Deutsches
Forschungszentrum
für Künstliche
Intelligenz GmbH

Technical
Memo
TM-91-11

Generating Spatial Descriptions for Cross-modal References

Peter Wazinski

September 1991

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH

Postfach 20 80
D-6750 Kaiserslautern, FRG
Tel.: (+49 631) 205-3211/13
Fax: (+49 631) 205-3210

Stuhlsatzenhausweg 3
D-6600 Saarbrücken 11, FRG
Tel.: (+49 681) 302-5252
Fax: (+49 681) 302-5341

Deutsches Forschungszentrum
für
Künstliche Intelligenz

The German Research Center for Artificial Intelligence (Deutsches Forschungszentrum für Künstliche Intelligenz, DFKI) with sites in Kaiserslautern und Saarbrücken is a non-profit organization which was founded in 1988 by the shareholder companies ADV/Orga, AEG, IBM, Insiders, Fraunhofer Gesellschaft, GMD, Krupp-Atlas, Mannesmann-Kienzle, Philips, Siemens and Siemens-Nixdorf. Research projects conducted at the DFKI are funded by the German Ministry for Research and Technology, by the shareholder companies, or by other industrial contracts.

The DFKI conducts application-oriented basic research in the field of artificial intelligence and other related subfields of computer science. The overall goal is to construct *systems with technical knowledge and common sense* which - by using AI methods - implement a problem solution for a selected application area. Currently, there are the following research areas at the DFKI:

- ❑ Intelligent Engineering Systems
- ❑ Intelligent User Interfaces
- ❑ Intelligent Communication Networks
- ❑ Intelligent Cooperative Systems.

The DFKI strives at making its research results available to the scientific community. There exist many contacts to domestic and foreign research institutions, both in academy and industry. The DFKI hosts technology transfer workshops for shareholders and other interested groups in order to inform about the current state of research.

From its beginning, the DFKI has provided an attractive working environment for AI researchers from Germany and from all over the world. The goal is to have a staff of about 100 researchers at the end of the building-up phase.

Prof. Dr. Gerhard Barth
Director

Generating Spatial Descriptions for Cross-modal References

Peter Wazinski

DFKI-TM-91-11

This work has been supported by a grant from The Federal Ministry for Research and Technology (FKZ ITW-8901 8).

Generating Spatial Descriptions for Cross-modal References

Peter Wazinski[1]

Abstract

We present a localisation component that supports the generation of cross-modal deictic expressions in the knowledge-based presentation system WIP. We deal with relative localisations (e.g., "The object to the left of object X."), absolute localisations (e.g., "The object in the upper left part of the picture.") and corner localisations (e.g., "The object in the lower right corner of the picture"). In addition, we distinguish two localisation granularities, one less detailed (e.g., "the object to the left of object X.") and one more detailed (e.g., "the object above and to the left of object X."). We consider corner localisations to be similar to absolute localisations and in turn absolute localisations to be specialisations of relative localisations. This allows us to compute all three localisation types with one generic localisation procedure. As elementary localisations are derived from previously computed composite localisations, we can cope with both localisation granularities in a computationally efficient way. Based on these primary localisation procedures, we discuss how objects can be localised among several other objects. Finally we introduce group localisations (e.g., "The object to left of the group of other objects.") and show how to deal with them.

[1]Current address: SFB 314, Department of Computer Science, University of Saarbrücken, 6600 Saarbrücken, Germany. Email: wazinski@cs.uni-sb.de.

Contents

1 Introduction

The increasing amount of information to be communicated to users of complex technical systems nowadays makes it necessary to find new ways to present information. Neither the variety of all possible presentation situations can be anticipated nor it is further adequate to present the required information in a single communication mode, such as either text or graphics. Therefore, the automatic generation of multimodal presentations tailored to the individual user has become necessary. Current research projects in artificial intelligence like SAGE ([RMM90]), FN/ANDD ([MR90]), COMET ([FM90]) and WIP ([WABGR91]) reflect the growing interest in this topic.

For the knowledge-based presentation system WIP, the task is the generation of a multimodal document according to the formal description of the communicative intent of the planned presentation and a set of generation parameters. The current scenario for WIP is the generation of instructions for using an espresso-machine. A typical fragment of an instruction manual for an espresso machine is shown in figure 1.

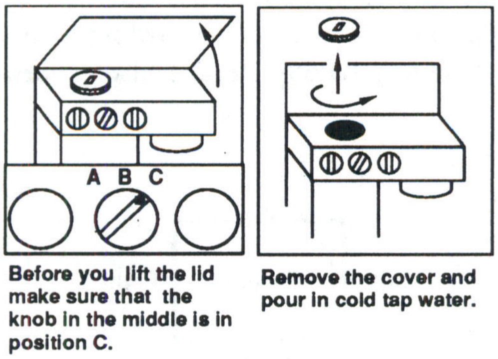

Figure 1: Fragment from an instruction manual

Cross-modal deictic expressions, e.g., "the lid" or "the knob in the middle," help to establish the coreferentiality between the entities mentioned in the text and shown in the picture as well ([WAGR91]). The use of spatial relationships such as "the knob in the middle" simplifies the generation of referring expressions that have to identify a particular object in a picture. Obviously these spatial relationships cannot be computed in advance because they depend on the projection parameters for the picture, e.g., the viewpoint, which in turn themselves depend on the communicative intent of the document to be planned[2].

[2]Even if the projection parameters are constant, it is not feasible to compute all possible relative localisations from a combinatoric point of view.

The localisation component described in this paper was developed in order to support the generation of cross-modal deictic referring expressions. All procedures are fully implemented and were recently integrated into the first WIP prototype. They are coded in Common Lisp and run under Genera 8.0 on a MacIvory. A testbed called LOC-SYS was also developed: it allows the convenient generation and manipulation of rectangle scenes like the examples given in this paper.

Before we describe the methods which underlie the various localisation procedures, in the following section we present our views about localisation phenomena and introduce the terminology used in the rest of this paper.

2 Object localisation

A lot of work has been done on 'object localisation' and its linguistic complement, 'spatial prepositions'. Wunderlich/Herweg ([Wun82], [WH]) and Herskovits ([Her85]) provide linguistic approaches to the semantics of spatial prepositions. NL-systems like NAOS ([NN86]), HAM-RPM ([HHJW80]), SWYSS ([HS84]) and CITYTOUR ([ABHR85],[ABHR86]) address various issues regarding computational aspects. Schirra ([Sch91]) and Habel/Pribbenow ([HP88],[Pri90]) also incorporate relevant work from cognitive psychology.

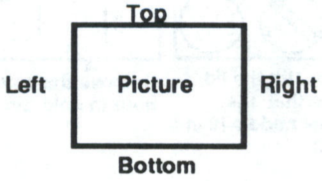

Figure 2: The deictic reference frame

In our approach, we concentrate on the requirements for localising objects in pictures. We assume that the user can see the picture containing the objects to be localised and we do not deal with the problem of anticipating possibly wrong visualisations of the user in the case he/she cannot see the picture. We do not deal with possible intrinsic orientations of depicted objects (c.f. [Ret88]) and assume the deictic reference frame of a common viewer (c.f. figure 2). Together with every localisation, we compute a so-called applicability degree from the intervall [0..1]. The applicability degree is not only used to generate linguistic hedges (c.f. [Lak72]) as in SWYSS or CITYTOUR, but also for selecting the 'best' localisation from a set of alternatives. The localisations computed on our system are two-dimensional localisations in the sense that they are based on the 2D-projection of a picture and not on its possible 3D-representation. In the rest of this section we will describe the localisation phenomena we take into account and introduce our terminology.

2.1 Relative and absolute localisations

The objects shown in part A of figure 3 can be localised as follows:

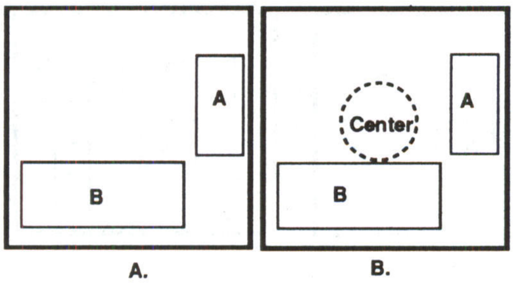

Figure 3: Localising objects in a picture

(1) "Object A is on the right side of the picture."

(2) "Object B is in the lower part of the picture."

(3) "Object A is to the right of Object B."

(4) "Object B is below Object A."

Sentences (1) and (2) are considered to contain **absolute localisations**: an object is localised by stating its *absolute* position in the picture. Sentences (3) and (4) are examples of **relative localisations**: an object is localised by stating its position *relative* to another object. The object to be located will be called the **primary object** (LO for short). The object that serves as reference for locating the primary object is called **reference object** (REFO for short).

How can we explain the similarity between absolute and relative localisations, between "on the right side of the picture" and "to the right of Object B"? Our hypothesis is:

> Absolute localisations are specialisations of relative localisations in the sense that for absolute localisations the center of the picture functions as an implicit reference object.

Part B of figure 3 shows how the absolute localisation of part A can be explained as a relative localisation by assuming a circle-shaped center: "Object A is on the right side of the picure." is equivalent to "Object A is to the right of the center of the picture."

2.2 Elementary and composite localisations

Whereas the unambiguous localisations of the objects in figure 3 could be achieved by naming either the horizontal ("on the right side", "to the right of") or vertical relation ("in the lower part", "below"), figure 4 shows a situation in which it is necessary to give both the horizontal and vertical position of the object with respect to the reference object:

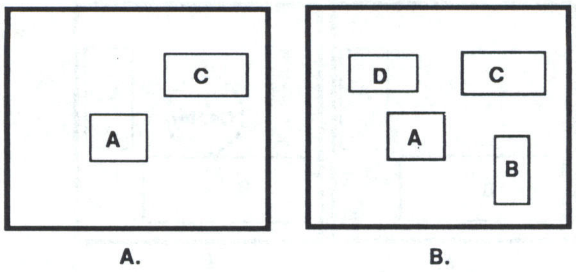

Figure 4: Elementary and composite localisations

In part A of figure 4, it is sufficient to describe object C as the object "to the right of" or "above" object A. But in part B, both descriptions would be ambiguous, because "to the right of" or "above" could refer to object D or B respectively. The only possibility to localise C unambiguously is to describe it as being "above and to the right" of A.

Localisations where either the horizontal or vertical relation is given will be called **elementary localisations**. If both relations are stated together, we will call it a **composite localisation**.

The localisation types introduced so far — absolute vs. relative and elementary vs. composite — are orthogonal. Therefore, an absolute or a relative localisation can be further subcategorized as being an elementary or a composite localisation.

Composite localisations cannot always be applied, e.g., in figure 3 object B cannot be localised as "the object in the lower left part of the picture." Criteria for the applicability of composite localisations will not be examined further in this paper as this would lead to more complex questions, e.g., whether an object can be localised at all. A detailed discussion of these problems is given in [Waz91].

2.3 The construction of the horizontal and vertical reference frame

One important feature of the localisation procedures is the division of the horizontal and vertical reference frame into three parts. The reason for this are 'center'-localisations as shown in figure 5: In all pictures, object A can be localised as the object "in the center of the picture." In order to integrate this observation with the elementary vs. composite distinction we divided the horizontal and vertical dimension into three parts: 'top', 'horizontal center' and 'bottom' and 'left', 'vertical center' and 'top' respectively (c.f. figure 6). Under these conditions the 'center'-localisation in the left part of figure 5 can be analysed as a composite ('vertical center','horizontal center')-localisation. For the picture in the middle it is an elementary 'vertical center'-localisation and for the right one an elementary 'horizontal center'-localisation. When transforming these different localisations into a surface string they all become the same: "in the center of the picture."

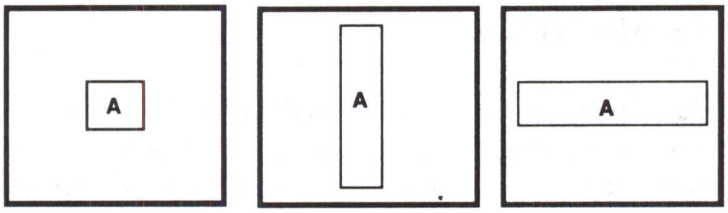

Figure 5: Center localisations

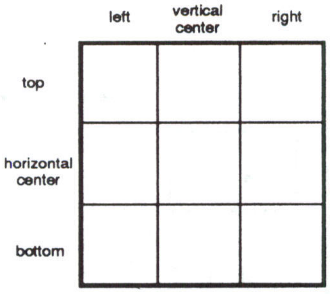

Figure 6: Horizontal and vertical reference frame

Figure 7 shows that it is also useful to adopt the this partition scheme for relative localisations: B would usually be described as the object "to the right of A" and C as the object "above and to the right of A." With respect to the partition scheme a ('right', 'top')-localisation can be applied to C and a ('right', 'horizontal center')-localisation to B. The former matches exactly with the surface string. The latter can be matched with "to the right of A" by assuming that the 'center'-part of a composite localisation is a special part of a composite localisation that does not appear at the linguistic level.

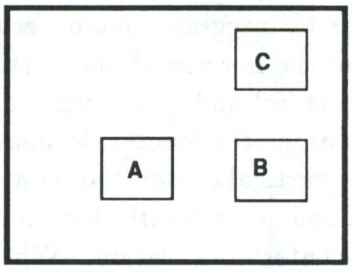

Figure 7: Center localisations and relative localisations

2.4 Corner localisations

An additional localisation type that can be used to localise objects in pictures is the **corner localisation**: if an object is placed in one of the four corner regions of the picture it can be localised as, e.g., "the object in the *left upper corner* of the picture."

The difference between absolute composite localisations and corner localisations is illustrated in figure 8: While object B can be localised as being "in the lower right corner of the picture" it is not possible to use a corner localisation for A. In that case, only "in the left upper part of the picture" could be used. This means that the applicability of a corner localisation implies the applicability of the corresponding absolute composite localisation but not vice versa.

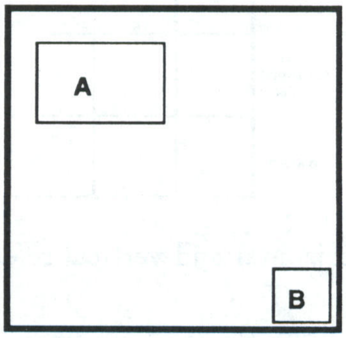

Figure 8: Corner localisations vs. absolute composite localisations

3 Basic localisation procedures

In this section we present matrix-oriented localisation procedures for absolute and relative localisations. As mentioned in section 2.2, both the horizontal and vertical relation of the primary object are given in case of a composite localisation. This suggests that composite localisations are composed of elementary localisations. The procedures presented here, though, behave differently: for the sake of efficiency they compute the composite localisations first and derive the elementary localisations from these previously computed localisation results.

3.1 Absolute localisations

We approximate the center of the picture with a rectangle whose horizontal and vertical extension is one third of the horizontal and vertical extension of the picture. Figure 9 shows the construction of the horizontal and vertical reference system according to the rectangular center region.

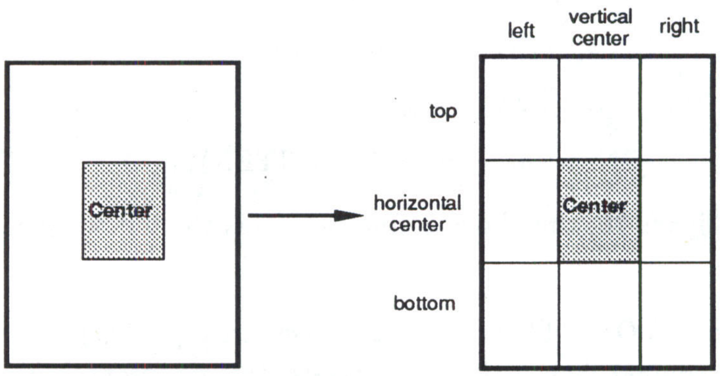

Figure 9: The construction of the horizontal and vertical reference system

Before describing the evaluation function for composite localisations, we give a few definitions:

- The horizontal reference system is abbreviated by XLOC = {left, x-center, right}, the vertical one by YLOC = {top, y-center, bottom}. Composite localisations are denoted by CLOC = XLOC $\times$ YLOC . Both reference systems together are described with ULOC = XLOC $\cup$ YLOC .

- The constant CENTER denotes the center rectangle of a given picture.

- POLY denotes the set of all polygons that can appear in a picture. For given polygons P_1 and P_2 the associative and commutative operator $\cap$, $\_\_ \cap \_\_$: POLY $\times$ POLY $\mapsto$ POLY computes the intersection polygon. The empty polygon is denoted by $P_\emptyset$. The following holds: $\forall P \in$ POLY : $P_\emptyset \cap P = P \cap P_\emptyset = P_\emptyset$.

- The function PR , PR : CLOC $\times$ POLY $\mapsto$ POLY , computes the rectangle corresponding to a given composite localisation and the rectangle partition of the picture induced by a given polygon. For example PR ((left,top), CENTER) computes the upper left rectangle according to the partition scheme shown in figure 9.

- $\Re$ denotes the set of the real numbers. Given a polygon P, the function f, f : POLY $\mapsto \Re$ computes the area of a polygon. It is $f(P_\emptyset) = 0$.

The applicability degree of a composite localisation evaluates how good the position of the object in question is described by that particular localisation. We define the applicability degree as the portion of the area of the object that lies in the rectangle of the picture that corresponds to the composite localisation and the rectangle partition of that picture. Thus we can define A_c as follows:

$$A_c : \text{CLOC} \times \text{POLY} \mapsto \Re$$
$$A_c\,(l, LO) = \frac{f(p)}{f(\text{LO})}$$
$$\text{with} \quad p = PR\,(l, \text{CENTER}) \cap \text{LO}$$

For object LO in figure 10, the above definition yields the following results:

A_c ((left, top), LO) = 1/12, A_c ((x-center, top), LO) = 1/6,
A_c ((left, y-center), LO) = 1/4, A_c ((x-center, y-center), LO) = 1/2 and
A_c (l, LO) = 0 for all other l $\in$ CLOC as $f(p) = f(P_\emptyset) = 0$

For elementary localisations we adopted an analogous definition: the applicability degree A_e of an elementary localisation is determined by the portion of the area of the object that lies in the corresponding row or column of the picture. As already mentioned at the beginning of this section we can write A_e in terms of A_c :

$$A_e^x : \text{XLOC} \times \text{POLY} \mapsto \Re$$
$$A_e^x(l_x, \text{LO}) = \sum_{l_y \in \text{YLOC}} A_c\,((l_x, l_y), \text{LO})$$
$$A_e^y : \text{YLOC} \times \text{POLY} \mapsto \Re$$
$$A_e^y(l_y, \text{LO}) = \sum_{l_x \in \text{XLOC}} A_c\,((l_x, l_y), \text{LO})$$

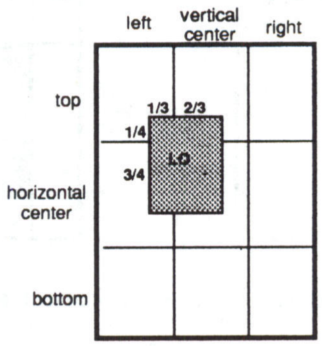

Figure 10: Computing absolute localisations

$$A_e : \text{ULOC} \times \text{POLY} \mapsto \Re$$
$$A_e\,(l, \text{LO}) = \begin{cases} A_e^x(l, \text{LO}) & \text{if } l \in \text{XLOC} \\ A_e^y(l, \text{LO}) & \text{if } l \in \text{YLOC} \end{cases}$$

A_e^x and A_e^y compute the applicability for the horizontal and vertical dimension by summing up the applicability degrees of the corresponding composite localisations. They are combined in A_e order to have a function that is defined on both dimensions, i.e., ULOC .

With respect to figure 10 we get.

A_e (top, LO) = A_c ((left, top), LO) + A_c ((x-center, top), LO) = 1/4,
A_e (y-center, LO) = A_c ((left, y-center), LO) + A_c ((x-center, y-center), LO) = 3/4,
A_e (left, LO) = A_c ((left, top), LO) + A_c ((left, y-center), LO) = 1/3,
A_e (x-center, LO) = A_c ((x-center, top), LO) + A_c ((x-center, y-center), LO) = 2/3.

As argued in paragraph 2.4 corner localisations are similar to composite ('left'/'right', 'top'/'bottom')-localisations, but less general. This property can be modelled by corner regions that are smaller than the corner regions for absolute localisations. In turn, these corner regions correspond to a larger center as shown in figure 11. Thus we can compute corner localisations just by changing the size of the center.

Instead of 1/3 as for absolute localisations we take 4/5 of the horizontal and vertical extension of the picture for the extended center.

3.2 Relative localisations

The localisation procedure for relative localisations is similar to the one for absolute localisations. One major difference is that now the construction of the horizontal and vertical reference frame is done with respect to a given reference object and not to the

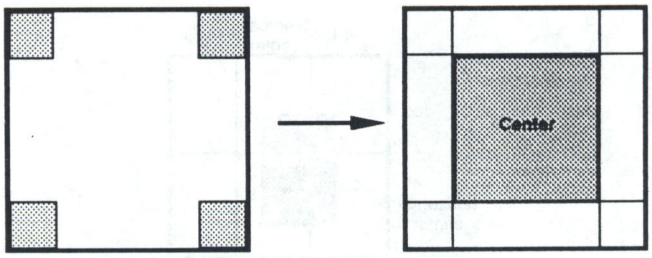

Figure 11: The relation between corner and center regions

implicit assumed center of the picture (c.f. figure 12). The second difference concerns the computation of the applicability degree: for relative localisations, not only the portion of an area is taken into account, but also the distance between the primary object and the reference object.

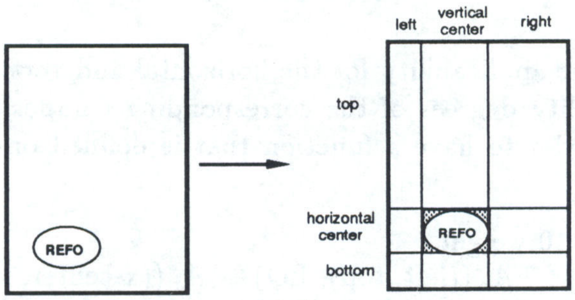

Figure 12: The construction of the reference frame for relative localisations

The basic idea for the evaluation of the distance between primary object and reference object is adopted from the CITYTOUR system: first we compute the center of gravity for the primary object. Then we determine its coordinates with respect to the reference system established by the reference object. Finally we use these coordinates for the computation of the applicability degree. Figure 13 illustrates the various factors that affect the applicability of an 'above'-localisation:

1. The applicability degree decreases with an increasing vertical distance. In Part A of figure 13 the applicability degree for "P_1 is above REFO" is greater than for "P_2 is above REFO."
2. The applicability degree decreases with an increasing horizontal distance. In Part B the applicability degree for "P_3 is above REFO" is greater than for "P_4 is above REFO."

3. If the horizontal and vertical distances increase by the same amount, then the applicability degree decreases more with the increasing horizontal distance than with the increasing vertical distance. This is shown in Part C: the applicability degree for "P_6 is above REFO" is greater than for "P_7 is above REFO", although the vertical distance between P_5 and P_6 and the horizontal distance between P_5 and P_7 are equal.

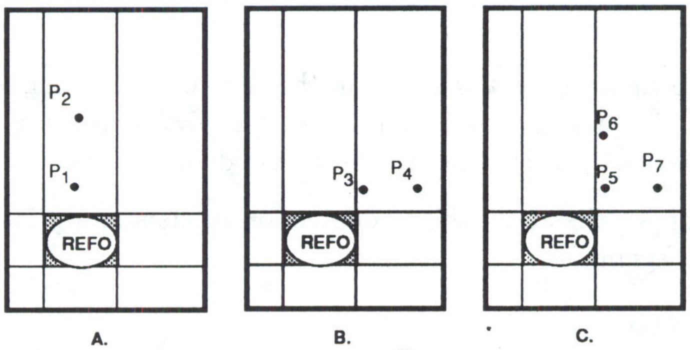

Figure 13: Evaluating the distance of a point

Let *eval* denote the function that evaluates the distance between a point and a rectangle according to the criteria mentioned above. Let further POINT denote the set of all points within a picture and RECT $\subseteq$ POLY the set of all rectangular polygons. Then the signature of *eval* can be written as[3]:

$$eval : \text{CLOC} \times \text{POINT} \times \text{RECT} \mapsto \Re$$

Now we are almost able to define the function A_c , which computes the applicability degree of a composite localisation. Let CG , CG : POLY $\mapsto$ POINT , compute the center of gravity for a polygon and let further SR , SR : POLY $\mapsto$ RECT , compute the smallest surrounding rectangle for a polygon. Then the applicability degree A_c of a composite

[3] In reality *eval* is slightly more complicated because it maps into $\Re \times \Re$ and not only into $\Re$. The reason for this is that the different evaluation of increasing vertical and horizontal distances can result in different evaluations for points to which both a horizontal or vertical localisation can be applied. E.g., P_7 in figure 13 would get a different evaluation for an 'above'- than for a 'right of'-localisation. We abstract from this detail in order to make the principle of the procedure clearer.

localisation can be defined as:

$$A_c \ : \text{CLOC} \ \times \text{POLY} \ \times \text{POLY} \ \mapsto \Re$$

$$A_c\ (l, \text{LO}, \text{REFO}) = w\ eval\ (l, CG\ (p), SR\ (\text{REFO}))$$

$$\text{with} \quad p = PR\ (l, \text{REFO}) \cap \text{LO}$$

$$w = \frac{f(p)}{f(\text{LO})}$$

p is the part of the primary object that lies in the rectangle corresponding to the composite localisation l. The factor w weighs the result of *eval* according to the portion of the area of the primary object that lies in the rectangle corresponding to l.

Now the definition of A_e, the applicability degree for an elementary localisation, can be given in terms of A_c again:

$$A_e^x \ : \text{XLOC} \ \times \text{POLY} \ \times \text{POLY} \ \mapsto \Re$$
$$A_e^x\ (l_x, \text{LO}, \text{REFO}) = \sum_{l_y \in \text{YLOC}} A_c\ ((l_x, l_y), \text{LO}, \text{REFO})$$
$$A_e^y \ : \text{YLOC} \ \times \text{POLY} \ \times \text{POLY} \ \mapsto \Re$$
$$A_e^y\ (l_y, \text{LO}, \text{REFO}) = \sum_{l_x \in \text{YLOC}} A_c\ ((l_x, l_y), \text{LO}, \text{REFO})$$
$$A_e \ : \text{ULOC} \ \times \text{POLY} \ \times \text{POLY} \ \mapsto \Re$$
$$A_e\ (l, \text{LO}, \text{REFO}) = \begin{cases} A_e^x\ (l, \text{LO}, \text{REFO}) & \text{if} \quad l \in \text{XLOC} \\ A_e^y\ (l, \text{LO}, \text{REFO}) & \text{if} \quad l \in \text{YLOC} \end{cases}$$

This means that the applicability degree A_e for a primary object LO is the sum of the composite localisations for the corresponding row or column of the reference frame.

For figure 14 we get the following results:

$$\begin{aligned} & A_c\ ((\text{x-center}, \text{top}), \text{LO}, \text{REFO}) \\ = \ & \tfrac{1}{3}\ eval\ ((\text{x-center}, \text{top}), P_1, SR\ (\text{REFO}) = \tfrac{1}{3} * 0.7 = 0.23 \\ & A_c\ ((\text{right}, \text{top}), \text{LO}, \text{REFO}) \\ = \ & \tfrac{2}{3}\ eval\ ((\text{right}, \text{top}), P_2, SR\ (\text{REFO}) = \tfrac{2}{3} * 0.65 = 0.43 \\ & A_c\ (l, \text{LO}, \text{REFO}) \\ = \ & 0 \text{ for all other l} \in \text{CLOC as } w = \frac{f(p)}{f(\text{LO})} = \frac{0}{f(\text{LO})} = 0 \end{aligned}$$

$$
\begin{aligned}
& A_e\,(\text{top}, \text{LO}, \text{REFO}) \\
= \; & A_c\,((\text{x-center}, \text{top}), \text{LO}, \text{REFO}) + A_c\,((\text{right}, \text{top}), \text{LO}, \text{REFO})
\end{aligned}
$$

$$
\begin{aligned}
& A_e\,(\text{right}, \text{LO}, \text{REFO}) \\
= \; & A_c\,((\text{right}, \text{top}), \text{LO}, \text{REFO})
\end{aligned}
$$

$$
\begin{aligned}
& A_e\,(\text{x-center}, \text{LO}, \text{REFO}) \\
= \; & A_c\,((\text{x-center}, \text{top}), \text{LO}, \text{REFO})
\end{aligned}
$$

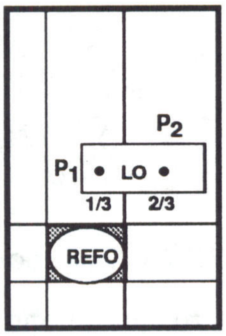

Figure 14: Computing relative localisations

4 A generic localisation procedure for absolute and relative localisations

The similarities between the localisation procedures discussed in the previous section allow us to design one generic localisation procedure that can be specialised to a procedure for absolute, relative or corner localisations.

Given the primary object LO and the reference object REFO the first step is to determine the 3×3 matrix $M^{\cap}$, which contains the intersection polygons of LO and the partial rectangles in the picture with respect to REFO. For relative localisations, REFO varies, for absolute localisations and corner localisations the parameter is set to either the normal or the extended center area (c.f. section 3.1).

$$M^{\cap}_{x,y} = PR\,((x,y), \text{REFO}) \cap \text{LO} \qquad \text{for all} \quad x \in \text{XLOC}\;, y \in \text{YLOC}\;.$$

The second step is the computation of the evaluation matrix M^A, which contains the applicability degrees of the composite localisations. The computation requires a function E, $E : \text{POLY} \times \text{POLY} \times \text{POLY} \mapsto \Re$. E corresponds exactly to the function A_c for absolute and relative localisations in section 3.1 and 3.2. The only difference results from the previous computation of $M^{\cap}$: the subexpression $p = PR\ ((x,y), \mathrm{REFO}) \cap \mathrm{LO}$ is factored from A_c and therefore computed only once.

$$M^A_{x,y} = E(M^{\cap}_{x,y}, \mathrm{LO}, \mathrm{REFO})$$

The third step is the computation of the elementary localisations. The vector $\vec{X}$ contains the evaluations of the horizontal localisations and $\vec{Y}$ the ones for the vertical localisations:

$$\vec{X}_x = \sum_{y \in \mathrm{YLOC}} M^A_{x,y}$$
$$\vec{Y}_y = \sum_{x \in \mathrm{XLOC}} M^A_{x,y}$$

This means that we have $\vec{X}_{l_x} = A_e\ (l_x)$ for l_x in XLOC and $\vec{Y}_{l_y} = A_e\ (l_y)$ for l_y in YLOC .

Finally, we can determine the best composite and elementary localisation and their applicability degrees by computing the maximum value of $M^{\cap}$ and $\vec{X}$ or $\vec{Y}$ respectively.

5 Localising objects in a complex scene

In the previous sections we considered pictures with a minimal number of objects. In order to deal with more complex object configurations the localisation procedures presented above have to be extended. The new task is no longer "Localise LO with respect to REFO!" but "Given a set of REFO candidates, choose the best one for LO!"

In order to reduce the search space for REFO candidates, first a kind of 'between'-test is applied to the set of possible reference objects. The idea behind this test is that an exclusion procedure based on simple geometric overlapping tests can be performed more efficiently than a comparison of applicability degrees that have to be computed by the rather complex localisation procedures. An example is given in figure 15: When searching for a suitable reference object for object A in figure 15, object D would be ruled out because object B is found in the 'between'-area of A and D.

The determination of the best reference object raises the problem of ambiguity. Not only is the applicability degree of a localisation important, but also whether the use of the reference object would result in an ambiguous localisation. In Part A of figure 16 object D could be localised as being either "above A" or "to the right of D." But the first localisation is ambiguous because both, C and D, are "above A."

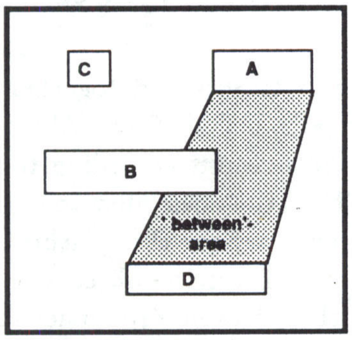

Figure 15: Search space reduction for complex object configurations

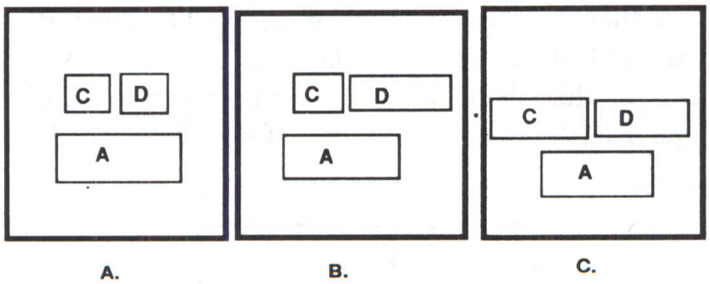

Figure 16: Ambiguous reference objects

With respect to elementary and composite localisations we distinguish three cases of ambiguity:

1. In Part A of figure 16, the localisation of object C or D would be ambiguous with respect to A because for both objects the composite localisations, (x-center, top), are equal.
2. In Part B a composite localisation cannot be applied to object D (neither "D is above and to the right of A" nor "D is immediately above A" are adequate) and its elementary localisation, 'top', is part of the composite localisation, (x-center, top), of object C.
3. In Part C a composite localisation can be applied neither to C nor to D and their elementary localisations, 'top', are equal.

6 Localising groups of objects

Control knobs and switches are often grouped together in a control panel in order to provide for easier operation of technical devices. Moreover spatially adjacent objects can also be grouped as one perceptual unit according to the 'law of the good gestalt' in Gestalt psychology ([MW78]). Thus the possibility to generate localisations with respect to a given group structure is neccessary for the "naturalness" of a localisation. Besides this, group localisations are also useful if the objects in the immediate neighbourhood of the primary object have exactly the same properties (c.f. [WJH78]). In this case, the primary object can be localised with respect to its group and has not to be localised with respect to the whole scene, which could have resulted in an ambiguous localisation.

For our localisation procedures this means that groups can function as a reference object as well as a primary object. In addition, objects can be localised absolutely with respect to the group in which they are contained in. In figure 17 object B would be localised as the object "to the right of the group of the other objects." Vice versa we can say "The group of objects to the left of object B" and we can localise object A as being "the object in the upper left region of the group of objects to the left of B."

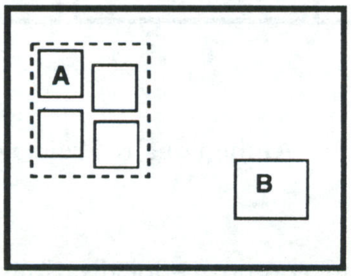

Figure 17: Group localisations

The last example also illustrates the hierarchical character of group localisations: An object can be localised absolutely within a group. This group might be localised again within a surrounding group or — if there is none — this group can localised relatively to another (group of) object(s).

The algorithm for group localisations cannot detect group hierarchies. Instead it expects a tree representation of the group hierarchy as an input. The output consists of two parts: According to the depth of the group tree the algorithm computes a chain of absolute localisations. In addition the outermost surrounding group of the primary object is localised relatively to an optional (group of) reference object(s).

7 Conclusions

We have introduced a unifying approach for absolute, relative and corner localisations of objects in pictures. In addition, the use of a special partition scheme for the reference frame of a preposition allows us to deal with two different localisation granularities for absolute and relative localisations. By defining the evaluation functions for elementary localisations in terms of the evaluation functions for the corresponding composite localisations, we have been able to design one procedure that handles all three localisation types and both localisation granularities efficiently. Furthermore, we have given a solution to the problem of localising an object within a complex configuration on the basis of this localisation procedure. Finally, we have shown how our system deals with group localisations.

References

[ABHR85] E. André, G. Bosch, G. Herzog, T. Rist: CITYTOUR – Ein natürlichsprachliches Anfragesystem zur Evaluierung räumlicher Präpositionen. Abschlußbericht des Fortgeschrittenenpraktikums, Department of Computer Science, University of Saarbrücken, 1985.

[ABHR86] E. André, G. Bosch, G. Herzog, T. Rist: Characterising Trajectories of Moving Objects using Natural Language Path Descriptions. In: *Proc. of the 7th ECAI*, pp. 1–8, 1986.

[FM90] S. K. Feiner, K. R. McKeown: Coordinating Text and Graphics in Explanation Generation. In: *Proc. 8th AAAI*, pp. 442–449, 1990.

[Her85] A. Herskovits: Semantics and Pragmatics of Locative Expressions. *Cognitive Science*, 9:341–378, 1985.

[HHJW80] W. v. Hahn, W. Hoeppner, A. Jameson, W. Wahlster: The Anatomy of the Natural Language Dialogue System HAM-RPM. In: L. Bolc (ed.): *Natural Language Based Computer Systems*, pp. 119–254. München: Hanser, 1980.

[HP88] C. Habel, S. Pribbenow: Gebietskonstituierende Prozesse. LILOG-Report 18, IBM Germany, 1988.

[HS84] M. Hußmann, P. Schefe: The Design of SWYSS, a Dialogue System for Scene Analysis. In: L. Bolc (ed.): *Natural Language Communication with Pictorial Information Processing.* München: Hanser McMillan, 1984.

[Lak72] G. Lakoff: Hedges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts. In: J.N. Levi, G.C. Phares (eds.): *Papers from the 8th regional Meeting of the Chicago Linguistics Society*, pp. 183–228. University of Chicago, Chicago, IL, 1972.

[MR90] J. Marks, E. Reiter: Avoiding Unwanted Conversational Implicatures in Text and Graphics. In: *Proc. 8th AAAI*, pp. 450–455, 1990.

[MW78] G.M. Murch, G.L. Woodworth: *Wahrnehmung.* Stuttgart: Kohlhammer, 1978.

[NN86] B. Neumann, H.-J. Novak: NAOS: Ein System zur natürlichsprachlichen Beschreibung zeitveränderlicher Szenen. *Informatik Forschung und Entwicklung*, pp. 83–92, 1986.

[Pri90] S. Pribbenow: Interaktion von propositionalen und bildhaften Repräsentationen. In: C. Habel, C. Freksa (eds.): *Repräsentation und Verarbeitung rämlichen Wissens*, pp. 156–174. Berlin: Springer, 1990.

[Ret88] G. Retz-Schmidt: Various Views on Spatial Prepositions. *AI Magazine*, 9(2):95–105, 1988.

[RMM90] S. Roth, J. Mattis, X. Mesnard: Graphics and Natural Language as Components of Automatic Explanation. In: J. W. Sullivan, S. W. Tyler (eds.): *Intelligent User Interfaces*, pp. 207–239. Reading, MA: Addison Wesley, 1990.

[Sch91] J. Schirra: A Contribution to the Reference Semantics of Spatial Prepositions: The Visualization Problem and its Solution in VITRA. In: *Proceedings of the IAI Workshop "On the Semantics of Prepositions in Natural Language Processing.* Mouton, de Gruyter, to appear 1991. Also available as Technical Report 75, SFB 314, Department of Computer Science, University of Saarbrücken.

[WABGR91] W. Wahlster, E. André, S. Bandyopadhyay, W. Graf, T. Rist: WIP: The Coordinated Generation of Multimodal Presentations from a Common Representation. In: O. Stock, J. Slack, A. Ortony (eds.): *Computational Theories of Communication and their Applications.* Berlin: Springer, to appear 1991.

[WAGR91] W. Wahlster, E. André, W. Graf, T. Rist: Designing Illustrated Text: How Language Production is Influenced by Text and Graphics. In: *Proc. 5th Conf. of the European Chapter of the Association for Computational Linguistics (EACL)*, pp. 8–14, 1991.

[Waz91] P. Wazinski: Objektlokalisation in graphischen Darstellungen. Master's thesis, Universität Koblenz-Landau, Abt. Koblenz/DFKI Saarbrücken, 1991.

[WH] D. Wunderlich, M. Herweg: Lokale und Direktionale. In: A. v. Stechow, D. Wunderlich (eds.): *Handbuch der Semantik*. Königstein Ts.: Athenäum Verlag, forthcoming.

[WJH78] W. Wahlster, A. Jameson, W. Hoeppner: Glancing, Referring and Explaining in the Dialougue System HAM-RPM. *American Journal of Computer Linguistics, Microfiche 77*, pp. 53–67, 1978.

[Wun82] D. Wunderlich: Sprache und Raum. *Studium Linguistik*, 12:1–19, 1982.

Deutsches
Forschungszentrum
für Künstliche
Intelligenz GmbH

DFKI
-Bibliothek-
PF 2080
6750 Kaiserslautern
FRG

DFKI Publikationen

Die folgenden DFKI Veröffentlichungen oder die aktuelle Liste von erhältlichen Publikationen können bezogen werden von der oben angegebenen Adresse.
Die Berichte werden, wenn nicht anders gekennzeichnet, kostenlos abgegeben.

DFKI Publications

The following DFKI publications or the list of currently available publications can be ordered from the above address.
The reports are distributed free of charge except if otherwise indicated.

DFKI Research Reports

RR-90-01
Franz Baader: Terminological Cycles in KL-ONE-based Knowledge Representation Languages
33 pages

RR-90-02
Hans-Jürgen Bürckert: A Resolution Principle for Clauses with Constraints
25 pages

RR-90-03
Andreas Dengel, Nelson M. Mattos: Integration of Document Representation, Processing and Management
18 pages

RR-90-04
Bernhard Hollunder, Werner Nutt: Subsumption Algorithms for Concept Languages
34 pages

RR-90-05
Franz Baader: A Formal Definition for the Expressive Power of Knowledge Representation Languages
22 pages

RR-90-06
Bernhard Hollunder: Hybrid Inferences in KL-ONE-based Knowledge Representation Systems
21 pages

RR-90-07
Elisabeth André, Thomas Rist: Wissensbasierte Informationspräsentation:
Zwei Beiträge zum Fachgespräch Graphik und KI:
1. Ein planbasierter Ansatz zur Synthese illustrierter Dokumente
2. Wissensbasierte Perspektivenwahl für die automatische Erzeugung von 3D-Objektdarstellungen

24 pages

RR-90-08
Andreas Dengel: A Step Towards Understanding Paper Documents
25 pages

RR-90-09
Susanne Biundo: Plan Generation Using a Method of Deductive Program Synthesis
17 pages

RR-90-10
Franz Baader, Hans-Jürgen Bürckert, Bernhard Hollunder, Werner Nutt, Jörg H. Siekmann: Concept Logics
26 pages

RR-90-11
Elisabeth André, Thomas Rist: Towards a Plan-Based Synthesis of Illustrated Documents
14 pages

RR-90-12
Harold Boley: Declarative Operations on Nets
43 pages

RR-90-13
Franz Baader: Augmenting Concept Languages by Transitive Closure of Roles: An Alternative to Terminological Cycles
40 pages

RR-90-14
Franz Schmalhofer, Otto Kühn, Gabriele Schmidt: Integrated Knowledge Acquisition from Text, Previously Solved Cases, and Expert Memories
20 pages

RR-90-15
Harald Trost: The Application of Two-level Morphology to Non-concatenative German Morphology
13 pages

RR-90-16
Franz Baader, Werner Nutt: Adding Homomorphisms to Commutative/Monoidal Theories, or: How Algebra Can Help in Equational Unification
25 pages

RR-90-17
Stephan Busemann: Generalisierte Phasenstrukturgrammatiken und ihre Verwendung zur maschinellen Sprachverarbeitung
114 Seiten

RR-91-01
Franz Baader, Hans-Jürgen Bürckert, Bernhard Nebel, Werner Nutt, Gert Smolka: On the Expressivity of Feature Logics with Negation, Functional Uncertainty, and Sort Equations
20 pages

RR-91-02
Francesco Donini, Bernhard Hollunder, Maurizio Lenzerini, Alberto Marchetti Spaccamela, Daniele Nardi, Werner Nutt: The Complexity of Existential Quantification in Concept Languages
22 pages

RR-91-03
B.Hollunder, Franz Baader: Qualifying Number Restrictions in Concept Languages
34 pages

RR-91-04
Harald Trost: X2MORF: A Morphological Component Based on Augmented Two-Level Morphology
19 pages

RR-91-05
Wolfgang Wahlster, Elisabeth André, Winfried Graf, Thomas Rist: Designing Illustrated Texts: How Language Production is Influenced by Graphics Generation.
17 pages

RR-91-06
Elisabeth André, Thomas Rist: Synthesizing Illustrated Documents A Plan-Based Approach
11 pages

RR-91-07
Günter Neumann, Wolfgang Finkler: A Head-Driven Approach to Incremental and Parallel Generation of Syntactic Structures
13 pages

RR-91-08
Wolfgang Wahlster, Elisabeth André, Som Bandyopadhyay, Winfried Graf, Thomas Rist: WIP: The Coordinated Generation of Multimodal Presentations from a Common Representation
23 pages

RR-91-09
Hans-Jürgen Bürckert, Jürgen Müller, Achim Schupeta: RATMAN and its Relation to Other Multi-Agent Testbeds
31 pages

RR-91-10
Franz Baader, Philipp Hanschke: A Scheme for Integrating Concrete Domains into Concept Languages
31 pages

RR-91-11
Bernhard Nebel: Belief Revision and Default Reasoning: Syntax-Based Approaches
37 pages

RR-91-12
J.Mark Gawron, John Nerbonne, Stanley Peters: The Absorption Principle and E-Type Anaphora
33 pages

RR-91-13
Gert Smolka: Residuation and Guarded Rules for Constraint Logic Programming
17 pages

RR-91-14
Peter Breuer, Jürgen Müller: A Two Level Representation for Spatial Relations, Part I
27 pages

RR-91-15
Bernhard Nebel, Gert Smolka: Attributive Description Formalisms ... and the Rest of the World
20 pages

RR-91-16
Stephan Busemann: Using Pattern-Action Rules for the Generation of GPSG Structures from Separate Semantic Representations
18 pages

RR-91-17
Andreas Dengel & Nelson M. Mattos: The Use of Abstraction Concepts for Representing and Structuring Documents
17 pages

RR-91-19
Munindar P. Singh: On the Commitments and Precommitments of Limited Agents
15 pages

RR-91-20
Christoph Klauck, Ansgar Bernardi, Ralf Legleitner FEAT-Rep: Representing Features in CAD/CAM
48 pages

RR-91-22
Andreas Dengel: Self-Adapting Structuring and Representation of Space
27 pages

RR-91-23
Michael Richter, Ansgar Bernardi, Christoph Klauck, Ralf Legleitner: Akquisition und Repräsentation von technischem Wissen für Planungsaufgaben im Bereich der Fertigungstechnik
24 Seiten

RR-91-24
Jochen Heinsohn: A Hybrid Approach for Modeling Uncertainty in Terminological Logics
22 pages

RR-91-25
Karin Harbusch, Wolfgang Finkler, Anne Schauder: Incremental Syntax Generation with Tree Adjoining Grammars
16 pages

RR-91-26
M. Bauer, S. Biundo, D. Dengler, M. Hecking, J. Koehler, G. Merziger:
Integrated Plan Generation and Recognition
- A Logic-Based Approach -
17 pages

RR-91-27
A. Bernardi, H. Boley, Ph. Hanschke, K. Hinkelmann, Ch. Klauck, O. Kühn, R. Legleitner, M. Meyer, M. M. Richter, F. Schmalhofer, G. Schmidt, W. Sommer:
ARC-TEC: Acquisition, Representation and Compilation of Technical Knowledge
18 pages

RR-91-30
Dan Flickinger, John Nerbonne:
Inheritance and Complementation: A Case Study of *Easy* Adjectives and Related Nouns
39pages

DFKI Technical Memos

TM-89-01
Susan Holbach-Weber: Connectionist Models and Figurative Speech
27 pages

TM-90-01
Som Bandyopadhyay: Towards an Understanding of Coherence in Multimodal Discourse
18 pages

TM-90-02
Jay C. Weber: The Myth of Domain-Independent Persistence
18 pages

TM-90-03
Franz Baader, Bernhard Hollunder: KRIS: Knowledge Representation and Inference System -System Description-
15 pages

TM-90-04
Franz Baader, Hans-Jürgen Bürckert, Jochen Heinsohn, Bernhard Hollunder, Jürgen Müller, Bernhard Nebel, Werner Nutt, Hans-Jürgen Profitlich: Terminological Knowledge Representation: A Proposal for a Terminological Logic
7 pages

TM-91-01
Jana Köhler: Approaches to the Reuse of Plan Schemata in Planning Formalisms
52 pages

TM-91-02
Knut Hinkelmann: Bidirectional Reasoning of Horn Clause Programs: Transformation and Compilation
20 pages

TM-91-03
Otto Kühn, Marc Linster, Gabriele Schmidt: Clamping, COKAM, KADS, and OMOS: The Construction and Operationalization of a KADS Conceptual Model
20 pages

TM-91-04
Harold Boley:
A sampler of Relational/Functional Definitions
12 pages

TM-91-05
Jay C. Weber, Andreas Dengel , Rainer Bleisinger: Theoretical Consideration of Goal Recognition Aspects for Understanding Information in Business Letters
10 pages

TM-91-08
Munindar P. Singh: Social and Psychological Commitments in Multiagent Systems
11 pages

TM-91-09
Munindar P. Singh: On the Semantics of Protocols Among Distributed Intelligent Agents
18 pages

TM-91-10
Béla Buschauer, Peter Poller, Anne Schauder, Karin Harbusch: Tree Adjoining Grammars mit Unifikation
149 pages

TM-91-11
Peter Wazinski: Generating Spatial Descriptions for Cross-modal References
21 pages

DFKI Documents

D-89-01
Michael H. Malburg, Rainer Bleisinger:
HYPERBIS: ein betriebliches Hypermedia-Informationssystem
43 Seiten

D-90-01
DFKI Wissenschaftlich-Technischer Jahresbericht 1989
45 pages

D-90-02
Georg Seul: Logisches Programmieren mit Feature -Typen
107 Seiten

D-90-03
Ansgar Bernardi, Christoph Klauck, Ralf Legleitner: Abschlußbericht des Arbeitspaketes PROD
36 Seiten

D-90-04
Ansgar Bernardi, Christoph Klauck, Ralf Legleitner: STEP: Überblick über eine zukünftige Schnittstelle zum Produktdatenaustausch
69 Seiten

D-90-05
Ansgar Bernardi, Christoph Klauck, Ralf Legleitner: Formalismus zur Repräsentation von Geo-metrie- und Technologieinformationen als Teil eines Wissensbasierten Produktmodells
66 Seiten

D-90-06
Andreas Becker: The Window Tool Kit
66 Seiten

D-91-01
Werner Stein , Michael Sintek: Relfun/X - An Experimental Prolog Implementation of Relfun
48 pages

D-91-03
Harold Boley, Klaus Elsbernd, Hans-Günther Hein, Thomas Krause: RFM Manual: Compiling RELFUN into the Relational/Functional Machine
43 pages

D-91-04
DFKI Wissenschaftlich-Technischer Jahresbericht 1990
93 Seiten

D-91-06
Gerd Kamp: Entwurf, vergleichende Beschreibung und Integration eines Arbeitsplanerstellungssystems für Drehteile
130 Seiten

D-91-07
Ansgar Bernardi, Christoph Klauck, Ralf Legleitner
TEC-REP: Repräsentation von Geometrie- und Technologieinformationen
70 Seiten

D-91-08
Thomas Krause: Globale Datenflußanalyse und horizontale Compilation der relational-funktionalen Sprache RELFUN
137 pages

D-91-09
David Powers and Lary Reeker (Eds):
Proceedings MLNLO'91 - Machine Learning of Natural Language and Ontology
211 pages
Note: This document is available only for a nominal charge of 25 DM (or 15 US-$).

D-91-10
Donald R. Steiner, Jürgen Müller (Eds.)
MAAMAW'91: Pre-Proceedings of the 3rd European Workshop on „Modeling Autonomous Agents and Multi-Agent Worlds"
246 pages
Note: This document is available only for a nominal charge of 25 DM (or 15 US-$).

D-91-11
*Thilo C. Horstmann:*Distributed Truth Maintenance
61 pages

D-91-12
Bernd Bachmann:
$\mathrm{H}^{\mathrm{iera}}\mathrm{C}_{\mathrm{on}}$ - a Knowledge Representation System with Typed Hierarchies and Constraints
75 pages

D-91-13
International Workshop on Terminological Logics
Organizers: Bernhard Nebel, Christof Peltason, Kai von Luck
131 pages

D-91-14
Erich Achilles, Bernhard Hollunder, Armin Laux, Jörg-Peter Mohren: 𝒦ℛℐ𝒮 : 𝒦nowledge ℛepresentation and ℐnference 𝒮ystem
- Benutzerhandbuch -
28 Seiten

Generating Spatial Descriptions for Cross-modal References

Peter Wazinski

TM-91-11

Technical Memo