

Connecting Visual and Verbal Space: Preliminary Considerations Concerning the Concept 'Mental Image'

Jörg R.J. Schirra

SFB 314, VITRA
Fachbereich 14 - Informatik
Universität des Saarlandes
D-6600 Saarbrücken 11
Federal Republic of Germany
e-mail: joerg@cs.uni-sb.de

November 30, 1992

Abstract

AI research concerning the connection between seeing and speaking mainly employs what is called reference semantics. Within this framework, the notion of 'mental image' is often used while explaining how somebody not situated in the same perceptual context is able to anchor his understanding of an utterance describing the scene visually perceived by the speaker. We give a foundation for considering mental images as propositions with respect to a certain *field* of concepts: these fields have to provide a *syntactically dense* set of concepts distinguishing locations. The use of such propositions in the reference semantic explanations of understanding utterances about visually perceived scenes is motivated by applying Kant's idea of the introduction of new types of objects: we conceive spatial relations as relations only applicable to *sortal objects*, i.e., individuated objects which are *synthetically introduced* on a syntactically dense field providing their potential locations. The concept 'mental image' which results from these preliminary studies is applied to two current projects in AI, one dealing with the semantics of particular spatial prepositions, and the other more generally concerned with the logic of the connection between visual and verbal space.

Paper for the 4th *International Workshop on Time, Space, Movement,
and Spatio-Temporal Reasoning*, September 4–8, 1992, Bonas, France

Connecting Visual and Verbal Space

Preliminary Considerations Concerning the Concept ‘Mental Image’

Jörg R.J. Schirra, Saarbrücken, FRG

1 Using Mental Images: a Naive Framework

The relation between language and world has attracted the interest of scientists since the beginning of science. Nowadays AI encounters a more restricted version of this question when addressing the problem of the connection between space in its visual and verbal forms. A typical example from our ordinary life is the task of a radio sports reporter: he has to give to his audience a verbal description of spatio-temporal configurations (among other aspects) which he accesses by his visual sense. Usually, the spatial entities to be considered are intuitively classified into those concerning the relations between two or more concrete objects – *spatial relations* – and those additionally integrating temporal aspects – *spatial events*. The reporter’s behavior is often explained by means of the *concept of ‘reference semantics’*, i.e., his report is viewed as primarily anchored in his perceptions. The combination of this understanding with the communicational aspect of language leads us to the conception of *mental images*.

Following G.H. Mead, one central aspect of conscious communication is that in the speaking individual, the same reaction is triggered as is in the other individuals (cf. [Mea62, p. 68ff.]): the speaker has to adopt the role of the others in the communicative act. In order to be language, what is said has to be understood by the speaker, as well, and he must be influenced by it in the same way as the others. This is essential since speakers never mention explicitly everything actually communicated: the phenomena of ellipses and anaphora, presuppositions and conversational implicature are just the tips of the iceberg. In AI, partner modeling provides the corresponding explanatory tool. Therefore, we should not examine the reporter without his communicative counterpart, the audience. Radio sports report listeners are involved in cognitive activities approximately converse to the speaker: they have to understand what the reporter has said. However, the conception of reference semantics cannot be simply transferred in order to explain their understanding (and the behavior resulting from it), since the situational setting explicitly excludes any perception of the reported events by the audience: as is indicated by Fig. 1, this problem is usually solved by means of the concept of ‘mental images’ (cf. [Sch93]). In order to explain how a listener understands the report which is grounded in the visual perception of the speaker, the listener is assumed to construct a *visual mental model* which substitutes percepts of objects not perceptually present: as a German linguist wrote in 1969, ‘*the radio reporter has solved his task only if he describes the reality of a sports event so vividly and obviously to the listener that the listener believes he sees that reality*’ (cf. [Dan69, p. 94]). Following Piaget’s detailed classification of mental images, we are concerned with *anticipatory kinetic images*, i.e., images of situations and movements which were not perceived before (cf. [PI66, Chapter 1, § 1]).

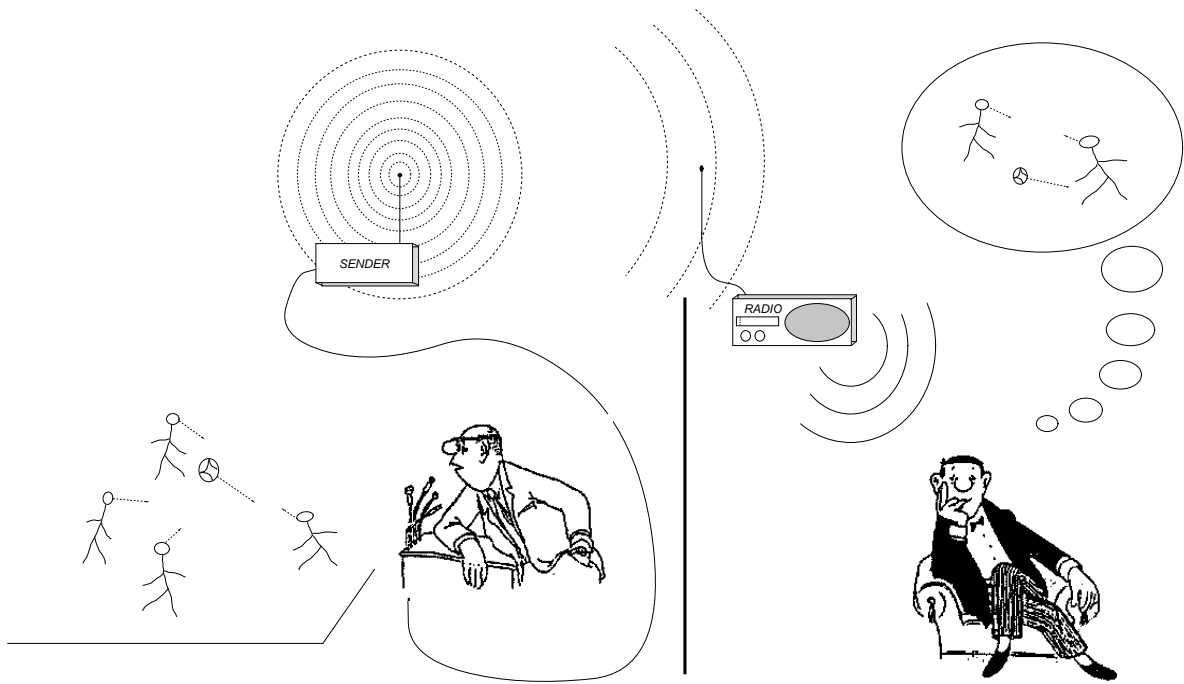


Figure 1: Mental Images as Explanations

In the following, I elaborate this approach to explaining the connection between visual (perceptual) and verbal (communicative) aspects of space by means of the concept of a ‘mental image’.

2 The Logical Components of an Utterance

We first concentrate on the verbal aspect of space which not only mediates between speaker and hearer, but also is usually viewed to be interpersonally accessible – in contrast to percepts or mental images. While performing his task a radio sports reporter utters primarily singular declarative sentences¹ about concrete things, sentences like (S 1) and (S 2):

S 1 *Miller is not standing in the right half of field.*

S 2 *Miller is doing a double pass with Smith.*

Following the constructivist analysis of logic ([KL73]), singular declarative phrases are logically divided into four parts: one is called the involved set of *nominators* which refer to some given (i.e., already mutually known) individual objects, like ‘Miller’ or ‘the right half of field’. The expression ‘nominator’ was introduced in [Lor70, p. 214] with the particular aim of obtaining a more distinct separation between logical and linguistical analyses.² In comparing the

¹Here, ‘singular’ is opposed to ‘general’: singular sentences concern single objects; general sentences use quantified variables with respect to a domain of objects;

²Four types of nominators are distinguished: demonstrator (‘this’), indicators (‘today’, ‘you’), proper names (‘Miller’), and definite descriptions which use a predicator to specify the object (either in combination with the demonstrator (deictical description; ‘this ball’) or pure (‘the ball’)); cf. [KL73, III.8–9];

logical and the linguistic terms, the nominators correspond mainly to the noun phrases of the sentence.

The second part is called the *predicator* (cf. [Car47, p. 6]). The function of predicators is to introduce a standard gauge with respect to which the considered objects are rated: a dimension of distinction to be communicated – i.e., proposed to be newly established in the interpersonal discourse universe. In the examples above, we employ the predicators ‘to stand in’ and ‘to do a double pass with’. In most cases, the predicators are linguistically given mostly by means of the predicates of such a sentence (in a broad sense, sometimes with adjectives and adverbs).

A *logical copula* is the third part of any assertion (cf. [KL73, I.4, and p. 90]): its function is to bind the nominators as arguments of the predicator, and perform the assertion. The logical copulae have to be clearly distinguished from the linguistic copula which essentially has the function of an unspecific verb binding adjectives as predicates to a subject, e.g., in English the verb ‘to be’ (at least in some uses). Traditionally, two copulae are studied in logic, corresponding to the two sides of a (binary) distinction, and formally expressed by the symbols ε for ascribing the predicator to that set of nominators, and ε' for denying it. Copulae represent the minimal logical residue of the performative aspect of an asserting utterance – they also reflect the corresponding reactions in that language game: accepting or refusing the assertion.

Finally, sentences are uttered in a particular *situational context* which determines their meaning: in particular the nominators cannot be understood without a context:³ following Strawson’s explications of the reference relation, a nominator does not simply represent an object by means of a one-to-one relation, but it *picks out* an object from a certain given contextual domain of objects of the discourse universe, therefore conceptually requiring a one-to-many relation (cf. [Str71, p. 17ff.] and [Tug76, p. 369ff.]). Since nominators correspond to those objects *mutually known* by speaker and hearer, contexts depend crucially on what has been communicated earlier (cf. [Joh76, p. 50f.]). Along this *horizontal dimension* of explanation (as I want to call it), contexts are viewed mainly as the comprehension of the text up until that moment (cf. [Kam90]). In the case of the radio sports reporter, the context of his utterances is alternatively conceived as the set of things which he visually perceives. In general, the minimal logical residue of the context of an assertion is represented by the (usually implicit) self-referencing psychological component of that utterance – e.g., ‘I see that’ + (S 1) – which basically establishes the relation to the speech situation by means of the reference to the speaker (“I”) and his understanding of being situated there. Here, the expression *vertical dimension* is used to explain contexts by referring to the visual field of the speaker.

In this paper, the representation of an utterance by explicitly and distinctly mentioning its four logical components is called the *proposition* corresponding to that utterance. Thus, the proposition – or logical form – of an utterance is a particular kind of abstraction. Thus, the proposition – or logical form – of an utterance is a particular kind of abstraction, like ‘the verbal manifestation’ or ‘the syntactical structure’ of an utterance (cf. [KL73, III.7]), an abstraction mainly used to explain the understanding of such an utterance. We use the scheme given in formula 1 for propositions: as a variable for copulae, we use the Greek character κ and parentheses around the whole propositional schema; the context is given in the abstract schema by means of the Greek capital letter Δ (reminder for ‘domain of objects’). Formula 2 expresses the propositional transcription of assertion (S 1):

³Correspondingly, Kamp speaks of *the need to replace the older absolute notion of propositions by that of a proposition relative to a context*. (cf. [Kam86, p. 13f.]);

$$(\{n_1, n_2, \dots, n_i\} \quad \kappa \quad P)_\Delta \quad (1)$$

$$(\{Miller, the\ right\ half\ of\ field\} \quad \varepsilon' \quad being\ in)_{(Radio\ reporter\ A, Time\ \tau)} \quad (2)$$

3 Predicators, Concepts, and Levels of Explanations

Explaining the understanding of an utterance is essentially based on its logical components: the context provides (horizontally or vertically) the multitude of objects among which the nominators identify some. The predicator draws the listener's attention to a distinction applicable to the selected objects, e.g., restricting their attributes or relations. The copula requests him to revise the context appropriately. This revision results in a new context for the subsequent utterance which simultaneously corresponds to the listeners' understanding of the utterance considered. We especially may assume that all implicit information and implicatures drawn by the audience are elaborated in the process of the revision. In general, the subsequent sentence may use the newly established distinctions to specify its objects by means of definite descriptions. The other kinds of nominators are derived from this referring to an already established distinction (cf. [Tug76, p. 326ff.], [Joh76, p. 50f.]).

We want to explain how mental images are (could or should be) used to explain the function of at least one of the logical components of an utterance like (S 1) or (S 2); i.e., we want to explain why a certain kind of explanation of the understanding of such an utterance is adequate. Since the utterances in our focus are characterized by their *spatial* predicators, we concentrate on this component. How can mental images be used to explain the revision process on the context triggered by the spatial predicators?

In order to explain the use of a verbal expression, generally the *meaning* of the expression is mentioned: the meaning of a predicator is called a *concept* – an interpersonally accessible abstract reference point which allows us to rate the validity of the communicated distinction (cf. [Ros90]). Assertions which are used to discuss *about* concepts – i.e., assertions with a nominator referring to a concept and a predicator highlighting or introducing attributes of that concept – are called *methodological* with respect to that concept. Sentences with a predicator corresponding to that concept are called *empirical* (with respect to the concept). An important use of methodological sentences with respect to spatial concepts is called *Spatial Reasoning* (or *Spatio-temporal Reasoning*, respectively).⁴ The empirical sentences with respect to spatial concepts – like (S 1) and (S 2) – form the *Object Level*. The corresponding methodological sentences declare distinctions of the abstract reference points of the Object Level, i.e., relations and attributes of spatial concepts. The set of concepts specified here outlines the verbal aspect of space: therefore, this level of sentences is called the *Spatial Level*. Rather obviously, such conclusions drawn by Spatial Reasoning have to be founded, as well: i.e., we have to consider methodological sentences of a higher level: methodological sentences with respect to

⁴We here use this expression for syllogism-like combinations of a methodological sentence about spatial concepts with corresponding empirical sentences, e.g., the following rather simple conclusion:

$$\frac{(\{the\ church, the\ book\ shop\} \varepsilon\ Being\ Left), (\{Being\ Left, Being\ Right\} \varepsilon\ Being\ Converse)}{(\{the\ book\ shop, the\ church\} \varepsilon\ Being\ Right)}$$

predications of concepts. This is called the *Methodological Level* (methodological in the closer sense, namely independent of a specific concept). The core of this level is clearly the concept of ‘concept’ as such: we can use the properties which we commonly ascribe to a concept by virtue of being a concept for founding the explanatory power of methodological sentences about that concept, e.g., sentences on the Spatial Level used in Spatial Reasoning. Our question as to the explanatory power of mental images belongs to this level: we are interested in the relation between the concept ‘mental image’ and the concept ‘concept’.

Due to the restriction on the length of this paper, we cannot go into the illuminating historical development of the contemporary concept of ‘concept’ from the ancient understanding as an eternal Form, through the scholastic conception of the three types of universals – before, in, and after the particulars – and the explanation of concepts as an autonomously created private mental entity to be found in the Philosophy of Enlightenment, although we conceive them as highly relevant for our theme. A corresponding discussion with respect to Spatial Reasoning is given in greater detail in [Sch94, Sect. 4], a general investigation is to be found in [Ros90]. In the following, we solely pick out essential components of Kant’s conception seen from a more modern non-mentalistic perspective.

4 Concepts and Mental Images

The relation between the concepts ‘concept’ and ‘mental image’ originates essentially from the mentalistic framework of the Philosophy of Enlightenment: *‘the word ‘concept’ (or other words used as equivalent) was conceived as referring to a mental phenomenon which is generated autonomously by the human mind, directly accessible in the ‘inner world’ (or consciousness) of the human being, and which should enable the human beings who have it to prove the validity of predicative statements’* (cf. [Ros90, Vol. II, p. 11]). In the dawn of this position, Descartes and especially Locke understood a concept to be a mental image, or more precisely, a *prolongated perception* of a corresponding particular which serves as a *prototype* for similar particulars. However, this interpretation ran quickly into severe problems (cf. [Ros90, Vol. II, p. 55ff.]). Integrating parts of this idea with Leibniz’s conception of a concept to be a human faculty, i.e., a *mental program* for recognizing corresponding instances, Kant in the heydays of the Philosophy of Enlightenment presented an elaborated theory of a two-fold mental construction: first, he considers a human faculty of constructing concepts which, second, themselves are mental faculties to construct *intuitions*, i.e., mental representatives of instances, or more colloquially: mental images (cf. [Kan65, B741f./A713f., A105, and B180/A141]).

Kant’s second step, the construction of mental images of instances, was resuscitated in contemporary Cognitive Science by Johnson-Laird under the name of *mental models* (cf. [JL83]). Fig. 2 illustrates the connection to the explanation given above: in the mentalistic tradition, the context of an utterance from which we start is interpreted as a mental image (or model); the nominators of the utterance under investigation are expected to refer to elements of that image; its predicator is used to communicate a distinction with respect to a concept – that is here, a certain mental faculty. By means of this faculty, the contextual mental image is transformed into (the image of) a concrete instance of the concept. Therefore, all implicatures of the application of the corresponding distinction in the given context have to be present in the resulting model. But they are not yet explicitly accessible. This can only be achieved by means of the concepts corresponding to those implied distinctions: they have to be applied to the mental image in order to recognize explicitly the implicatures, and thereby also transform the image into the representative of the context of the following sentence (cf. [Pri91]; the recognition

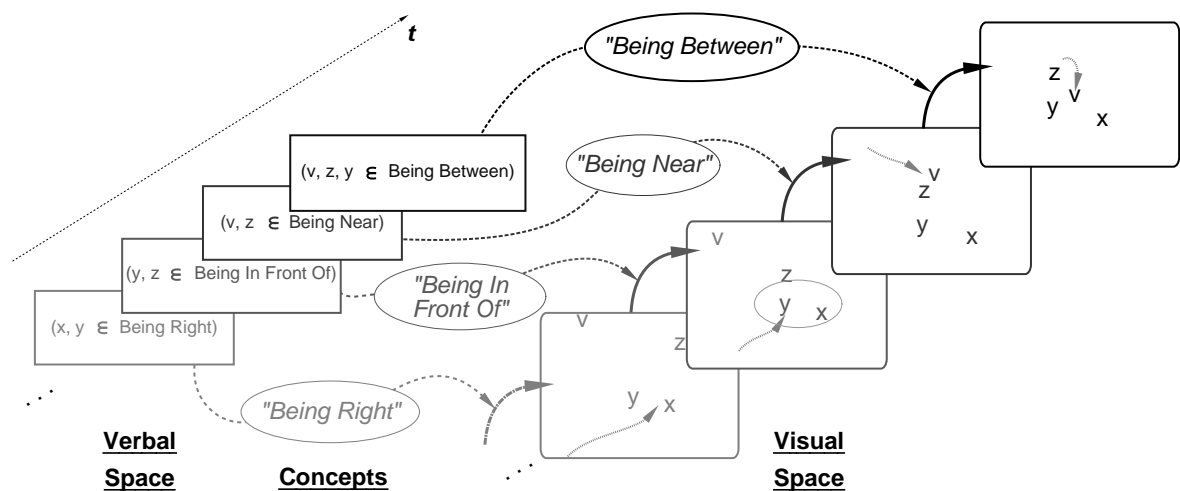


Figure 2: Concepts Mediating Between Verbal and Visual Space

of the implicatures is not shown in Fig. 2).

Though, in order to avoid the solipsistic consequences of conceiving concepts as private (i.e., explicitly *not* interpersonally accessible) mental entities, we should follow Wittgenstein (cf. [Wit53, §§ 656 and 25] and [Ros90, Vol. III, p. 35ff.]): the mentalistic terminology of the Philosophy of Enlightenment has to undergo a *linguistic turn*: mental phenomena have to be reinterpreted as a particular form of methodological communicative behavior used to explain other verbal behavior. ‘*We describe the technique of using words by means of words*’ (cf. [Wit89, p. 128]). Concepts should be understood as abstractions of verbal behavior which are used to interpret, explain, and interpersonally coordinate this kind of behavior. Instead of being mental entities existing before the corresponding verbal behavior, they are derived from it, and therefore secondary to it.

The important question remains: where do these faculties to construct or revise mental models come from, and what do they look like. Kant’s answer is, as was mentioned above: they also are autonomously created – or synthesized – by the human mind.

5 The Making of a Distinct Concept

By ‘synthetically creating a concept’, Kant does not mean the definition of a concept on a given set of elementary concepts which all distinguish properties of a given type of object, i.e., concepts within one *field*.⁵ In this case, the composed concept is already completely *performed* in the set of elementary concepts: it is merely a symbolic abbreviation for a particular combination of the elementary distinctions, since in fact no new kind of distinction is introduced. Kant had in mind a totally different kind of introducing concepts, which really *extends* the capacity to recognize: the ‘*making of a distinct concept*’ instead of merely ‘*making a concept distinct*’ (cf. [Kan60, Introduction, § VIII]). He wants to *synthesize* a completely new field of concepts, i.e., concepts

⁵A field of concepts – also called an incompatibility domain – puts together concepts which are applicable to one type of objects but cannot be applied simultaneously to a particular object of that type; colors form such an incompatibility domain: a one-colored object cannot be simultaneously red and green; in fields of graded concepts, like the spatial field, usually also the incompatibility is a matter of degree;

distinguishing the attributes of a completely new kind of objects, by combining several given but unconnected fields of concepts. Like the construction of mental models, this construction has to be adjusted to a reference point, as well: in constructing a field of concepts, we use, so to speak, a description of our goal – a ‘blueprint’ – which specifies the internal relations between the concepts of that field, and thus, restricts its possible realizations by means of a synthesis of other fields of concepts which are already given (in Kant’s terminology: ‘real’; cf. [ibidem]). For example, the introduction of the rational numbers as a combination of two (sets of) integers (counter and denominator) or of the complex numbers as a pair of real numbers should be viewed as a synthesis in the sense of Kant. All fields of concepts have an internal structure: we with Kant are now mainly interested in the external relations between the internal structures of several fields. Kant calls the specification of the relations between the concepts within one field its *logical rules*, e.g., in the form of an axiomatic system generating that field. The set of rules connecting several fields to create a new field is called the (*transcendental*) *schema* of the new field (cf. [Kan65, B174/A135]).

Feeling obliged to make the linguistic turn, we shift the focus of our attention from the construction of a concept understood as a private mental entity to the corresponding explanations we could give for the explanative power of a concept conceived as an abstraction of verbal behavior: in order to explain why the concepts of a certain field (e.g., the field of spatial concepts) can be used to explain the utterances of the Object Level (e.g., by means of Spatial Reasoning) we could remain within that field of concepts, employing merely the internal definitions and logical rules of that field: because the concepts of ‘being right’ and ‘being left’ are converse, we can conclude that, if an object is to the left of another object, the latter is to the right of the former. In this case, the conversivity of the two concepts cannot be explained, as well. However, applying Kant’s understanding, we additionally could consider the schema of that field which then is viewed as receiving its internal structure from other fields of concepts: the logical rules of the former can be reduced to the logical rules of the latter. With respect to the example above, this allows us to even give an explanation for the conversivity of ‘being left’ and ‘being right’: it has to be a consequence of the particular combination of the concepts of the simpler fields with their characteristic logical structures. Correspondingly, *judgments a priori*, which in Kant’s terminology correspond approximately to propositions of methodological utterances – e.g., ‘The concept of ‘being in’ is transitive in the following cases ...’ – are distinguished with respect to the involved kind of foundation in analytic and synthetic judgments. *Analytic judgments a priori* are founded by means of the logical rules within one field of concepts – its *internal* structure. *Synthetic judgments a priori* integrate additionally the schema connecting several such fields in their explanation, i.e.: synthetic judgments a priori are propositions which have to be founded by explaining the logical rules of a concept of one field *externally* by the logical rules of the concepts of other primarily independent fields by means of the schema.⁶ With a synthetic judgments a priori, like ‘This concept has those properties’, a certain *genetic* aspect is contributed to argumentations about the attributes of a concept: the claim is laid for a foundation by means of a sentence like ‘*because we conceive this concept in its field as constructed in such a way (schema!) from concepts with such attributes (logical rules).*’

There exists a close parallel between the distinction between logical rules and schemata of a field of concepts on the one side, and the distinction between *specification* and *implementation* of an

⁶This is a rather simplified presentation of Kant’s concept of ‘synthetic judgments a priori’; for a detailed description: cf. [Ros91], and even more elaborated in [Ros90, Vol. II, Chapter 3];

Abstract Data Type (ADT) on the other side: the specification of an ADT, e.g., by means of a set of algebraic equations (cf. [EM85]), declares axiomatically the attributes of and relations between data types internal to an ADT. The implementation furthermore reduces an ADT to one or more others: data types of the implemented ADT are projected to data types of the implementing ADTs: this is the ‘schema’ of the implementation, so to speak. By means of this schema, the specification of the implemented ADT can be *founded* on the specifications of the implementing ADTs: the equivalent of a methodological sentence in the theory of ADTs, like, for example, ‘Multiplication of any rational with the rational multiplicative identity unit results in the same rational’ can be understood both analytically – because it follows from the specification of that ADT *Rationals* – or synthetically – because it is a direct consequence of the particular combination of two (sets of) elements of the ADT *Integers* to one element of *Rationals* and the definition of the rational multiplication on such tuples by means of the operations of *Integers*.

6 The Horizontal Dimension of Explanation

We now can interpret the sentences of the Spatial Level (i.e., the methodological sentences with respect to spatial concepts, like the rules of transitivity of the concept ‘being in’ or the rules of conversivity between the concepts of the projective prepositions) as analytic judgments expressing the internal structure of that field. We also may use this level primarily to logically explain the adaptation of the context resulting from a new utterance: to that purpose, we describe the context – i.e., what we assume to be the common knowledge of speaker and hearer – by a set of sentences empirical with respect to spatial concepts.⁷ Methodological sentences (among the logical rules of the Spatial Level) corresponding to the predicator of the new utterance are selected and used to add further statements to the context in the syllogism-like manner of Spatial Reasoning, thus making explicit the implicatures of the utterance in that context. This kind of explanation is often called misleadingly ‘propositional’ in Cognitive Science (cf. [Pri91, (e.g. Sect. 2.2.1)]). We call it the *horizontal dimension of explanation*, since the context is based totally on the analysis of what was said before, and its revision takes place within merely one field of concepts. Note that, for example, the ‘empirical’ rules of transitivity for the use of the concept ‘being in’ given by Vieu (for the contemporary French; cf. [Vie91, p. 225ff.]) can be used exactly in this manner without considering any other field of concepts: the objects of the spatial field have different possible attributes which govern the transitivity of ‘being in’.

7 The Vertical Dimension of Explanation

We also may interpret the sentences of the Spatial Level as synthetic judgments: for example, we may say that the concept ‘being in’ is in certain cases transitive and in other not, *because* it is introduced – or implemented, if we use the terminology of ADTs – in a particular way on concepts of other fields with their characteristic internal structures or specifications. [Vie91] again provides a very good example to demonstrate this kind of explanation, and simultaneously to focus on the two fields of concepts which we conceive as crucial for implementing the Spatial Level, as well: the field of configurational Gestalt concepts (geometrical level), and the field of

⁷In fact, a context should be conceived as one compound proposition, logically composed from a set of elementary propositions; we ignore here the characteristics of the logical junctors involved, since they do not contribute any important features to the present discussion;

functional part-whole concepts (functional level). As was said above, all concepts of a field fit to a certain kind of object the attributes of which they distinguish: this kind of object constitutes the whole field. Therefore, the synthetic combination of several fields to a new field corresponds to the establishing of a new type of object combining or merging attributes of the objects underlying the implementing fields of concepts. Following Vieu, spatial concepts like ‘being in’ only can distinguish relations between objects which have both perceptible configurations and non-perceptible functions. The rules of transitivity mentioned above are used as the core of the ‘blue-print’ for implementing the spatial field.

Vieu’s field of geometrical concepts distinguishes objects called ‘individuals’ (adopted from Clarke’s calculus of individuals) by means of their configurations. Although logically including Euclidean geometry, this field is too weak for the specification of the spatial field; e.g., it is not rich enough to show analytically the particularly restricted transitivity of ‘being in’. In contrast to the classical Euclidean geometry based on the primitive object ‘point’, in Vieu’s specification of the geometrical level, points of time and space are derived from the basic concept ‘individual’ which approximately corresponds to the concept ‘Gestalt’ in Gestalt psychology – an undivided whole distinguished as figure from the ground.

Vieu’s field of functional relations spans around the concept of an object which consists of functional parts or is a functional part of other objects. The relation between a ring of gold and the substance of gold used in the ring, the relation between a forest and its trees, the relation between a car and its constructive parts all exemplify different versions of the functional part-whole relations considered on this level. Note that such a particular whole may stay the same although some of its parts may change.

The spatial field centers around the concept of an object which both has configurational properties and is involved in functional part-whole relations: every object of the spatial field can be projected onto an object of the functional field and an individual of the geometrical field (cf. [Sch94, Sect. 4.2.3.h]). This also means that all the parts such an object has due to the relations of its projection to the functional level also have a configurational counterpart on the geometrical level. In fact, by means of the synthesis of these two fields of concepts, Vieu introduces as the backbone of the spatial field a concept of ‘object’ equivalent to Strawson’s concept of ‘*sortal universals*’, i.e., that kind of object we usually have in mind when we use the expression ‘concrete object’ in the close sense: perceptible, countable, persistent over time even if not perceived (cf. [Str71], [Tug76, p. 453], [Sch94, Sect. 4.3]). Sortal concepts like ‘chair’, ‘car’, or ‘human being’ are distinguished from concepts like ‘fog’, ‘red’, ‘fast’, ‘water’, ‘gold’. As a characteristic, sortal concepts contain a criterion to identify and distinguish different individual objects of the same sort, and thus firstly enable us to point to one such individual *as an individual*, or to count several of them: sortal concepts *individuate* their instances. Similarly to two red objects which are not distinguishable already by their being red alone, the functional parts of a car, for example, do not already distinguish one car clearly from another one, since they both have – within a certain range – the same functional structure, and are therefore functionally undistinguishable. Only the different geometrical components of two instances of ‘car’, or with Vieu’s words: their different *histories*, allow us to distinguish both, e.g., by means of pointing. On the other hand, it is not the mere configuration which makes something a car, but the functional relations between the configurational parts.

Founding the properties of spatial concepts synthetically thus means to explain them by means of the interaction of the properties of the geometrical and the functional field. We call this aspect of explanation the *vertical dimension*, since the synthesis *constructs higher*, i.e., more

complicated fields of concepts, from simpler ones. Any set of propositions – or context – on the spatial field of concepts can be vertically explained as a synthesis of a set of propositions on the geometrical field with a set of propositions on the functional field: each spatial proposition predicating on a sortal object is projected to configurational propositions predicating on the – perceptible – histories of the sortal objects, and functional propositions predicating on its mereological relatives.

The geometrical level provides those concepts used to describe the (essentially visually) perceptible attributes of sortal objects, aspects which are assumed to be already commonly given, that is interpersonally established. Common between whom? Remember that we here are involved in the endeavor to determine the concept of ‘perception’ and ‘communication’ used by somebody who explains somebody else what the exemplary radio reporter is doing: this person while founding synthetically the concepts of the spatial field which he ascribes to the radio reporter uses the geometrical level as a field of concepts he and his vis-à-vis already agree upon. In this case, the interpretation of a context of the geometrical field as a corresponding context of the spatial field (with an appropriate functional presupposition) can be viewed as an explanation of perception: the geometrical field provides in this case the visual aspects of space. Moreover: since the projection of a spatial context onto the geometrical level results in a set of propositions concerning the configurational relations between individuals, we should look for arguments to take this projection to be an *image*.

8 Images and the Syntax of the Pictorial

Why are we inclined to call contexts of the geometrical field images, but not the contexts on the functional or spatial fields? The relevant criterion presented by N. Goodman ([Goo68]) restricts the set of concepts in a field, i.e., the set of predicators used there, and the relations between them. In other words, the structures of contexts possible on that field, that is, the *syntax* governing corresponding propositions (in a very general sense of syntax; cf. [Sch91, p. 88]) gives us the clue for the pictorial: contexts of such fields with a characteristic syntax are images.

Goodman suggests to conceive *syntactical density* of the system of predicators expressing the geometrical field as an essential and for our purpose (though not generally) sufficient criterion for pictorial systems (cf. [Goo68, p. 226]). A system of predicators is called syntactically dense ‘if it provides for infinitely many characters [i.e., predicators] so ordered that between each two is a third’ (cf. [Goo68, p. 136]). Scholz ([Sch91, p. 97]) remarks that this does not mean that any single context of such a field contains an unbound number of propositions with infinitely many predicators; it only is necessary that – on the Methodological Level – the field as such provides the potential of an unrestricted number of densely ordered concepts some of which are used in a particular context. With that, two individual contexts may be similar to any degree without being identical. This is true for ordinary pictures since for example any spot of color in a picture may be moved a very small distance to get another similar image: no matter how small the distance is, we always can think of a picture with that dot shifted only half the distance. In consequence, the proper application of such a predicator is not strictly decidable in the computational sense: after a finite time, the ‘conceptual neighborhood’ of concurrent predicators is reduced but not eliminated.

Vieu’s geometrical level logically includes Euclidean Geometry: that is, its specification

permits to define the concept of a zero-dimensional point with the corresponding attribute: location. Since the Euclidean concept of a location corresponds in fact to the concepts of the densely ordered real numbers, they form a syntactically dense field of concepts: for example, for any pair of predicators ‘having location x ’ and ‘having location y ’ (with the ‘internal’ real numbers x and y), the predicator ‘having location $\frac{(x+y)}{2}$ ’ is also in the field and, in the above sense, *between* the former two.⁸ In contrast to that, the set of predicators expressing the functional field is not densely ordered: the mereological predicators can be clearly distinguished, as is the case for the set of predicators of the spatial field: the concepts ‘being left’ and ‘being near’ may be applicable to the same particular case, but in the description of such a case on the spatial field, we clearly can distinguish the propositions ($\{X, Y\} \varepsilon \textit{Being Left}$) and ($\{X, Y\} \varepsilon \textit{Being Near}$) within finite time.

Last but not least, we call these pictorial contexts on the geometrical level *mental* because we use them in explanations of conscious behavior. This conception of the mental is not private, as was the understanding in the framework of the Philosophy of Enlightenment, but (like the whole explanation) interpersonally accessible. ‘Mental image’ in this understanding is not defined in opposition to ‘proposition’, since a mental image is conceived as a set of propositions of a certain field of concepts with a characteristic attribute – syntactical density.⁹

9 The Connection between Visual and Verbal Space

Let us finally consider the interaction between the horizontal and vertical dimensions of explanation of the understanding of an utterance in its context: if we integrate a reference to the schema of the spatial field to the horizontal (‘propositional’) explanation given above, i.e., if we change from analytic to synthetic judgments, the revision of the context by means of the predicator’s concept is now explained in three steps (cf. Fig. 3):

- First, the proposition of the utterance is transformed into a corresponding structure of sets of propositions on the lower fields following the schema of the spatial field (1b). According to our definition above, in the geometrical field, the context of the utterance corresponds to a mental image (1a): we here meet again the perceptual aspects of Kant’s conception of intuitions and Johnson-Laird’s understanding of (visual) mental models which both represent the context in their explanative systems.
- Second, the revision of the context by means of the spatial concept communicated by the predicator of the utterance takes place on the lower fields (2a & b): the corresponding projections of the context are revised by those concepts of the lower fields implementing the spatial concept thereby taking into account the interactions between the implementing fields given by the schema. The horizontal dimension of explanation is shifted down one level, so to speak.
- Third, the resulting mental models on the lower fields – especially the mental image of the geometrical level – are synthesized back to the spatial field where they form the wanted

⁸In any particular context of the geometrical level of Vieu, only a finite number of points is realized – the construction of points in Clarke’s calculus of individuals results only in the *relevant* points for that context (cf. [Vie91, p. 130f.]); but infinitely many contexts can be constructed which are similar to any degree to that context; cf. also [McG84, p. 213ff.: (‘Gesamtheit und System’)];

⁹Note that this criterion clearly excludes approaches to mental images based on discrete cell matrices;

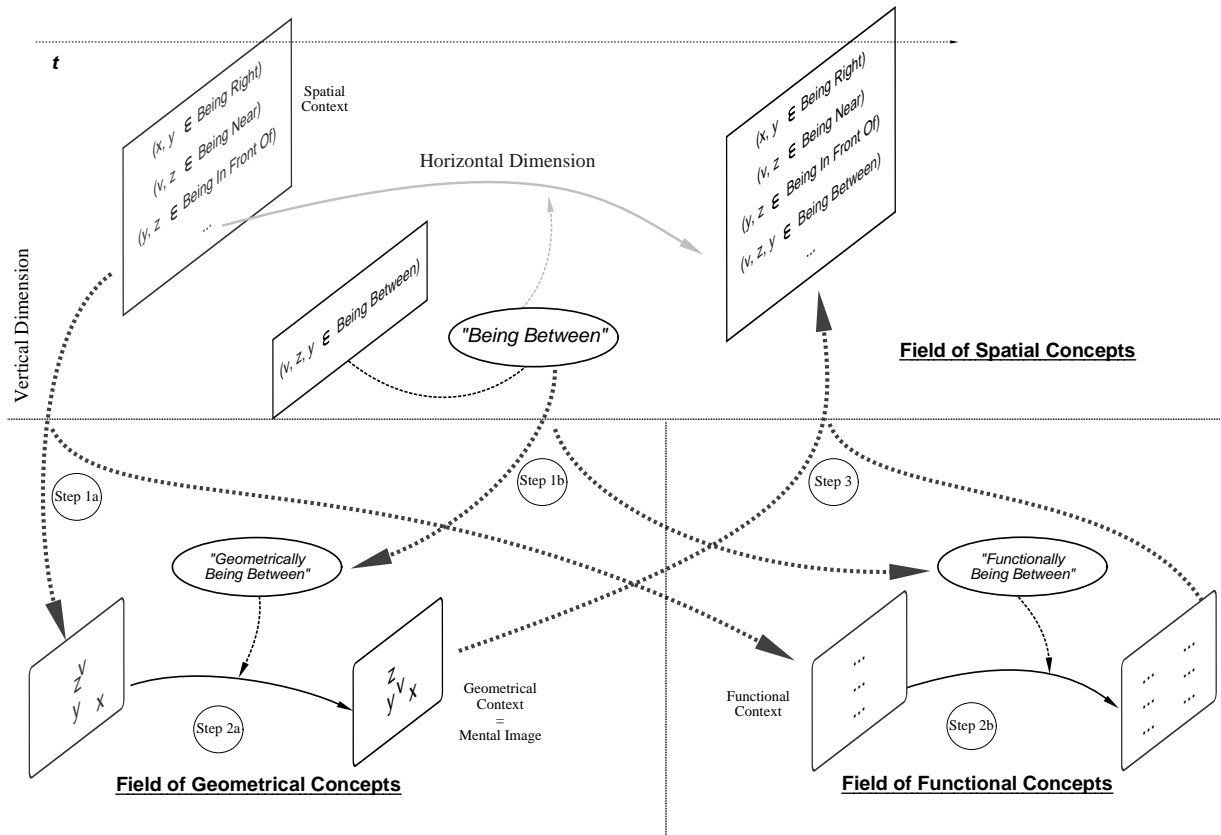


Figure 3: Interaction of Vertical and Horizontal Explanations

mental model used as the context for the subsequent utterance (3): this corresponds to the already mentioned application of the spatial concepts to make explicit the implicatures in the revised model.

The analytic explanation of the transformation of a context by a new predication is explained by a set of associated transformations in the implementing fields of concepts and the corresponding shifts between the implemented and the implementing fields. Thus, the concept ‘mental image’ as we understand it here comes into the explanative game if the field of concepts we are investigating is viewed as synthesized using a syntactically dense field of concepts – like the geometrical level of View.

A corresponding integration of the vertical and horizontal dimensions of explaining spatial cognition is exemplified by the system SOCCER of the project VITRA: in this case, the exemplary radio sports reporter from the beginning is considered (cf. Fig. 1). The explanation of the visual perception, which is part of the foundation by reference semantics of the utterances of the radio reporter, starts from the field of spatial concepts and views it as implemented on a cascade of lower fields of concepts down to a special field conceived as the common basis to describe the visual field of the reporter in VITRA: the field of the ‘digitized image sequences’ describing the output of a video camera. This rather poor field contains the concept of a very special and unusual kind of object: an instantaneous dot with a location (relative to the camera-system)

and an intensity (grey-value) – the only elementary attributes of the objects on this level. The field of spatial concepts in SOCCER (cf. [Sch93], [Her93]) is based on elementary concepts like distance and direction which allow us to define simplified versions of the well-known static spatial relations, like ‘being in’, ‘- at’, ‘- near’, ‘- to the left’, etc. These concepts rely on a reduced version of sortal objects: rigid objects persistent over time, with form, location, and velocity. The concept ‘distance’ is also used here in the temporal sense – as duration of the phases of events: in combination with the spatial relations, it is involved in defining the concepts of spatial events, like ‘doing a double pass with’. These relations are internal to the spatial field and form the logical rules of that field in SOCCER. They are synthetically based on the syntactically dense field of concepts of the digitized image sequences. The system XTRACK (cf. [Kol92]) describes the mapping from the simplified version of sortal concepts used on the spatial field in VITRA to the field of ‘object candidates’ – corresponding approximately to Vieu’s ‘histories’ – and a field of simple part-of relations.¹⁰ These object candidates are projected down to ‘digitized image sequences’ by the system ACTIONS (cf. [Sun88], [SBSZ87], and [HSE⁺89]) which incorporates as schema Gestalt principles, like grouping of similar elements and good continuation. With these two schemata, the concept of ‘perception’ in this explanatory framework finds the reporters spatial utterance in the description of his situational context on the level of the ‘digitized image sequence’: starting from the digitized image sequence as the primary description of the visual field commonly ascribed to the radio reporter, maximally coherent interpretations are constructed successively on the higher fields. Finally, some propositions of the spatial field are chosen to be communicated and transformed into a corresponding verbal manifestation – following the concept of ‘reference semantics’.¹¹

As was mentioned above, Mead expects that any adequate theory of communication explaining the behavior of a speaker also has to consider the audience in a particular way: the speaker has to be conceived as somebody who sets himself in the position of his audience – playing anticipatorily its role in the language game – in order to consciously communicate. In VITRA, this demand is redeemed by means of the listener model ANTLIMA (cf. [Sch93]): we focus here only on the static spatial relations, as in sentence (S 1), although spatial events as in (S 2) are dealt with, as well. The understanding of the audience is modeled with the three steps described above: first, the proposition of the (planned) utterance is projected to the lower levels synthesizing the spatial field: i.e., from restrictions of the spatial interaction with other objects to restrictions of the locations of the objects (corresponding to the result of XTRACK); this transformation – the schema of the corresponding spatial concept – is encoded in ANTLIMA by means of functions called ‘TyPoFs’ (cf. below and Fig. 4); second, the context of the planned utterance is revised on the lower level, i.e., as mental image, following the schema: the locations of the objects are chosen by means of a hill climbing algorithm ruled by the TyPoFs and depending on the contextual positions (cf. Fig. 4 illustrating the influence of three different geometrical contexts (starting positions) on the selected location); third, the schemata of the spatial concepts are applied to construct the context on the spatial field with the explicit implicatures. Finally in the listener model, the resulting context modeling the anticipated understanding of the audience has to be compared with the intended understanding, i.e., what has been perceived: the differing propositions are used in an *anticipation feedback*

¹⁰At present, XTRACK is not yet integrated in SOCCER;

¹¹The field of spatial concepts is used also as a basis to implement even higher fields of concepts, especially those around the concept of an object involved in intentional activities and autonomous self-control – like the players of a soccer game – and teams of such objects (cf. [RS91]);

loop to find the selections of propositions and their verbalizations uttered.

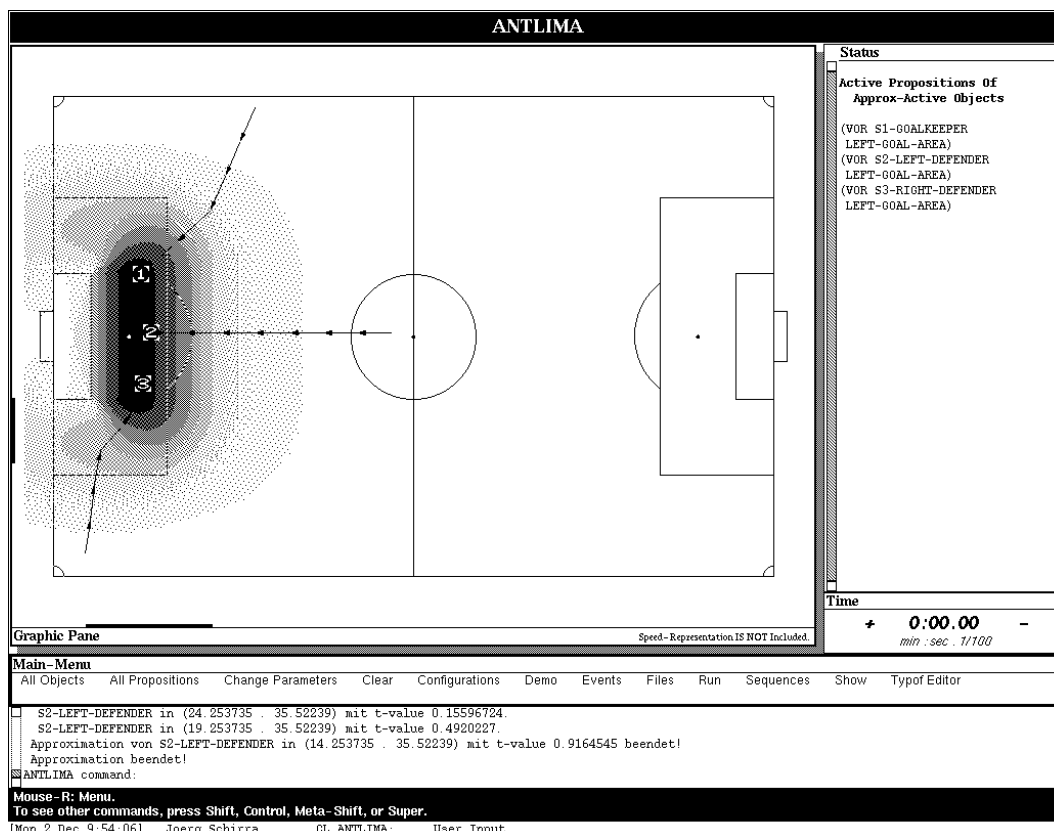


Figure 4: Typicality distribution of the spatial relation ‘player in front of the goal area’

In this case, using a syntactically dense field of concepts to explain a syntactically sparse field results in two interdependent characteristics. First, the concepts of spatial relations are *graded*, i.e., although an unbounded number of locational propositions is associated with one spatial proposition, they are not all associated in the same strength: some locations are more typical with respect to a spatial relation than others, or viewed from the other direction: the spatial description is better applicable to some locational descriptions than to others. The functions which encode in ANTLIMA this mapping are therefore called *Typicality Potential Fields*: for the concerned spatial proposition, they associate to every corresponding proposition on the geometrical field a typicality value – a real number $\in [0..1]$ graphically represented in Fig. 4 by means of grey values. The gradation allows us to formulate an important selection criterion which directs the communication between the speaker and his audience: both expect that the *most typical interpretation* of an utterance is intended, and that discrepancies must be mentioned explicitly. Second, the syntactical density of the geometrical field restricts the possibilities to find the resulting contexts in a characteristic manner: as was said before, the predicators in syntactically dense fields cannot be distinguished by means of a decidable algorithm. Correspondingly in ANTLIMA, we are not able to select the undoubtedly optimal positions. But a procedure of stepwise approximation on the basis of the gradedness of the spatial concepts – hill climbing – enables us to find at least an ‘almost best’ location.

10 Conclusion

On the preceding pages, we presented in a preliminary manner how the concept of ‘mental image’ can be introduced and applied as a meta-explanative tool in reference semantics: on the basis of Kant’s distinction between analytic and synthetic methodological propositions and Goodman’s concept of ‘syntactical density’, we determined the concept of ‘mental image’ to distinguish the contexts of a syntactically dense field of concepts used to synthetically found the field of spatial concepts. The validity of this conception was motivated by applying it successfully to two – on first view rather different – approaches in AI: Vieu’s logical analysis of the predicator ‘being in’, and the project VITRA concerned with the connection between vision and language systems. Mental images play a crucial role if we integrate the concept of ‘reference semantics’ anchoring utterances in the situational (here visual) context with Mead’s conception of conscious communication: if communication depends on the speaker’s anticipation of the audience’s comprehension, the reference semantical connection between visual and verbal space demands for the reconstruction of perceptually absent referents in the form of mental images: ANTLIMA, the listener model in VITRA, exemplifies this idea.

M. Beckenkamp, K. Buchholz, E. Hays, W. Maaß, P. Olivier, S. Pribbenow, G. Retz-Schmidt, A. Ros, P. Sablayrolles, D. Schmauks, and L. Vieu have to be heartily thanked for their comments and discussions on the theme of this paper.

References

- [Car47] Rudolf Carnap. *Meaning and Necessity*. Chicago, 1947.
- [Dan69] H. Dankert. *Sportsprache und Kommunikation – Untersuchungen zur Struktur der Fußballsprache und zum Stil der Sportberichterstattung*. Tübinger Vereinigung für Volkswissenschaften e.V., Tübingen, 1969.
- [EM85] Hartmut Ehrig and B. Mahr. *Fundamentals of Algebraic Specification*. Springer, Berlin, 1985.
- [Goo68] Nelson Goodman. *Languages of Art – An Approach to a Theory of Symbols*. Indianapolis, 1968.
- [Her93] Gerd Herzog. Utilizing Interval-Based Event Representations for Incremental High-Level Scene Analysis. In M. Aurnague, A. Borillo, M. Borillo, and M. Bras, editors, *Semantics of Time, Space, and Movement*, pages 453–463, Toulouse, 1993. Groupe “Langue, Raisonnement, Calcul”, CNRS und die Universitäten Paul Sabatier und Le Merail. Working Papers of the 4th International Workshop, Château de Bonas, 1992.
- [HSE⁺89] Gerd Herzog, C.-K. Sung, E. André, W. Enkelmann, H.-H. Nagel, T. Rist, W. Wahlster, and G. Zimmermann. Incremental Natural Language Description of Dynamic Imagery. In W. Brauer and C. Freksa, editors, *Wissensbasierte Systeme*, pages 153–162, Berlin, 1989. Springer.
- [JL83] Philip N. Johnson-Laird. *Mental Models*. Cambridge University Press, Cambridge, UK, 1983.
- [Joh76] F. Grant Johnson. *Referenz und Intersubjektivität – Beitrag zur philosophischen Sprachpragmatik*. Suhrkamp, Frankfurt/M., 1976.
- [Kam86] Hans Kamp. Belief Attribution and Context: Comments on Robert Stalnaker. Technical Report FNS-Bericht-86-13, Universität Tübingen, FNS, Tübingen, 1986.
- [Kam90] Hans Kamp. On the Representation and Transmission of Information: Sketch of a Theory of Verbal Communication Based On Discourse Representation Theory. In E. Klein and F. Veltman, editors, *Natural Language and Speech, Symposium Proc.*, pages 135–158. Springer, Berlin, 1990.
- [Kan60] Immanuel Kant. Logik. In W. Weischedel, editor, *Werke in 12 Bänden*, volume VI. Wiesbaden, 1960.
- [Kan65] Immanuel Kant. *Critique of Pure Reason*. St Martin’s Press, New York, 1965. (orig. by Macmillan & Co, Ltd., 1929), Transl.: Norman Kemp Smith.
- [KL73] Wilhelm Kamlah and Paul Lorenzen. *Logische Propädeutik – Vorschule des vernünftigen Redens*. BI Wissenschaftsverlag, München, Wien, Zürich, 1973.
- [Kol92] Dieter Koller. *Detektion, Verfolgung und Klassifikation bewegter Objekte in monokularen Bildfolgen am Beispiel von Straßenverkehrsszenen*. Infix, St. Augustin, 1992. (Doktorarbeit).
- [Lor70] Kuno Lorenz. *Elemente der Sprachkritik – Eine Alternative zum Dogmatismus und Skeptizismus in der Analytischen Philosophie*. Suhrkamp, Frankfurt/M., 1970.

- [McG84] B.F. McGuinness, editor. *Ludwig Wittgenstein und der Wiener Kreis – Gespräche, aufgezeichnet von Friedrich Waismann*. Suhrkamp, Frankfurt/M., 1984. Wittgenstein Werkausgabe Band 3 (stw 503).
- [Mea62] George Herbert Mead. *Mind, Self, and Society. From the Standpoint of a Social Behaviorist*. Chicago University Press, Chicago, IL, 1962. (ed. by Charles W. Morris, orig. publ. in 1934).
- [PI66] Jean Piaget and B. Inhelder. *L’image mentale chez l’enfant*. Presses Universitaires de France, Paris, 1966.
- [Pri91] Simone Pribbenow. *Zur Verarbeitung von Lokalisierungsausdrücken in einem hybriden System*. PhD thesis, Univ. Hamburg, Hamburg, 1991.
- [Ros90] Arno Ros. *Begründung und Begriff – Wandlungen des Verständnisses begrifflicher Argumentationen*. Felix Meiner Verlag, Hamburg, 1990. (3 volumes).
- [Ros91] Arno Ros. Kants Begriff der synthetischen Urteile a priori. *Kant-Studien*, 82(2):146–172, 1991.
- [RS91] Gudula Retz-Schmidt. Recognizing Intentions, Interactions, and Causes of Plan Failures. *User Modeling and User-Adapted Interaction*, 1(1):173–202, 1991.
- [SBSZ87] Jörg R.J. Schirra, G. Bosch, C.K. Sung, and G. Zimmermann. From Image Sequences to Natural Language: A First Step towards Automatic Perception and Description of Motions. *Applied AI*, 1(3):287–305, 1987.
- [Sch91] Oliver R. Scholz. *Bild, Darstellung, Zeichen – Philosophische Theorien bildhafter Darstellung*. Kolleg Philosophie. Verlag Karl Alber, Freiburg/München, 1991.
- [Sch93] Jörg R.J. Schirra. A Contribution to Reference Semantics of Spatial Prepositions: The Visualization Problem and its Solution in VITRA. In C. Zelinsky-Wibbelt, editor, *The Semantics of Prepositions – From Mental Processing to Natural Language Processing*, pages 471–515. Mouton de Gruyter, Berlin, 1993. (also as VITRA Report No. 75, 1990).
- [Sch94] Jörg R.J. Schirra. *Sprachliche Bildbeschreibung als Verbindung von visuellem und sprachlichem Raum – Eine interdisziplinäre Untersuchung von Bildvorstellungen in einem Hörermodell*. PhD thesis, Fachbereich Informatik, Universität des Saarlandes, Saarbrücken, 1994. forthcoming.
- [Str71] Peter F. Strawson. *Logico-Linguistic Papers*. London, 1971.
- [Sun88] C.-K. Sung. Extraktion von typischen und komplexen Vorgängen aus einer langen Bildfolge einer Verkehrsszene. In H. Bunke, O. Kübler, and P. Stucki, editors, *Mustererkennung 88*, Berlin, 1988. Springer.
- [Tug76] Ernst Tugendhat. *Vorlesungen zur Einführung in die sprachanalytische Philosophie*. Suhrkamp, Frankfurt/M., 1976.
- [Vie91] Laure Vieu. *Sémantique des relations spatiales et inférences spatio-temporelles: Une contribution à l’étude des structures formelles de l’espace en Langage Naturel*. PhD thesis, Institut de Recherches en Informatique de Toulouse, Univ. Paul Sabatier, Toulouse, 1991.
- [Wit53] Ludwig Wittgenstein. *Philosophical Investigations*. Macmillan, New York, 1953.

[Wit89] Ludwig Wittgenstein. Ursache und Wirkung. Intuitives Erfassen. In Joachim Schulte, editor, *Wittgenstein: Vortrag über Ethik und andere kleine Schriften*, pages 101–139. Suhrkamp, Frankfurt/M., 1989.

Parts of two drawings from *Loriot's Heile Welt* (Diogenes, Zürich, 1973) were used for Fig. 1;

Paper appeared in M. Aurnague, A. Borillo, M. Borillo, and M. Bras (eds.): *Semantics of Time, Space, and Movement: Working Papers of the 4th International Workshop, Château de Bonas*, Toulouse: Group “Langue, Raisonnement, Calcul”, CNRS and the universities Paul Sabatier and Le Merail, 1993, p. 105–121.