix. Below we will describe our assumptions about the model, the representation of objects, the definition of relations and the inference process.

3.1 Constructing the graph

The texts we consider describe spatial scenes in a room. Therefore the graph initially contains the nodes for the *room* and its parts (objects are discussed below). We represent the background knowledge about the relations between the subparts in the arcs from the *room*-node to the nodes for the parts. Relations between two walls, e.g., can be computed. The constraints with respect to the

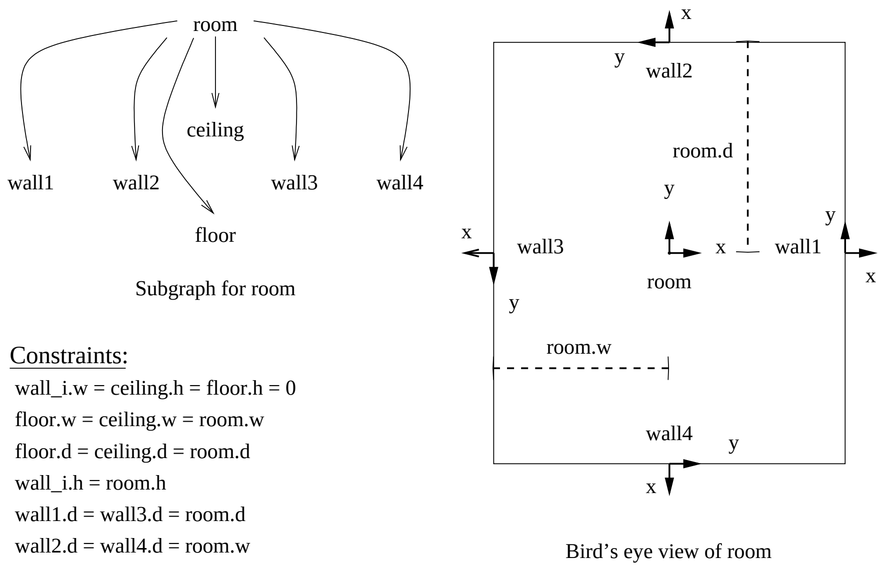

Fig. 2. Default constellation for *room* object(s)

extensions of the subparts are stored in a constraint-table. The default layout of the room with the location of the coordinate systems, the constraints, and the initial graph are shown in Fig. 2. The extensions of the objects are denoted by *w*idth, *d*epth and *h*eight, respectively. Spatial relations are given by propositions like *left_of(Torsten, refrigerator)* for sentence (1). To insert this arc into the graph, we have to check whether there are already nodes for *Torsten* and *refrigerator*. If not, the descriptions of the objects have to be looked up (see below) and the nodes must be created. Then the definition of the relation *left_of* is looked up, the constraints for *left_of(Torsten, refrigerator)* are computed and are stored in the constraint-table. Now the directed arc from *refrigerator* to *Torsten* is added to the graph and is labelled with *left* and the transformation matrix. If a new node is introduced, the background knowledge that all objects are *in* the room, is represented by an arc from the object to the *room* node.

Objects. Objects are represented by *cuboids* and *cylinders*. Simple symmetrical bodies are commonly used to build up objects in programs for spatial inference (Waltz and Boggess 1979) and for the description of office environments (Jörding and Wachsmuth 1996). As we don't get any visual input, we feel we don't need more complicated objects. We don't use default values for the extensions of objects, but it is possible to give upper and lower bounds for the dimensions. The coordinate system is located in the center of the object. The extension is always measured from the origin of the coordinate system to the border of the object (see Fig. 2). The variables for the extensions of objects are *w*idth, *d*epth and *h*eight for cuboids, and *r*adius and *h*eight for cylinders. At present the only object which can be decomposed into its building parts, is the *room*. In Fig. 2, the arcs from the *room* node are labelled with the special relation *has part* (for lack of space, the labels are not shown in the picture) and the corresponding transformation matrix. The relation *has part* is used to ensure that for the mentioned *room* object, all sub-objects are included in the graph. The transformations are fully specified by the extension of the *room*. As a wall has no intrinsic axes for front/back and left/right, the orientation of these axes is arbitrary. We have chosen these coordinate systems for the walls so that the definition for the relation *at_the_wall* is the same for all walls. The constraints on the matrices and on the variables for the extensions of the (sub)objects ensure that the parts "fit together". A more detailed description of objects may become necessary when we want to construct depictions for a given model.

As stated in Sect. 2, to build a mental model from a text, it is necessary to know whether an intrinsic or a deictic interpretation of a relation is intended. Therefore it is essential to know which readings are possible. We provide the information which of the axes of the coordinate system correspond to intrinsic axes of the object. Currently we assume an intrinsic use of spatial propositions. In the future we hope to get results which allow to decide whether an allocentric or egocentric reference frame must be used for the interpretation of a proposition.

Relations. The definition of a relation consists of a name and the constraints for the transformation matrix. These constraints can restrict the rotation of the localized object with respect to the relatum and describe a region of R^3. In Fig. 3, a possible definition for the relation *right* is shown. In this case B must be fully inside the defined region. Because of the variables in the transformation matrix and the unknown extensions of objects we cannot use collision detection procedures to decide whether B is in the depicted region *right* of A. We simplify the problem by considering the smallest cuboid that is parallel to A and contains B. The size of this cuboid can be computed with the function max which takes an object and a direction vector as arguments. The result is still a symbolic expression containing the rotation θ.[3]

The constraints for the above definition of *right* are

$$\begin{aligned}
\Delta_x^{(A,B)} &\geq A.\max(1,0,0) + B.\max(R_1) \\
\Delta_y^{(A,B)} &\leq \quad (\Delta_x^{(A,B)} - A.\max(1,0,0) - B.\max(R_1)) \\
&\qquad + (A.\max(0,1,0) - B.\max(R_2)) \\
\Delta_y^{(A,B)} &\geq -[(\Delta_x^{(A,B)} - A.\max(1,0,0) - B.\max(R_1)) \\
&\qquad + (A.\max(0,1,0) - B.\max(R_2))]
\end{aligned}$$

The function max gets a vector w.r.t. A's coordinate system as argument and computes the maximum distance from the object's origin to the border of the bounding box in this direction. For A, $\max(1,0,0) = A.w$ for a cuboid, and

[3] Sometimes we may be able to restrict the possible orientations with defaults or background knowledge.

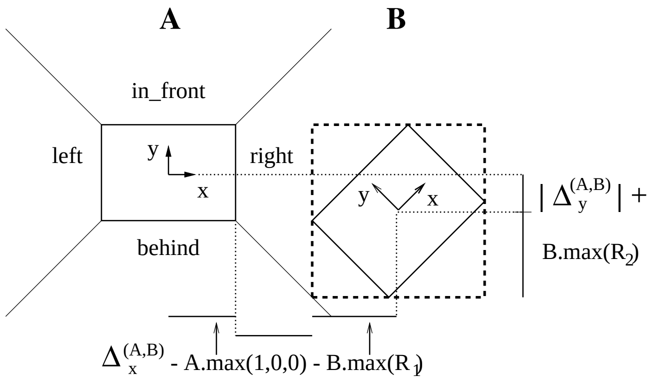

Fig. 3. A possible definition for *right*

$\max(1,0,0) = \max(0,1,0) = A.r$ for a cylinder. As can be seen in the picture, we don't get an exhaustive set of relations when we consider extended objects and require that the whole object is positioned within the region. This means we will often have situations where the relation (i.e. a transformation matrix) between two objects can be computed, but the corresponding arc cannot be labelled with the name of a defined relation. An example is shown in Fig. 4. For these specific positions of the objects and the above definitions, B is *right* of A, and C is *left* of B, but the relation between A and C cannot be named, because C overlaps the sector for *right* and the sector for *behind*.

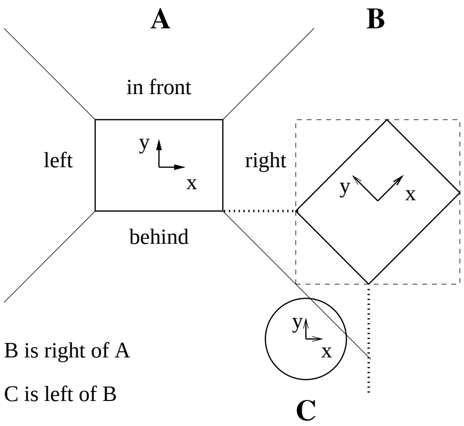

Fig. 4. Bird's eye view of a constellation of three objects

Finding linguistically justified constraints for spatial relations is not a goal of our work. We are experimenting with different sets of relations. In principle it is no problem to include results from psychological experiments that report overlapping, nonsymmetric regions of different sizes for the four basic directions

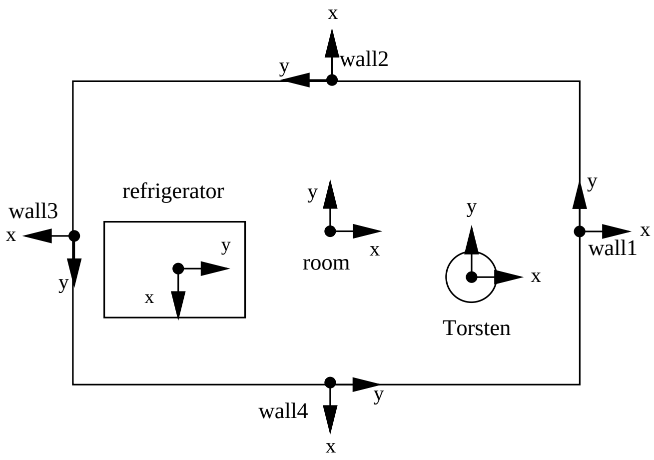

Fig. 5. Possible situation for "The refrigerator stands on the left of Torsten"

(Franklin, Henkel, and Zangas 1995; Vorwerg 1997; Vorwerg and Rickheit 1998). Whether the models that are possible for these regions are psychologically justified for text understanding tasks is an open problem. The constraints for the transformation matrix can be equations or inequations containing the elements of the matrix and the extensions of the two related objects. A possible topic of research is the introduction of functional roles for objects which would enable us to include context dependent information.

For our example sentence (1), one situation is depicted in Fig. 5. We have used the following assumptions: a) the protagonist faces a wall. As walls are so far undistinguished, we have set the coordinate system of Torsten parallel to that of the room. b) The refrigerator is oriented with its front side (positive y-axis) towards Torsten. We get the following transformation matrices (abbreviations: *To*rsten, *re*frigerator, *w*idth, *d*epth, *h*eight, and *r*adius).

$$T^{(room,\ To)} = \begin{pmatrix} 1 & 0 & 0 & \Delta_x^{(\text{room},\text{To})} \\ 0 & 1 & 0 & \Delta_y^{(\text{room},\text{To})} \\ 0 & 0 & 1 & \text{To.h} - \text{room.h} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

with the constraints

$$\begin{aligned} -(\text{room.w} - \text{To.r}) &\leq \Delta_x^{(\text{room},\text{To})} \leq \text{room.w - To.r} \\ -(\text{room.d} - \text{To.r}) &\leq \Delta_y^{(\text{room},\text{To})} \leq \text{room.d - To.r} \end{aligned}$$

$$T^{(To,\ re)} = \begin{pmatrix} 0 & 1 & 0 & \Delta_x^{(\text{To},\text{re})} \\ -1 & 0 & 0 & \Delta_y^{(\text{To},\text{re})} \\ 0 & 0 & 1 & \text{re.h} - \text{To.h} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

with the constraints

$$-(\Delta_x^{(\mathrm{room,To})} + \mathrm{room.w} - \mathrm{To.r} - 2\mathrm{re.d}) \leq \Delta_x^{(\mathrm{To,re})} \leq -(\mathrm{To.r} + \mathrm{re.d})$$

$$\begin{aligned}(\Delta_x^{(\mathrm{To,re})} + \mathrm{To.r} + \mathrm{re.d}) - (\mathrm{To.r} - \mathrm{re.w}) = \\ (\Delta_x^{(\mathrm{To,re})} + \mathrm{re.d} + \mathrm{re.w}) \leq \Delta_y^{(\mathrm{To,re})} \leq -(\Delta_x^{(\mathrm{To,re})} + \mathrm{re.d} + \mathrm{re.w})\end{aligned}$$

and

$$T^{(room,\ re)} = \begin{pmatrix} 0 & 1 & 0 & \Delta_x^{(\mathrm{room,re})} \\ -1 & 0 & 0 & \Delta_y^{(\mathrm{room,re})} \\ 0 & 0 & 1 & \mathrm{re.h} - \mathrm{room.h} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

with the constraints

$$\begin{aligned} -(\mathrm{room.w} - \mathrm{re.d}) \leq \Delta_x^{(\mathrm{room,re})} &\leq \Delta_x^{(\mathrm{room,To})} - \mathrm{To.r} - \mathrm{re.d} \\ -(\mathrm{room.d} - \mathrm{re.w}) \leq \Delta_y^{(\mathrm{room,re})} &\leq (\mathrm{room.d} - \mathrm{re.w}) \end{aligned}$$

3.2 Inferences

To infer a relation between objects B and relatum A not explicitly given in a text, we have to find a directed path $B = n_0 - n_1 - \ldots - n_k = A$ in the graph. We can then multiply the transformation matrices to get

$$T^{(A,\ B)} = T^{(A, n_{k-1})} \times \cdots \times T^{(n_1, B)}$$

Constraint propagation yields the constraints for the computed matrix which have to be matched against the predefined relations if we are interested in an interpretation (a proposition label, e.g. *right*) for the arc.

It is a topic of research when and how (along which paths) relations between objects in the model are inferred. As stated in Sect. 2, this also depends on the reference frame chosen for the interpretation of a spatial proposition. If a global reference frame P exists (e.g. the *room* or the protagonist typically introduced in the text), it would be possible to construct a tree where only the relations between the objects and P are represented by arcs (and inferred at encoding time) whereas no relation between any two other objects might be represented, even if explicitly given in the text. The other extreme would be to build a complete graph with an arc between every pair of objects, even if most arcs cannot be labelled with the name of a relation (i.e. the transformation matrix does not satisfy the constraints of any predefined relation). At present we have implemented a medium strategy. The graph consists of those arcs for which relations are explicitly given, the arcs for the background knowledge (any object is *in* the room), and arcs that are inferred after a query to the program. The existence of an arc from an object to the *room* node can be regarded as choosing the *room* object as a global reference frame. From the graph in Fig. 6 below, it should be obvious that we have more than one fixed reference frame, because otherwise the arcs labelled *left* and *on*, respectively, would not be included. The

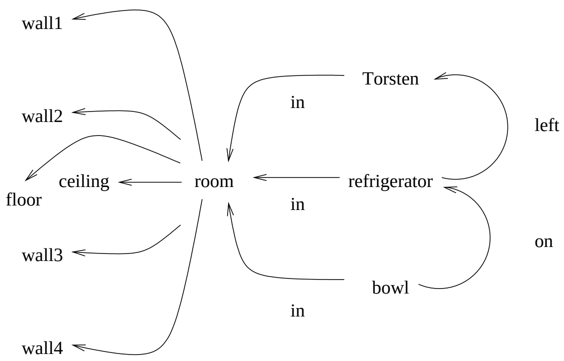

Fig. 6. The graph after two sentences

goal of this work is the implementation of several strategies dependent on the results of the psychological experiments described below.

If we continue our example with sentence (2), we have to add the arcs *bowl* $\longrightarrow$ *refrigerator* and *bowl* $\longrightarrow$ *room*. A *bowl* is modelled as a cylinder without intrinsic x- and y-axes. Therefore the orientation of its coordinate system is of no importance. For sake of convenience, we choose an orientation identical with that of the refrigerator. The graph is shown in Fig. 6.

For the arc *bowl* $\longrightarrow$ *refrigerator*, the rotation matrix is the identity matrix. The constraints are:

$$\begin{aligned} -(\mathrm{re.w} - \mathrm{bowl.r}) \leq \Delta_x^{(\mathrm{re,bowl})} &\leq (\mathrm{re.w} - \mathrm{bowl.r}), \\ -(\mathrm{re.d} - \mathrm{bowl.r}) \leq \Delta_y^{(\mathrm{re,bowl})} &\leq (\mathrm{re.d} - \mathrm{bowl.r}), \text{ and} \\ \Delta_z^{(\mathrm{re,bowl})} &= \mathrm{re.h} + \mathrm{bowl.h}. \end{aligned}$$

These constraints are propagated, so that we get

$$T^{(room,\ bowl)} = \begin{pmatrix} 0 & 1 & 0 & \Delta_x^{(\mathrm{room,bowl})} \\ -1 & 0 & 0 & \Delta_y^{(\mathrm{room,bowl})} \\ 0 & 0 & 1 & 2\mathrm{re.h} + \mathrm{bowl.h} - \mathrm{room.h} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

with the constraints

$$\Delta_x^{(\mathrm{room,re})} - (\mathrm{re.d} - \mathrm{bowl.r}) \leq \Delta_x^{(\mathrm{room,bowl})} \leq \Delta_x^{(\mathrm{room,re})} + (\mathrm{re.d} - \mathrm{bowl.r})$$
$$\Delta_y^{(\mathrm{room,re})} - (\mathrm{re.w} - \mathrm{bowl.r}) \leq \Delta_y^{(\mathrm{room,bowl})} \leq \Delta_y^{(\mathrm{room,re})} + \mathrm{re.w} - \mathrm{bowl.r})$$

To infer the relation between the bowl and Torsten (the relatum), we have to find a directed path from *bowl* to *Torsten*. In the figure there is no such path. Now, an advantage of the transformation matrix approach over propositional inference calculi is the fact that for every relation, the reverse relation can be computed by inverting the matrix[4].

[4] In general, the inverse matrix need not correspond to any predefined relation. For instance, in Fig. 5, *wall1* is left of *wall3*, and *wall3* is left of *wall1* for the relations

When we are allowed to follow arcs in both directions (the inverse arc is not inserted in the graph), we find the following four paths: *bowl* – *refrigerator* – *Torsten*, *bowl* – *room* – *Torsten*$^{-1}$, *bowl* – *refrigerator* – *room* – *Torsten*$^{-1}$, and *bowl* – *room* – *refrigerator*$^{-1}$ – *Torsten*, where the raised $^{-1}$ denotes inverting the preceding arc. Since the graph is always connected, there is at least one path between any two nodes. The search strategy for finding a path depends heavily on the organization of the mental model and our representing graph. In this first implementation, we can always take the shortcut via the *room* node, i.e., the path *bowl*–*room*–*Torsten*$^{-1}$. Now we can compute the matrix

$$
\begin{aligned}
&T^{(\text{Torsten},\text{bowl})} \\
&= T^{(\text{room},\text{Torsten})^{-1}} \times T^{(\text{room},\text{bowl})} \\
&= \begin{pmatrix} 1 & 0 & 0 & -\Delta_x^{(\text{room},\text{To})} \\ 0 & 1 & 0 & -\Delta_y^{(\text{room},\text{To})} \\ 0 & 0 & 1 & -(\text{To.h} - \text{room.h}) \\ 0 & 0 & 0 & 1 \end{pmatrix} \\
&\times \begin{pmatrix} 0 & 1 & 0 & \Delta_x^{(\text{room},\text{bowl})} \\ -1 & 0 & 0 & \Delta_y^{(\text{room},\text{bowl})} \\ 0 & 0 & 1 & 2\text{re.h} + \text{bowl.h} - \text{room.h} \\ 0 & 0 & 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} 0 & 1 & 0 & \Delta_x^{(\text{room},\text{bowl})} - \Delta_x^{(\text{room},\text{To})} \\ -1 & 0 & 0 & \Delta_y^{(\text{room},\text{bowl})} - \Delta_y^{(\text{room},\text{To})} \\ 0 & 0 & 1 & 2\text{re.h} + \text{bowl.h} - \text{room.h} - (\text{To.h} - \text{room.h}) \\ 0 & 0 & 0 & 1 \end{pmatrix}
\end{aligned}
$$

In the next step, we have to match the elements of the matrix against the constraints for all predefined relations (unless the query asks whether a particular relation holds). For the matrix above, it is easy to prove that *left_of(Torsten, bowl)* holds although the inference would be easier for the obvious path *bowl* – *refrigerator* – *Torsten*. Constraint solving is inherently difficult, because in the general case, there are inequations involving trigonometric functions. As shown above, a possible simplification of the problem is the use of defaults. Currently, we restrict the orientations of objects to multiples of 90° around the *z*-axis. This means that the coordinate systems of all objects become parallel or orthogonal. This makes the inferences easier without too much loss of expressive power. We are at present working on the problem of finding heuristics to prove trigonometric (in)equations for arbitrary rotations around the *z*-axis. We are also investigating the application of machine learning algorithms for constraintsolving (see Gips et al. 1998).

---

in Fig. 4, because the coordinate system of *wall3* is rotated 180° with respect to the coordinate system of *wall1*, and the relations are defined w.r.t. the intrinsic axes of the relatum.

In the above example, all possible situations are captured within one model. Underdeterminate descriptions may be handled by introducing disjunctive constraints, which no longer yield a unique model. Wysotzki et al. (1997) present an example of several pairwise incompatible models (all compatible with the description), which may eventually be disambiguated by information given later in the text. In the area of spatial reasoning, a similar problem arises if the premises do not license a straightforward conclusion. In this latter case there is evidence that human problem solvers prefer some conclusions over others (see Knauff et al. 1998). Schlieder (1996) and Berendt (1996) formulate cognitive principles in terms of the construction of *preferred mental models* in order to explain such preferences. In our domain we investigate if human text understanders prefer to infer object positions with respect to some reference frame as compared with some other. For this purpose we proposed two alternative organizational principles (see Sect. 2) that would lead to the construction of quite different graphs. Which of these graphs corresponds to the preferred mental model has to be verified empirically. The next section discusses the hypotheses that follow from these organizational principles, as well as one of our experiments and the results obtained so far.

4 An Empirical Approach to Online Inferences in Spatial Reference Frames

In this section we restrict our considerations to egocentric mental models, i.e. mental models constructed by a reader taking the protagonist's perspective. We primarily focus on the question whether the egocentric reference frame serves as an overall organizing principle, or if an egocentric mental model is provided with an allocentric reference frame. This section relies on the work done by Barbara Tversky and her colleagues on egocentric spatial mental models (Franklin and Tversky 1990, Franklin, Tversky, and Coon 1992, Bryant and Tversky 1992).

4.1 Spatial framework

In a series of experiments, Tversky and her colleagues demonstrated that object access in spatial mental models depends on the direction in which the object is located. In the experiments of interest here, the texts described objects located either around the reader (e.g., "object A is behind you") or around a protagonist. We will consider those instances where the reader took the protagonist's perspective. Along the horizontal plane, objects located in front of *ego* were accessed fastest. Throughout, objects located behind *ego* could be accessed faster than objects beside *ego*[5]. Tversky et al. explained these results by characterizing egocentric mental models as (internal) spatial framework models. A spatial

[5] While we examine only object arrangements along the horizontal plane, Tversky et al. additionally compared those objects with objects above and below *ego*. Objects above and below *ego* proved to be accessed fastest, as long as the recipient imagined herself as standing upright. Again, we limit our considerations to this last condition.

framework is a three-dimensional reference frame imposed on the surrounding space which is derived from an observer's body-axes. Differences in access latencies are explained in terms of dominance relations between the axes and their symmetry properties. While the above/below axis renders the primary orientation, the left/right axis is dominated by the front/back axis from which it is derived. Therefore, objects located on the front/back axis are accessed faster than on the left/right axis. Moreover, the front-back axis is asymmetric by keeping objects in front perceptually as well as functionally more salient, thus causing access to objects in front to be faster.

Access latencies are explained by properties of the egocentric reference frame, i.e. in terms of representation. Consequently, access latencies should exhibit the spatial framework pattern whenever the objects are localized directly within the egocentric reference frame. If access latencies deviate from the spatial framework pattern this indicates that the examined objects are not represented directly within an egocentric reference frame. With this argument at hand it is possible to empirically test the hypothesis that an egocentric mental model sets up an overall organizing principle as described above. As is stated by this principle, objects introduced in a text relative to a relatum distinct from the protagonist should nevertheless be located directly within the egocentric reference frame, already at the time of encoding by virtue of online inference. Consequently, all objects should be accessed as predicted by the spatial framework model[6]. Before we go into more detail, we will discuss an alternative model, whose predictions for access latencies deviate from the spatial framework pattern: the mental transformation model.

4.2 Mental transformation

In presenting the spatial framework model, Franklin and Tversky (1990) contrasted it to the mental transformation model. According to the mental transformation model, object access proceeds in much the same way as a perceptual search within a surrounding environment. Only objects in front are immediatly accessible. In order to access an object located someplace else one has to mentally re-orient (i.e. rotate) until the object lies in front and therefore becomes accessible. According to this model, access latencies increase as a function of the angle by which one has to re-orient in order to enable immediate access to the object. Fastest access along the horizontal plane is predicted for objects located in front of *ego*, as was predicted by the spatial framework model. With respect to objects behind and beside *ego*, the mental transformation model contradicts the spatial framework model. In order to access an object behind *ego*, one has to re-orient by 180°, which takes longer than to access an object on either side, which only requires a re-orientation by 90°.

[6] As far as we know, in the texts used by Tversky et al., the objects were always linguistically located directly within the egocentric reference frame. Therefore, the results do not call for an overall organizing principle in order to postulate immediate access within the egocentric reference frame.

It seems worthwhile to take a closer look at the presumptions of this idea of a mental transformation. As is evident from the findings by Tversky et al., objects located linguistically within the egocentric reference frame are immediately accessible without any need for a mental transformation. In terms of reference frames, a mental transformation has to be stated as a mental rotation of the egocentric reference frame against the surrounding space, i.e. an allocentric reference frame. So, it is only for objects not already represented directly within the egocentric reference frame that one may reasonably assume an access mediated by a mental transformation. As long as we rely on the egocentric reference frame to establish an overall organizing principle, there is no need for a mental transformation. On the other hand, if we do suggest such a need, we must admit for egocentric mental models in which objects are not always directly located within the egocentric reference frame but only within an allocentric one. This assumption concerning a twofold representation is then supplemented by a procedural principle. If object access within the egocentric reference frame fails, it has to proceed within the allocentric reference frame. This case demands an indirect access, maybe governed by a procedural principle such as mental transformation. In summary, we acknowledge the possibility of two concurrent organizing principles in case of an egocentric mental model: an egocentric reference frame connected with an allocentric reference frame.

If there are two concurrent organizing principles for egocentric mental models, how do they interrelate? We assume this as the question of how a reader gains orientation within a spatial mental model. As long as one considers solely the egocentric reference frame, the question of orientation does not arise, because *left* is necessarily on the left-hand side of *ego*. It is with respect to an allocentric reference frame that the question becomes a sensitive one. At the time the reader encodes sentence (1), given the protagonist's perspective, egocentric *left* coincides with the location of the refrigerator[7]. We claim that an object like the refrigerator, which may be assumed to be localized invariantly within the kitchen, enables the reader to establish her orientation within the spatial mental model by anchoring her egocentric reference frame within the allocentric room reference frame. An object serving this purpose we call an *anchoring* object (anchoring objects play roughly the role of local landmarks in navigation tasks; see Werner et al. (1997)). The location of an anchoring object in an egocentric mental model is always represented in at least two ways. It is specified with respect to the egocentric reference frame as well as with respect to the allocentric reference frame, thus interconnecting the two. As a consequence, anchoring objects are always represented directly within the egocentric reference frame and therefore directly accessible. Because of this, access latencies for anchoring objects should always exhibit the spatial framework pattern. On the other hand, there may be objects in the mental model which are not located with respect to the egocentric

[7] This point becomes clear when the protagonist re-orients: in this case *left* is still left within the egocentric reference frame, but egocentrical *left* changes relative to the refrigerator and its constant position within (the allocentric reference frame of) the kitchen (see Hörnig, Claus, and Eyferth 1997).

reference frame, but merely with respect to the allocentric reference frame (we call such objects *critical objects*). If subjects have to access critical objects, latencies may deviate from the spatial framework pattern. Rather, critical objects may be accessed as predicted by the mental transformation model.

Our formal approach raises a severe argument against the egocentric reference frame as an overall organizing principle. In case of a re-orientation of the protagonist, the reader taking the protagonist's perspective has to execute the re-orientation within the egocentric mental model. As a consequence, every edge from an object node in the graph linked to the *ego*-node has to be updated online, since the mental model represents the actual egocentric situation, which changes in case of a re-orientation. If our approach to represent mental models is valid, a re-orientation in an egocentric mental model will be more complex - and, in terms of computation, more costly - as more edges have to be updated. This makes it highly implausible that a reader infers every object location within the egocentric reference frame online, leading to a maximal number of edges to be updated. On the other hand, if object access within the allocentric reference frame requires a mental transformation, which in turn demands an update of all of the edges from the anchoring objects (whether continuously or in discrete steps), it turns out to be a highly complex process too. Mental transformation as an access principle is even more complex in that the reader has to keep her actual orientation in mind. Moreover, if the reader has to generate the current location of a given critical object by means of a mental transformation, she has to determine the object location in the allocentric reference frame given her current orientation (before she initiates mental transformation) by "subtracting" the mental transformation she has performed in order to access the object in front of her (after having finished the mental transformation).

4.3 An empirical investigation

To illustrate our empirical approach we report one of our latest experiments. Our primary goal is to compare access to anchoring objects with critical objects in egocentric mental models. If the egocentric reference frame sets up an overall organizing principle, both types of objects are represented directly within the egocentric reference frame at the time of testing. Therefore, both types of objects are immediately accessible and should exhibit the same latency pattern.

Material. We constructed eight experimental texts which described a protagonist standing in the middle of a room. Along the horizontal plane there were four anchoring objects placed in all four directions of the protagonist (front, behind, left, right). Anchoring objects were not introduced as in (1), taking the protagonist as the relatum. Instead we used sentences like "An der Wand links von Torsten steht der Kühlschrank." ("At the wall to Torsten's left, there is the refrigerator."). Thus, in order for the so-called anchoring object to be localized directly within the egocentric reference frame, it is already necessary for the reader to infer the spatial relation between the refrigerator and Torsten, i.e. that the refrigerator is positioned on the left of Torsten. In addition, four critical objects were introduced, each of them combined with one of the anchoring objects as in (2). For every text there were four versions to systematically vary the direction, in which the object combinations (anchoring and critical object) were placed. The order of mentioning objects was held constant across the versions. The four walls were described one after another, by introducing the anchoring object, followed by the critical object. Subjects read the texts on a PC screen. Reading times were self-paced.

Procedure. In most of the cases, Tversky et al. presented direction probes to their subjects who had to remember the object located in that direction. Of course, we could not adapt this paradigm, because direction probes would be ambiguous in that there are always two objects located in each direction: an anchoring object and a critical object. We therefore adopted the procedure used by Bryant and Tversky (1992), who presented their subjects object probes and subjects had to remember the location of the named object. Our subjects were presented an object name of either an anchoring object or a critical object on the PC screen and had to respond in two successive steps. As soon as they knew the location of the object, they had to press the enter-button (RT1). Afterwards they had to specify the remembered direction by pressing one of four keys indicating front, behind, left and right (RT2). For both steps response times were measured. In the subsequent presentation of the results, we report the latencies for the first step (RT1).

For every text, subjects were probed three times. The first probe followed the description of the object arrangement, with a filler sentence between them. After the first and second probe, the text was continued by mentioning a re-orientation of the protagonist either to his left, to his right or towards his back (in two of all cases the protagonist did not re-orient). Subjects were instructed to answer in accordance with the current orientation of the protagonist and were told that taking the protagonist's perspective would aid in this task. The object names chosen for probes as well as the order in which they were presented were the same for every text and subject. Test items differed only in the direction which had to be remembered, dependent on the text version. For every subject 24 answers were recorded: for both object types all of the four directions were probed three times.

Hypotheses. We predict different patterns of response times for anchoring objects and critical objects. With regard to anchoring objects we suppose that object positions with respect to the egocentric reference frame are inferred at encoding time. Anchoring objects are therefore represented directly within the egocentric reference frame. As was demonstrated by Tversky and colleagues, objects directly located within the egocentric reference frame exhibit access latencies according to the spatial framework pattern. Especially, objects behind should be accessed faster than objects beside (left or right). In contrast we assume critical objects not to be located directly within the egocentric reference

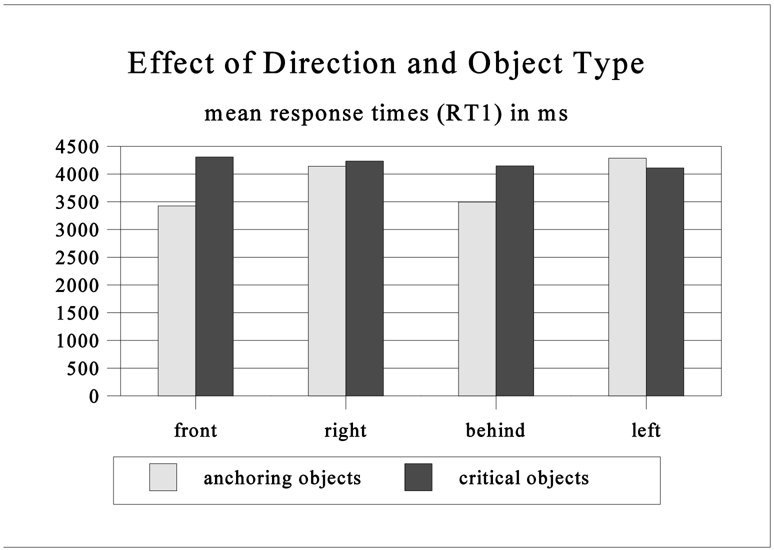

Fig. 7. Mean response times (RT1) for anchoring objects and critical objects as a function of direction

frame, i.e. the location of a critical object with respect to the egocentric reference frame is not inferred online. Critical objects are solely represented within the allocentric reference frame and are not accessible directly within the egocentric reference frame. Our general prediction is that access latencies for critical objects deviate from the spatial framework pattern. Specifically, we suggest that access latencies in this case follow the predictions of the mental transformation pattern, i.e. objects beside are accessed faster than objects behind. For both object types, objects in front should be accessed fastest.

A direct comparison of mean response times for anchoring objects and critical objects should reveal faster responses in case of the former than in the case of the latter. The task of the subjects is to indicate the direction in which an object is currently located. In the case of an anchoring object, the object location is claimed to be inferred at encoding time and updated online when reading about a re-orientation of the protagonist. So, it has only to be read off the mental model. In contrast, the location of a critical object is supposed not to be inferred at encoding time but rather has to be inferred at test time.

Results. Analysis of response times was based on the correct responses (71%) of 28 subjects and limited to RT1. Mean response times were significantly shorter for anchoring objects (3838 ms) than for critical objects (4200 ms) [$F(1, 27) = 10.36$; $p < .01$]. While object direction yielded no overall main effect [$F(3, 81) = 1.98$; $p > .10$], it interacted with object type [$F(3, 81) = 2.69$; $p = .05$]. Separate analyses for both object types revealed an influence of object direction only for anchoring objects [$F(3, 81) = 4.51$; $p < .01$] but not for critical objects [F(3,

81) < 1]. As is illustrated in Fig. 7, for critical objects there was no difference at all in response times as a function of object direction. In contrast, anchoring objects were responded to faster if located in front (3426 ms) or behind (3496 ms) - front being only slightly faster than behind - than if located on the left (4287 ms) or right (4141 ms).

Discussion. Our results clearly indicate different access times for anchoring objects and critical objects. Latencies for anchoring objects mainly exhibit the spatial framework pattern in that objects located on the front/back axis are responded to faster than on the left/right axis. However, objects behind are judged as fast as are objects in front, a finding not conforming to the expected pattern. While the dominance of the front/back axis over the left/right axis did affect latencies, the asymmetry of the front/back axis yielded no influence. This lack of influence in the last case might be due to the higher complexity of the mental models in our case (eight objects in four directions) as compared with those of Tversky et al. (five objects in five directions). Yet the effectiveness of the dominance of the front/back axis speaks in favor of egocentric mental models.

With critical objects, object access did not exhibit the slightest tendency of a spatial framework pattern. As we argued above, access to critical objects does not conform to the predictions of the spatial framework model because critical objects are not directly represented within the egocentric reference frame, and therefore are not immediately accessible. We take this as a strong argument against the egocentric reference frame as an overall organizing principle. The relative positions of the critical objects with respect to the egocentric reference frame are not inferred online. This finding agrees with our assumption that the positions of critical objects are determined with respect to an allocentric reference frame, even in egocentric mental models. Nevertheless, critical objects were obviously not accessed in a manner of a mental transformation. Because for critical objects response times were not influenced by their location at all, we may not ultimately decide if these objects are really accessed within an allocentric reference frame.

On average, anchoring objects are responded to faster than critical objects, as predicted. But since mean response times for critical objects are as high as those obtained for anchoring objects located on the left/right axis, it is difficult to explain the overall disadvantage for the critical objects in terms of an inference necessary at test time. Subjects might have been immediately able to discriminate critical objects from anchoring objects, hence being able to circumvent to try to access them within the egocentric reference frame. But with respect to our initial hypothesis this explantion is not satisfactory. The task of our subjects was to respond with the egocentrically defined location of the probed object. To access the object is not sufficient for this purpose if the object location in the mental model is not egocentrically defined, as we assume in the case of critical objects. Since we expected a time-consuming inference necessary for critical objects, this would mean that they are accessed faster than anchoring objects on the left/right axis, because their positions are assumed to be egocentrically defined.

In summary, we have evidence that for anchoring objects located on the front/back axis their location is easiest to determine. At least for those anchoring objects we maintain that they are immediately accessible within the egocentric reference frame and that their location may be directly read off the mental model, since it is inferred or updated online. We leave open the possibility that the egocentric reference frame is anchored within the allocentric reference frame solely by the objects located on the front/back axis but that there are no anchoring objects on the left/right axis.

5 Conclusion and Further Work

First results of our attempts to experimentally explore and to model mechanisms that allow for a representation of spatial configurations are implemented in the model prototype SPACE/0 (cf. Wiebrock et al. 1998). Since model construction with SPACE/0 does not put any constraints on the selection of arcs that built up the graph, the resulting model may realize multiple perspectives simultaneously. That is, we may represent a spatial scene in a way that cannot possibly be *perceived* under any perspective in the real world. The perceptible part of a spatial configuration is restricted by the viewpoint of an observer. In extending our approach to the construction of depictions (planned as a next step), principles for constructing graphs for a selected visual perspective tied to a given frame of reference have to be specified. Likewise, spatial mental models constructed in text understanding are not restricted to a specific viewpoint in the same way as depictions. Instead - even under an egocentric perspective - the mental model can represent the space surrounding *ego* simultaneously. But the whole surrounding space cannot possibly be visible from an egocentric viewpoint.

Generating depictions from a mental model is closely connected to the problem of inferring spatial relations. In graph representations that allow a "direct" transformation into depictions, all object positions have to be represented with respect to a unique - viewpoint dependent - reference frame. That is, the graph has to contain a unique node (representing the viewpoint), which is directly connected to every other node (representing the objects of a spatial configuration). Such graph representations are more restrictive than the class of mental models constructed by SPACE/0 and the class of egocentric mental models constructed by human text understanders. To transform a mental model into a "depictable" graph, some relations represented explicitly in the mental model have to be deleted and others have to be inferred.

In expanding our model by generating depictions, we hope to gain more precise guidelines for the assessment of mental models concerning spatial inferences as well as for techniques of visualization guided by descriptive terms for several fields of application.

References

Ambler, A. P. and R. J. Popplestone (1975). Inferring the Positions of Bodies from Specified Spatial Relationships. *Artificial Intelligence 6*, 157–174.

Anderson, J. and G. Bower (1973). *Human Associative Memory.* Washington, D.C.: Winston.

Berendt, B. (1996). The utility of mental images: How to construct stable mental models in an unstable image medium. In U. Schmid, J. Krems, and F. Wysotzki (Hrsg.), *Proceedings of the First European Workshop on Cognitive Modeling*, Forschungsberichte des Fachbereichs Informatik, Report No. 96-39, S. 97–103. TU Berlin.

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

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

Bühler, K. (1982). The deictic field of language and deictic words. In R. J. Jarvella and W. Klein (Hrsg.), *Speech, place, and action. Studies in deixis and related topics.* Chichester: John Wiley & Sons Ltd. English reprint. Originally published in 1934.

Franklin, N., L. A. Henkel, and T. Zangas (1995). Parsing surrounding space into regions. *Memory and Cognition 23*(4), 397–407.

Franklin, N. and B. Tversky (1990). Searching imagined environments. *Journal of Experimental Psychology: General 119*, 63–76.

Franklin, N., B. Tversky, and V. Coon (1992). Switching points of view in spatial mental models. *Memory and Cognition 20*(5), 507–518.

Garnham, A. (1985). *Psycholinguistics: Central Topics.* Routledge.

Gips, C., P. Geibel, S. Wiebrock, and F. Wysotzki (1998). Learning Spatial Relations with CAL5 and TRITOP. In preparation.

Glasgow, J. and D. Papadias (1992). Computational Imagery. *Cognitive Science 16*, 355–394.

Guesgen, H. W. (1989). Spatial reasoning based on Allen's temporal logic. Technical Report TR-89-049, ICSI, Berkeley, Cal.

Guesgen, H. W. (1997). Towards hybrid spatial reasoning. In T. P. Martin and A. L. Ralescu (Hrsg.), *Fuzzy Logic in Artificial Intelligence. Towards Intelligent Systems*, LNAI 1188, S. 197–206. Springer. IJCAI'95 Workshop, Montreal CDN, Selected Papers.

Hernández, D. (1991). Relative representation of spatial knowledge: the 2-D case. In D. M. Merk and A. Frank (Hrsg.), *Cognitive and Linguistic Aspects of Geographic Space*, S. 373–385. Dordrecht: Kluwer.

Hörnig, R., B. Claus, and K. Eyferth (1997). Objektzugriff in Mentalen Modellen: Eine Frage der Perspektive. In C. Umbach, M. Grabski, and R. Hörnig (Hrsg.), *Perspektive in Sprache und Raum.* Deutscher Universitätsverlag, Wiesbaden.

Johnson-Laird, P. and R. Byrne (1991). *Deduction.* Hillsdale, NL: Lawrence Erlbaum.

Johnson-Laird, P. N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference and Consciousness.* Cambridge: Cambridge University Press.

Jörding, T. and I. Wachsmuth (1996). An Antropomorphic Agent for the Use of Spatial Language. In *Proceedings of ECAI'96-Workshop on Representation and Processing of Spatial Expressions*, S. 41–53.

Knauff, M., C. Schlieder, R. Rauh, and G. Strube (1998). Mental Models in Spatial Reasoning. This volume.

Lenzmann, B. and I. Wachsmuth (1996). A user–adaptive interface agency for interaction with a virtual environment. In G. Weiss and S. Sen (Hrsg.), *Adaptation and Learning in Multi–Agent Systems*, LNAI 1042, S. 140–151. Springer Verlag.

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

Mittal, N. and A. Mukerjee (1995). Qualitative Subdivision Algebra: moving towards the Quantitative. Technical Report TR 95-23, UT Austin, CS Dept.

Mukerjee, A. and N. Mittal (1995). A qualitative representation of frame-transformation motions in 3-dimensional space. In *Proceedings of the IEEE Conference on Robotics & Automation.*

O'Keefe, J. (1989). Computations the hippocampus might perform. In *Neural connections, mental computation*, S. 225–284. Cambrdige, MA: MIT Press.

Paivio, A. (1971). *Imagery and Verbal Processes.* New York: Holt, Rinehart and Winston.

Pribbenow, S. (1993, Dezember). Räumliche Inferenzen und Bilder. *KI 4*, 7–13.

Renz, J. and B. Nebel (1997). On the Complexity of Qualitative Spatial Reasoning: A Maximal Tractable Fragment of RCC-8. In *Proceedings of the 11th International Workshop on Qualitative Reasoning QR 97.*

Schlieder, C. (1996). A Computational Account of Preferences in Mental Model Construction. In U. Schmid, J. Krems, and F. Wysotzki (Hrsg.), *Proceedings of the First European Workshop on Cognitive Modeling*, Forschungsberichte des Fachbereichs Informatik, Report No. 96-39, S. 90–96. TU Berlin.

Schlieder, C. (1998). Two Approaches to Qualitative Spatial Reasoning: Constraint Satisfaction and Model Construction. Forschungsberichte des Fachbereichs Informatik, Report No. 98-2. Technische Universität Berlin. Results of the Colloquium in Berlin, june, 26.-27. 1997. http://ki.tu-berlin.de/˜sppraum/juni-paper.ps.

Thompson, S. (1991). *Type Theory and Functional Programming.* Reading, Massachusetts: Addison Wesley.

Vorwerg, C. (1997). Kategorisierung von Richtungsrelationen. In E. van der Meer et al. (Ed.), *Experimentelle Psychologie. 39. Tagung experimentell arbeitender Psychologen (TeaP), HU Berlin, 24. - 27.3.97*, Lengerich. Pabst Science Publishers.

Vorwerg, C. and G. Rickheit (1998). Typicality effects in the Categorization of Spatial Relations. This volume.

Waltz, D. and L. Boggess (1979). Visual analog representations for natural language understanding. In *Proc. of the 6th IJCAI*, S. 926–934.

Werner, S., B. Krieg-Brückner, H. A. Mallot, K. Schweizer, and C. Freksa (1997). Spatial cognition: the role of landmark, route, and survey knowledge in human and robots. In *Jahrestagung der Gesellschaft für Informatik, LNCS.*

Wiebrock, S., C. Gips, U. Schmid, and F. Wysotzki (1998). SPACE/0 – Documentation of a Program for Spatial Inference. In preparation.

Wysotzki, F., U. Schmid, and E. Heymann (1997). Modellierung räumlicher Inferenzen durch Graphen mit symbolischen und numerischen Constraints. In C. Umbach, M. Grabski, and R. Hörnig (Hrsg.), *Perspektive in Sprache und Raum.* Deutscher Universitätsverlag, Wiesbaden.