Conceptual and Semantic Knowledge in Language Production

Proceedings of a Workshop of the Special Collaborative Research Program 245
Language and Situation

Heidelberg, November 15th -- November 17th, 1995

Ralf Meyer-Klabunde & Christiane von Stutterheim (eds.)

Bericht Nr. 92

Band I

Arbeiten aus dem Sonderforschungsbereich 245 "Sprache und Situation",
Heidelberg/Mannheim

Wir danken der Deutschen Forschungsgemeinschaft für die Förderung unserer Arbeiten.

Information zum SFB 245 im WWW:
http:\\www.sfb245.uni-heidelberg.de

ISSN 0941/990X

Contents

Volume I

Preface

The volume comprises a collection of papers presented at the symposium *Konzeptuelles und semantisches Wissen in der Sprachproduktion* (Conceptual and Semantic Knowledge in Language Production). The symposium, organized by members of the Sonderforschungsbereich 245 *Sprache und Situation* (Language and Situation), was held on November 15th to 17th at the University of Heidelberg. Most of the papers are revised versions of the talks.

The goal of the conference was to bring together researchers from different disciplines to discuss questions of language production. The current state of the art is characterised by a number of very general models of speech production, while empirical research leading to a specific evaluation of these models has been relatively limited.

In particular, questions concerning the relation between conceptualisation and (language specific) semantics which formed the focus of the symposium have not yet found convincing solutions. The symposium thus provided a forum to bring together results from empirical studies in various domains and to discuss different theoretical models in the light of these findings. The following questions formed the guidelines for the presentations and the plenary discussions:

- Which levels and possible interfaces have to be distinguished on the way from conceptualisation to formulation?
- What form do processing units take at different levels of the system?
- In what way do language specific semantic structures affect conceptualisation processes, i.e. processes of information selection and linearisation?
- At what point(s) in the planning process and to what extent are assumptions about the listener integrated in the planning process?

These questions are dealt with in areas as different as neurophysiological research, artificial intelligence, psychology, and linguistics. They converge, however, in forming a very clear picture: We have only taken initial steps in the explanation of production processes; the papers present some of these steps and show some ways forward in future research.

Ralf Meyer-Klabunde Christiane von Stutterheim

We would like to thank the German Science Foundation (DFG) for supporting the symposium.

Conceptualisation in language production: How conceptualisation processes which are specific to language determine lexical access

Mary Carroll[1]
University of Heidelberg, Plöck 55, D-69117 Heidelberg

Abstract

The study reported on here investigates the relationship between different knowledge sources in language production and how they interact in determining the choice of a lexical item when naming and identifying objects in a communicative task (instructions). In language production, information relating to a specific state of affairs has to be mapped into a network of conceptual relations (spatial, temporal, modal, etc.). It is assumed that information organisation of this kind requires the generation of a conceptual structure or mental model which allows speakers to deal with knowledge from different sources in coherent terms. Although overall principles of organisation may be similar across languages, this does not apply to the specific concepts which languages provide in encoding such relations. The study looks at the way speakers of English and German organise conceptual material at this level for a specific task, and how the options which each language provides affect lexical access. The results show that the lexical items accessed in identifying the same set of objects differ, depending on the conceptual structure used to organise information. Speakers of English and German show language-specific differences in both the concepts used in organising information and the lexical items accessed, providing evidence of a language-specific level of conceptualisation or 'thinking for speaking' in language production.

1. Introduction

The decision to refer to an entity in a specific way is the product of a series of processes which draw on different knowledge sources. Extensive studies have been carried out by both psychologists and linguists on the factors determining lexical selection in different contexts. A number of studies deal with the way participants collaborate in making object references comprehensible. Results show how speakers tailor their contributions in accordance with listener-related factors such as shared knowledge and overall contextual relevance. Object features that are discriminatory in the context of use are of crucial importance in determining the lexical item selected (Olson 1970; Herrmann & Deutsch 1976; Clark & Wilkes-Gibbs 1986; Mangold & Pobel 1988). The role of linguistic factors such as

[1] This study was supported by the German Research Society, SFB 245.

given versus new information in texts has also been the subject of many investigations (Hawkins 1978; Clark & Marshall 1981; Givón 1983; Ariel 1990).

However, object reference has generally been studied in situations in which object discrimination formed a central communicative goal. Factors which lead to the choice of a lexical item in language production when object reference forms but one aspect of a communicative task have not been studied on a similar scale (cf. v. Stutterheim et al 1993; Kohlmann 1996). In complex communicative tasks, such as giving sets of instructions on how to assemble a model vehicle, object identification is but one aspect of the task with which the speaker has to contend. When instructing an addressee to carry out a series of actions in an assembly process, speakers have to organise different kinds of conceptual material to form a coherent whole (spatial information on where the pieces fit, function of each piece, types of action required). Information organisation of this kind has to be described in terms of the conceptual structures or mental models required to deal with it.

Mental models and their role in language have been developed mainly from the perspective of language comprehension (Johnson-Laird 1983; Greeno 1983; Morrow, Greenspan & Bower 1987). Such models are seen to represent the states of affairs portrayed in discourse. Models of information organisation in language production are few by comparison (Kamp 1981; Garrod & Sanford 1988; v.Stutterheim & Klein 1989; Slobin 1991). Garrod and Sanford argue that "language only relates to the world in a principled way through the mediation of mental models of that world. Hence utterances may only be given precise meanings in relation to models which can then be variously mapped onto the world to yield a variety of distinct interpretations." (ibid:147).

Models of the language production process as such all show a broad consensus on the main levels or stages of information processing (conceptualisation, formulation, articulation). They share the view that the processes which lead to the generation of a message draw openly on a broadly defined conceptual base (Levelt 1989; Herrmann & Grabowski 1994). In the model presented in Levelt (1989), for example, the encoding process starts with a conceptualiser which is described as "a highly open-ended system" involving very different aspects of the speaker as an acting person. The conceptualiser generates pre-verbal messages which are adapted to the encoding process. Messages differ from general cognitions in the sense that they are propositional, have thematic structure, and are structured so as to allow the expression of basic kinds of relations (temporal, modal, etc.). So the conceptualiser which is activated in generating a message encompasses knowledge which is very general in nature, as well as linguistic knowledge of a certain kind. Little is known, however, as to "how lemmas are activated by fragments of the message." (Levelt 1989:198).

In the language production model of Herrmann & Grabowski (1994) speakers first engage in a process whereby information in memory which is relevant to the task is focussed (focus memory). This is carried out by a central executive which includes both declarative and procedural components. Information in focus memory is present in the form of concept structures (ibid:300). Part of this information is selected and adapted for expression. Adaptation means that the selected information is checked for task-related criteria such as adequacy, informativeness, and replaced by alternative options of a similar category when necessary; adapted information is then linearised for expression. Adaptation plays a central role in this model of language production with provisions for feedback and regulatory processes between the different levels. Specific steps in the production process such as the choice of a lexical item are described in broad terms by the process of selection and adaptation.

The relationship between conceptual and linguistic structure, in as far as it relates to lexical choice, is stated as follows in Bierwisch and Schreuder (1992), drawing on Levelt's model. The choice of a lexical item is determined by

> "the speaker's mental model of his or her external environment, which incorporates all sorts of information - visual, auditory, motoric, inferential, background knowledge etc." (Bierwisch & Schreuder 1992:24)

In other words, the structures on which messages are based, and with this the processes governing the choice of lexical items, allow the speaker to access a wide range of concepts.

The study presented here sets out to investigate this question, looking at the types of concepts which speakers access in object reference when these are embedded in a complex communicative task. The question is do speakers have access to the full range of concepts available in memory, and their counterparts in the lexicon, bringing them into focus as required, or do the planning processes required in language production narrow the range in a systematic way. The study thus looks at how speakers of English and German organise conceptual material for a specific communicative task and how they proceed in dealing with sub-tasks (e.g. object reference) within this context.

In its broadest sense, the grammar of a language determines the types of relations which can be established in specific conceptual domains. Studies of the spatial expressions provided in languages and their grammatical form show significant constraints on the kinds of spatial relations that are represented in languages such as English and German (Talmy 1983; Jackendoff 1983; Habel, Herweg & Rehkämper 1989; Rauh (1991); Lang, Carstensen & Simmons 1991; Jackendoff & Landau 1991; Klein 1991; Becker 1994). The question is what factors determine the selections made in specific contexts in language production? Klein & v. Stutterheim (1987) have proposed a model of information organisation where one of the central factors which determine how information is selected for expression depends on how the speaker defines the communicative task. The task may be treated alternatively as requiring a narrative presentation of the events at issue, or it may be seen as better resolved by providing a description of the state of affairs, or by giving a set of instructions on how to change them. The information selected is that which answers a general question or quaestio. When a communicative task is seen as best resolved by providing a description of things, the information verbalised is selected and structured as an answer to the quaestio "what is x like?". This is a general question which holds for all descriptions. In giving a description, relevant information must thus be organised so as to allow the ascription of features to the entity specified. The quaestio thus isolates a general criterion which applies for all descriptive tasks. It provides guidelines on the type of conceptual material which has to find expression if a communicative task is to hold as a description. The guidelines are formulated for conceptual domains (spatial, temporal, object domain). It determines on a systematic basis what categories of information form the main structure (foreground) of a text (v.Stutterheim & Klein 1989), and imposes constraints on the types of abstract relations which can be established between entities and events which hold for the task as a whole.

If the communicative task is treated as requiring a set of instructions, the information which must be selected if the task is to hold as an instruction is determined by the conceptual material required in providing an answer to the question "how does one do x?" Other categories of information expressed, though possibly indispensable in making main structure information comprehensible, belong to what is termed side structure.

The distinction between "what must be said" and "how to say it" is common to many models of language production and language generation (McKeown 1985; Appelt 1985;

Herrmann 1986; Danlos 1987). The quaestio model provides criteria which determine what categories of information and types of relations should be expressed in different conceptual domains, given a specific communicative task. This model does not determine how the speaker should proceed in implementing conceptual material to solve problems such the linearisation of information (sets of propositions) in each specific description, instruction or narrative task. The different means which speakers typically implement or devise is an empirical question. Studies of the options chosen in giving instructions shows that although the temporal domain plays a fundamental role when linearising information, for example, (what to do first, next, etc.) speakers can also choose to establish additional levels of information organisation, linking required states of affairs from a predominantly spatial or functional perspective. Once a procedure is selected, however, this establishes constraints which are binding for the task (v. Stutterheim 1995).

Different forms of conceptual organisation which support coherence in a linear medium are indispensable to the process of information organisation in language production. They need not figure in mental representations which are of a more general nature. The present study thus relates to the conceptual structures which speakers generate when organising information for expression.

2. The Data

The task of identifying a series of objects was embedded in a set of instructions which involved the assembly of parts of the base of a model vehicle (see below). Speakers of English and German were asked to instruct another person (research assistant) in putting the parts together. Eight parts had to be named and identified, involving 8 separate assembly steps. The study was carried out in two stages with different groups of speakers. The participants were college students from various faculties.

Setting (a): The base of the unit was designed so it did not resemble any particular entity, nor did it have clearly defined features such as a front and back section. Participants were told, however, that it was part of a fork-lift truck. With the exception of a gear wheel, the parts under assembly were also non-descript, so some ingenuity was required in finding terms to describe them.

Since the same data were also used in a separate investigation to study the way speakers tailor their speech to the requirements of the addressee, the data in this setting (a) was collected with and without an addressee.

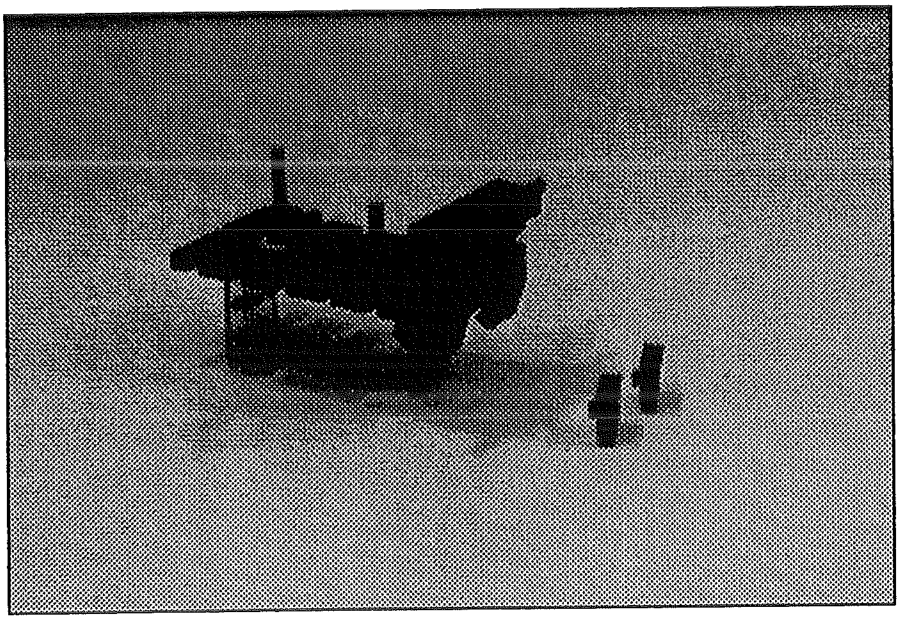
(a/c) Addressee present: The instructions were given to a research assistant who assembled the parts on-line as instructed (10 speakers of English; 10 speakers of German).

(a/d) Addressee absent: The participants were told that the assembly would be performed at a later date, using the instructions recorded on tape. (15 speakers of German).

Setting (b): The task was repeated with a remodelled base unit in which elements were added to give a salient front and back section. Some of the parts were similar to those used in the first setting, but new ones were also required to fit the added elements.
Participants: 10 speakers of English; 10 speakers of German. This setting was carried out without an addressee with both groups (10 speakers of English; 10 speakers of German)

2.1 Procedure

Demonstrations were given of the assembly, with no verbal interaction on the part of the researcher. The participant was then asked to practise the assembly before giving the instructions. A set of photos was also provided showing each part and its required position on the base of the unit. This eliminated the need to provide verbal assistance in the practice session.



a/c) Addressee present: During the instructions the speaker and addressee (research assistant) were seated opposite one another at the same table. The parts were hidden behind a small screen so the instructor could not simply point but had to give a full description when identifying the parts. Otherwise the instructors could see what the addressee was doing and intervene (verbally) where necessary.
a/d) and b) Addressee absent: The participants were told that the assembly would be performed later on the basis of the recorded instructions.

3. Organising information in the spatial domain and object domain

One of the central tasks for the instructor, in addition to identifying each part, is to specify the place on the unit at which each part should be placed. The frame of reference chosen to describe the different locations is not uniform. Roughly speaking, two types of frame were implemented.

Some speakers adopt a holistic or global perspective and view the entity as having front, middle and back sections (see tables below). The individual parts are located on the base of the unit in terms of these sub-spaces.

In the second case the frame of reference was structured on the basis of individual parts, starting with the parts already on the base, and the individual features provided by these and and each successive part placed on the unit. The locations which can be distinguished in this frame relate to individual places and their possible sub-spaces (inner space, surface space, adjacent space). They emerge on a point by point basis with each part mounted during the assembly.

3.1 Holistic frame

Ex. 1 *What you first want to get is the main body*
it's a red platform
with a few pieces stuck onto it
***the top** is where the two white circles are*

Ex. 2 *lets start with **the top part***
with the red that's on the top

Ex. 3 *also Du hast jetzt das Grundteil/ dieses grosse Teil*
okay you have now this base-piece/ this big piece
mit der schwarzen Kurbel dran
with the black crank there-on
*(..) also **vorne** ist immer da*
*okay **in front** ist always there*
wo die Kurbel ist
where the crank is
*und **hinten** ist*
*and **in back** is*
wo auf dieser Grundplatte noch nichts draufmontiert ist
where on this base-piece nothing is yet placed

Ex. 4 *also das Grundelement*
alright the basic element
das ist eine Trägerplatte mit zwei grauen Ecksteine
that is a platform with two grey corner-squares
*dann diese Trägerplatte hat **in der Mitte** Löcher*
*then this platform has **in the middle** holes*
*da **an dem rechten Ende** ist eine Kürbel mit einem Zahnrad*
*there **on the right end** is a crank with a gear wheel*

The spatial concept selected in the frame of reference for the spatial domain is based on the set of coordinate axes (vertical, sagittal, and lateral axes) which give top, bottom, front, back, left and right sections of the unit as a whole. The examples also show how speakers relate explicitly to features of the unit as an entity (platform with pieces on it). In the quaestio framework of analysis this is treated as information which relates to the object domain.

3.2 Point by point frame

Ex. 5 *we're going to start with the longest piece*
it's the biggest red piece

*and it slides **onto the two black parts***

Ex. 6 *I want you to take a red piece*
it's a long red piece with two poles
*and then put it **onto the two grey pieces***

Ex. 7 *das längliche rote Teil steckst Du auf die zwei grauen Blöcke*
*okay the long red piece place you **on the two grey blocks***

Attention is directed in this frame of reference to the individual pieces. Locations on the base are defined in terms of the places they delimit (the two grey pieces). Reference points such as the top (of the unit) or the front (of the unit) do not occur in this frame.

These two frames of reference represent different ways of conceptualising the spatial structure of the unit under assembly. In this sense they are different mental models of the state of affairs in question.

4. Lexical Access in Text Production

As mentioned above, the individual parts which have to be identified are for the most part unfamiliar and speakers have to find ways of identifying them. The first step in the analysis was to find out if the conceptual means used in the spatial domain correlate with the lexical means selected when identifying the individual parts. The means adopted (holistic versus point by point frame) thus function as an independent variable in the investigation of factors leading to the choice of lexical form.

The results show that the lexical forms accessed correlate with the way speakers structure the spatial domain.

Terms accessed with holistic organisation in spatial domain

English: *rectangle, square, block, gear, washer, wedge, circle, wheel, crank device*
German: *Quader* (block), *Block* (block), *Stein* (stone), *Schiene* (rail), *Achse* (axis), *Stift* (pin), *Ring* (ring)

If a holistic perspective is adopted, the terms accessed for each part also reflect a holistic perspective, relating to shape and function.

Terms accessed with point by point frame in the spatial domain

English: *piece, part*
German: *Teil* (part), *Stück* (piece)

In this case individual parts are described by the same general term (*piece* or *part*).
The parts are identified via additional attributes which are encoded as adjectives or in the form of a separate statement.

Ex. 8 first take the long rectangular red piece

Ex. 9 the first piece is the longest red piece
it has two eh feet on the bottom

These results apply to speakers of both languages. Descriptions which relate to form or function in a point by point frame are confined to objects with a familiar form (ring, gear). Before presenting these results in detail, however, there is a another factor which determines the extent to which terms relating to shape and form are accessed, given a holistic frame of reference.

4.1 Language specific differences in the concepts used to structure space holistically

Speakers who adopt a subjective or deictic viewpoint in determining what holds as the front, back, left and right side of the unit access more lexical terms which relate to shape or form when identifying the individual parts, compared to speakers who adopt an intrinsic perspective on what holds as front or back. As mentioned above, the unit under assembly did not have inherent features which would easily lead to the construal of a holistic frame of reference (setting a). Some speakers of German establish a front, back, left and right from an external viewpoint, specifying which side of the base holds as the front etc. A deictic front is typically the side closest to the front side of the speaker or other participant. This viewpoint is termed **deictic** and contrasts with an **intrinsic** viewpoint in which front and back sections are defined by inherent parts of an object and their intrinsic features (Bühler 1934; Fillmore 1982; Levelt 1984; Herskovits 1986; Graumann 1989; Garnham 1989; Klein 1991).

English and German speakers differ in the type of concept used to structure space, given a holistic frame. Deictic and intrinsic viewpoints are established by speakers of German, while speakers of English adopt intrinsic viewpoints only. Since speakers of English do not establish a deictic viewpoint with a holistic frame, they do not access novel, shape-related terms to the same extent as the German group.

The use of a deictic perspective is formally restricted in English via the morpho-syntactic structure of locative expressions (Carroll 1993; Carroll and v. Stutterheim 1993). Spatial expressions in English are closely linked to abstract inherent features which objects provide, features such as an object's boundaries, or its intrinsic orientation. Although deictic expressions are provided in the system of spatial expressions in English, their use is constrained by their structure. In morpho-syntactic terms spatial expressions occur as prepositions and adverbs. Prepositions form parts of a noun phrase and are thus grammatically bound to the nouns and objects with which they co-occur. Grammatically they form the head of the phrase. Significantly, prepositions cannot occur without explicit reference to the object to which they relate. In contrast, adverbs are not bound to nouns and objects in the same way. They can occur on their own without any reference to the underlying object or reference point. They can thus serve to convey spatial structures which are not closely determined by objects and their features. Spatial structures encoded via adverbs can overrule object features when projected onto reference objects. Adverbs are thus suitable candidates to encode spatial relations which are not object-centred and are used in German and English for deictic, subject-centred spatial structures. However, English has no spatial expressions which are genuine adverbs; spatial expressions consist of prepositions (on, in, above, below, for example), or prepositional phrases (*on top of, in front of, at the top of*). Deictic spatial expressions in English differ from their intrinsic counterparts in that they exhibit an adverbial-like structure. They are not genuine adverbs, however, but covert prepositional phrases. Structural means are thus in place in the system to limit the contexts in which deictic viewpoints can be implemented. The adverbial structure allows the application of deictic viewpoints in a restricted set of contexts only. (A detailed description of the grammatical structure of these spatial expressions is given in Carroll 1993 or Carroll & v. Stutterheim 1993 and cannot be repeated here.) Problems which arise with the adverbial character of deictic expressions in English can be briefly demonstrated on the basis of the following object description.

This is an L-shaped figure with a red square on top and a blue one underneath.

The two expressions *on top* and *underneath* behave like adverbs in this context (they do not require mention of the reference object *the figure*). But their prepositional character becomes evident when the reference object to which they apply is mentioned. The places delimited by the expressions change:

This is an L-shaped object with a red square on top ***of the figure*** *and a blue one underneath* ***the figure***.

The places delimited no longer form part of the figure as such. Object boundaries which remain vague in the description with adverbial variants come into focus once the object is mentioned explicitly. The places delimited (*on top of the figure*) are then explicitly situated beyond the outer boundary of the space delimited by the figure. Shifts of this kind do not occur with genuine adverbs (cf. Carroll & v. Stutterheim 1993), since adverbs cannot directly govern or form the head of a noun phrase (the cat is on top *the house). If a reference point/noun phrase has to be supplied, it is mapped into its own prepositional phrase which then specifies the sub-space in question (German: *oben* ***am Objekt****/obendran* (on top **at the object**/on top-there-at); *oben* ***im Objekt****/obendrin* (on top **in the object**/on top -there-in); *obndrüber* (on top-there-over), which specifies a subspace beyond the upper boundary of the object.

One context in which deictic concepts cannot be consistently applied in English is of crucial significance in the present task. They cannot be used to structure the place of an object as a whole, as the example with the L-shaped figure indicates. Whether speakers adopt a deictic or intrinsic viewpoint is reflected clearly in the type of expressions used to express spatial relations (English deictic: *in front, in back, behind*; intrinsic: *at the front; at the back;* German deictic: *vorne; hinten;* intrinsic: *an der Vorderseite; an der Rückseite*).

4.2 Accessing the lexicon

The following tables show the patterns of information organisation in the spatial domain and the kind of lexical term accessed for each of the 8 parts, beginning with German speakers (setting a/d: non-descript base, no addressee).

Given a holistic pattern of organisation which is deictic (4 speakers), terms which reflect form or function are accessed in 78% of the cases. These include terms for familiar objects such as gear or ring. Novel terms accessed for non-descript parts (*block, rectangle*) amount to 65%.

If a point by point frame is adopted where attention is focussed on individual places (5 speakers), the majority of terms accessed are simply *Stück* (piece) or *Teil* (part), (78%).

When the holistic frame is organised from an intrinsic perspective (5 speakers), the tendency to access terms relating to shape or function drops to 38%, compared to 78% for a holistic, deictic perspective.

German — Lexical terms accessed

Frame of Ref.	Speakers 15	No. Parts identified	Term:Shape related	Term:'piece'
Holistic Deictic	4	32	25/32 :78% novel :65%	7/32 :21%
Holistic Intrinsic	5	40	15/40 :38% novel :25%	25/40 :62%
Point by point	5	40	9/40 :22% novel :10%	31/40 :78%

(The remaining speaker lined up the pieces in order of assembly and simply referred to each one as "the next piece".)

The same pattern was found for holistic frames with a further set of 10 speakers (a/c), recorded this time with an addressee who carried out the assembly as instructed. (No speaker adopted a point by point perspective in this setting.)

German — Lexical Terms Accessed

Frame	Speakers 10	Parts	Term:Shape related	Term:'piece'
Holistic deictic	6	47	37/47:79% novel 31/47:66%	10/47 :21%
Holistic intrinsic	3	22	8/22 :36% novel 5/22 :23%	14/22 :64%
Point by point	-	-	-	-

(The remaining speaker made a number of false starts and was not considered in the analysis.)

The pattern for English speakers (setting a/c: nondescript object, addressee present) is similar in that the tendency to access terms which relate to shape or function is higher with a holistic pattern of organisation (33%), compared to a point by point frame (19%)
The percentage of cases in which form-related terms were accessed with an holistic intrinsic frame is less compared to German speakers.

English Lexical Terms Accessed

Frame	Speakers 10	Parts mentioned	Term: Shape related	Term: 'piece'
Holistic deictic	–			
Holistic intrinsic	3	24	8/24 :33% novel 5/24 :20%	16/24 :66%
Point by point frame	7	53	10/53 :19% novel 3/53 :6%	43/53 :81%

Speakers of English are more likely to start with the term *piece* or *part*, and supply shape-related information in the form of an analogy. This accounts for the lower rate of access for terms relating to shape or function (33%), compared to the German groups (38%, 36%) with an intrinsic frame.

Ex. 10 *okay now you've got* ***what looks like a bar***

Ex. 11 *you take the piece that looks like a sprocket*

Ex. 12 *you take the piece that looks like two eye beams*

No analogies were used to identify parts with a point by point frame in the English data. In the German data they occurred with only 2 out of a possible 40 parts. The use of an analogy is thus clearly linked to a holistic perspective.

In summary, the tendency to access lexical items which relate to the shape or function of individual objects is highest when speakers structure space holistically from a deictic perspective.

4.3 The role of the spatial domain in determining the lexical forms accessed

The same task was repeated with a base unit with intrinsic front and back sections in order to test the role of the spatial domain in determining lexical access. If the structure of spatial domain is the main determining factor lexical access, all speakers who adopt a holistic (intrinsic) frame of reference should access a certain number of shape-related terms.

As mentioned above, the base of the unit used in the first task was modified by adding other elements. Some new parts were required to fit the modified elements, but they were similar to those in the first task in that they did not resemble familiar objects. Since there were two gears in this setting, there are automatically more shape-related terms, compared to the first setting. Comparisons between the two settings are therefore based on the number of novel terms.

German

Lexical Terms Accessed

German	Speakers 10	Parts identified	Term:Shape related	Term:'piece'
Holistic deictic	2	16	12/16 :75% novel 8/16 :50%	4/16 :25%
Holistic intrinsic	8	64	18/64 28% novel 8/64:13%	46/64 :72%
Point by point	-			

The results show that an intrinsic perspective is adopted by the majority of speakers (8 out of 10) when the entity under assembly has salient front and back sections. However, an intrinsic frame of reference does not lead to the selection of novel items to the degree found in the previous settings (novel terms accessed in the first settings reach 23% and 25% versus 13% in the present setting). This pattern also applies to speakers who adopted a deictic perspective (first settings 66%;65% versus 50% in the present case).

English

Lexical Terms Accessed

Frame	Speakers 10	Parts identified	Term:Shape related	Term:'piece'
Holistic deictic	-			
Holistic intrinsic	10	80	20/80 :25% novel 8/80 :10%	60/80 :75%
Point by point	-			

The results for the English group show a corresponding pattern. The number of novel terms accessed drops from 20% to 10%.

The comparison for the two settings indicates that innovation in lexical access which comes with a holistic perspective is more likely to occur with a non-descript entity, as given in the first setting. This means that the lexical items accessed in identifying individual objects is determined by the way speakers conceptualise the object under assembly (object domain). The tendency to access novel terms for the different parts is more constrained when the spatial frame of reference is based on intrinsic features of the object. Speakers of English are bound to an intrinsic pattern of conceptualisation when organising information from a holistic viewpoint.

5. Discussion and Conclusions

The results show how lexical access is determined by the different ways in which information is organised conceptually for a specific communicative task. Speakers conceptualise the unit under assembly and its parts in different ways. The different mental models of the unit correlate with the way individual parts are identified in the assembly process. The parts are identified in either neutral or novel terms (*piece* versus *crank device* for the same object) depending on the conceptual structure established in organising information.

The unit can be represented **holistically** in terms of its inherent features and part-whole relations. In this case the frame of reference established to specify locations on the unit is structured in corresponding terms. The space delimited by the base is thereby divided into sub-spaces (front, back, left hand side, right hand side etc.) which encompass the space as a whole. Given this pattern of conceptualisation, speakers access novel terms for the individual pieces which have to be placed on the base of the unit. The terms reflect either inherent features of each piece or their function (*rectangle, axis, block, wedge*). In other words, the individual pieces are also viewed holistically on the basis of their place in a general scheme of things. An holistic frame of reference is not an ideal solution for lexical access in this task since the search for novel terms is frequently accompanied by marked pauses of two seconds or more, thus indicating that the selection of a holistic structure is not driven by lexical knowledge.

In the second pattern of information organisation observed, speakers conceptualise the unit **as an entity made up of individual parts**. The spatial concepts used to structure space are based on individual pieces (some of which were already on the unit at the outset) and the places they delimit. The other places which can function as reference points in the spatial frame of reference emerge in time on a point by point basis with each new piece placed on the base. Given a pattern of conceptualisation of this kind in which attention is directed to individual pieces and associated places, the pieces under assembly are all reduced to the same status and are referred to simply as a *part, piece, (Teil, Stück)*.

Object and form are treated as identical given a holistic perspective. The individual pieces are referred to by means of terms such as *block, ring, axis*. In point by point frames, on the other hand, the conceptual structure which dominates in the underlying model sets the stage for identification on the basis of a common structure (a piece) and a general catalogue of features - size, shape and colour. Distinguishing features are encoded attributively in the form of adjectives or as separate statements.

All speakers in this task participated in the same situation, they perceived the same set of objects, and all had the opportunity of manipulating them while practising the assembly. Nevertheless speakers choose to draw on some aspects of their knowledge only in organising information for expression in the communicative task. They may generate a mental model which leads them to ignore the possibility of simply calling each part a *piece*, for example. The study shows how models hold consistently in language production for the entire task. Lexical terms are accessed which fit a consistent schema or model of things. Options available in viewing conceptual material are thus selected or "narrowed" in a systematic way.

In existing models of language production the term 'language-specific' is used in different ways. It involves cognitive processes which relate to language processing in general (generation of messages with a propositional structure, for example) and thus stand in contrast to general cognitions; the term also relates to structures which are typical for a single language or language group. In existing models of language production the conceptualiser is seen to accomodate processes which hold in general for the expression of information through language, irrespective of individual structural requirements, leaving processes which are specific to any one language to the formulator.

The present study provides evidence of patterns of conceptualisation which relate to structural requirements of a specific language. Speakers of English have to relate to inherent features of an object in order to establish a frame of reference which applies to the object as a whole. The morpho-syntactic structure of the set of locative expressions (object-centred) available in English constrains the formats possible in the mental model for the spatial domain. Spatial structures such as the set of coordinate axes are generated from an intrinsic, object-centred viewpoint only by English speakers, given a holistic frame of reference. Frames of reference based on deictic, subject-centred viewpoints, for example, cannot be applied to structure places as a whole. English has deictic expressions but their morpho-syntactic features prevent their application in structuring the place of an object holistically (to give front, back, top and bottom sections). The application of a holistic spatial structure to cover the entire place delimited by an object is confined in the system to intrinsic viewpoints only. Objects must have features which warrant use of the structure applied.

Since an intrinsic object-centred model holds sway in information organisation, object-centred constraints also hold in accessing lexical terms to identify individual object parts. Highly novel terms are introduced in the form of an analogy (*you take the piece that looks like a sprocket*). An object-centred perspective constrains the eye of the beholder so to speak.

In contrast, speakers of German have the option of setting up frames of reference from a deictic, subject-centred perspective, defining a front and back sub-space from the viewpoint of an observer, fictive or otherwise. They do not have to relate to an intrinsic feature of an entity when structuring it holistically in terms of part-whole relations. The licence which a subjective standpoint allows is also reflected in the extent to which novel lexical terms are applied in identifying individual parts of the object in the assembly process, as well as in the relatively low occurrence of analogies.

When German speakers adopt an intrinsic, object-centred pattern of conceptualisation (reflected in the spatial expressions used), the lexical terms accessed compare with those found for English speakers. The terms accessed in identifying individual parts are constrained by the object domain. Novel terms are not applied to the extent observed with a deictic, subject-centred perspective.

The cross-linguistic comparison shows how conceptual knowledge is implemented in language-specific terms, thus providing evidence for the existence of a novel form of conceptualisation which sets guidelines for information organisation in language production. Speakers know that they can call an entity a *piece* in a task of this kind. This information is available in memory, but they may establish a conceptual structure in organising information for expression in which this does hold as a satisfactory solution. The conceptual structure does not allow open access to the range of means memory would provide, willingly or otherwise.

Can information organisation of this kind be handled solely at the level of an encoder or formulator? The formulator encompasses knowledge on a systemic level of the semantic forms and corresponding conceptual structures which a language provides. English has deictic and intrinsic means of expressing spatial relations, and with this the corresponding conceptual structures. But the constraints which hold relate to the contexts in which deictic forms can be applied. The semantic system provides the means to express deictic spatial relations but it does not deliver on the contexts in which the means available will fail to meet a required communicative goal. The means in question cannot function in a spatial frame of reference organised from a holistic perspective. If the encoder were the only source of language-specific knowledge of this kind, speakers would not know what type of spatial relations they were talking about until they had consulted the encoder. Knowledge of the goals which semantic concepts can serve belongs to the level of information organisation. The conceptualiser in language production is not merely a conduit which serves in mapping information focussed in memory into language-dependent formats of a general nature (predicate-argument structure). Information flow between the conceptualiser and formulator must occur on an interactive basis.

The results show how the underlying conceptual structure established in organising information provides speakers with far-reaching guidelines for the encoding process. The mental model provides a skeletal structure for the kind of lexical material which should find expression. Specifications go beyond general categories such as 'specific object' to list certain concepts (concepts underlying the option *piece* versus *crank device* for identification of the same object). The task of lexical search is channeled for the speaker to the selection of appropriate tokens within a certain category.

The quaestio approach sets out to model general patterns of information organisation of this kind in language production. The evidence underlines how languages require speakers to "adopt a particular framework for schematizing experience" (Slobin 1991:7) when "thinking for speaking".

References

Appelt, D.E. (1985). *Planning English Sentences*. Cambridge: Cambridge University Press.

Ariel, M. (1990). *Accessing noun-phrase antecedents*. London, New York: Routledge.

Becker, A. (1994). *Lokalisierungsausdrücke im Sprachvergleich*. Tübingen: Niemeyer.

Bierwisch, M. & Schreuder, R. (1992). From concepts to lexical items. *Cognition, 42*, 23-60.

Bock, K., & Levelt, W. (1994). Language Production. Grammatical Encoding. In M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics*. San Diego: Academic Press, 945-984.

Bühler, K. (1934). *Sprachtheorie*. Jena: Fischer.

Carroll, M. (1993). Deictic and Intrinsic Orientation in Spatial Descriptions: A Comparison Between English and German. In J. Altarriba (Ed.), *A Cross-Cultural Approach to Psychology*. New York: Elsevier.

Carroll, M., & von Stutterheim, C. (1993). The representation of spatial configurations in English and German and the grammmatical structure of locative and anaphoric expressions. *Linguistics, 31*, 1011-1041.

Clark, H., & Marshall, C.R. (1981). Definite reference and mutual knowledge. In A.K. Joshi, B.L. Webber, & I.A. Sag (Eds.), *Elements in Discourse Understanding*.

Cambridge, MA: MIT Press.
Clark, H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*, 1-39.
Danlos, L. (1987). *The Linguistic Basis of Text Generation*. Cambridge: Cambridge University Press.
Garrod, S.C., & Sanford, A.J. (1988). Discourse models as interfaces between language and the spatial world. *Journal of Semantics, 6*, 147-160.
Garnham, A. (1989). A unified theory of the meaning of some spatial relational terms. *Cognition, 31*, 45-60.
Givon, T. (Ed.), (1983). *Topic continuity in discourse: A quantitative crosslanguage study*. Amsterdam: John Benjamins.
Greeno, J.G. (1983). Conceptual entities. In D. Gentner & A.L. Stevens (Eds.), *Mental Models*. Hillsdale, NJ.: Lawrence Erlbaum.
Graumann, C.F. (1989). Perspective setting and talking in verbal interaction. In R. Dietrich & C.F. Graumann (Eds.), *Language processing in social context*. Amsterdam: North Holland. 95-113.
Habel, C., Herweg, M. & Rehkämper, K. (1989). *Raumkonzepte in Verstehensprozessen*. Tübingen: Niemeyer.
Hawkins, J. (1978). *Definiteness and indefiniteness: A study in reference and grammatical prediction*. London: Croom Helm.
Herrmann, T. (1989). Vor, hinter, rechts und links: das 6-H Modell. Psycholinguistische Studien zum sprachlichen Lokalisieren. Sonderforschungsbereich 245, Bericht Nr. 15. Lehrstuhl Psychologie III, Universität Mannheim.
Herrmann, T. & Deutsch, W. (1976). *Psychologie der Objektbenennung*. Bern: Huber.
Herrmann, T. & Grabowski, J. (1994). Sprechen: Psychologie der Sprachproduktion. Heidelberg: Spektrum, Akademischer Verlag.
Herskovits, A. (1986). *Language and spatial cognition*. Cambridge: Cambridge University Press.
Jackendoff, R. (1983). *Semantics and Cognition*. Cambridge MA: MIT Press.
Jackendoff, R. & Landau, B. (1991). Spatial Language and Spatial Cognition. In D.J.Napoli & J.A. Kegl (Eds.), *Bridges between Psychology and Linguistics: A Swathmore Festschrift for Lila Gleitman*. Hillsdale, NJ.: Lawrence Erlbaum. 145-169.
Johnson-Laird, P.N. (1983). *Mental Models*. Cambridge: Cambridge University Press.
Kamp, H. (1981). A theory of truth and semantic representation. In J.Groendijk, T. Janssen, & M. Stokhof (Eds.), *Formal methods in the study of language*. Mathematical Centre Tracts, Amsterdam.
Klein, W. (1991). Raumausdrücke. *Linguistische Berichte, 132*, 77-115.
Klein, W., & von Stutterheim, C. (1987). Quaestio und referentielle Bewegung in Erzählungen. *Linguistische Berichte, 109*, 163-184.
Kohlmann, U. (1996). Objektreferenzen in Beschreibungen und Instruktionen. Eine empirische Untersuchung zum Zusammenhang von Textstruktur, referentielle Bewegung und Formen von Objektreferenzen. Unpublished manuscript, University of Heidelberg.
Koelbing, H.G., Mangold-Allwinn, R., Barattelli, S., Kohlmann, U., Stutterheim, C.v. (1994). Wie beeinflußt der Ausführende den Instruierenden?: Zum Einfluß eines anwesenden Partners auf die Wahl von Objektreferenzen in Instruktionstexten. *Sonderforschungsbereich 245, Bericht Nr. 85*, Lehrstuhl Psychologie III, Universität Mannheim.

Lang, E., Carstensen, K.-U., & Simmons, G., (1991). *Modelling Spatial Knowledge on a Linguistic Basis*. Berlin: Springer Verlag.

Levelt, W. (1984). Some perceptual limitations on talking about space. In A. von Doorn, W. de Grind, & J. Koenderink (Eds.), *Proceedings of the Tenth International Congress of Phonetics Sciences*. Dordrecht: Foris.

Levelt, W. (1989). *Speaking. From intention to articulation*. Cambridge, MA.: The MIT Press.

Levelt, W., Schriefers, H., Vorberg, D., Meyer, A.S., Pechmann, T. & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming.*Psychological Review, 98*, 122-142.

Mangold, R. & Pobel, R. (1988). Informativeness and instrumentality in referential communication. *Journal of Language and Social Psychology, 7*, 181-191.

McKeown, K.R. (1985). *Text Generation: Using Discourse Strategies and Focus Constraints to Generate Natural Language Text*. Cambridge: Cambridge University Press.

Morrow, D. & Greenspan, S. (1989). Situation models and information accessibility. In N.E. Sharkey (Ed.), *Models of Cognition. A review of cognitive science*, 53-77. Norwood: Ablex.

Olson, D.R. (1970). Language and thought. Aspects of a cognitive theory of semantics. *Psychological Review 77*, 257-273.

Rauh, G. (Ed.), (1991). *Approaches to Prepositions*. Tübingen: Gunter Narr Verlag.

Slobin, D.I. (1991). Learning to think for speaking. *Pragmatics, 1 (1)*, 7-25.

Stutterheim, C. von & Klein, W. (1989). Text structure and referential movement. In R. Dietrich & C.F. Graumann (Eds), *Language Processing in Social Context*. Amsterdam: North Holland, 39-70.

Stutterheim, C. von (1995). Einige Prinzipien des Textaufbaus. Empirische Untersuchungen zur Produktion mündlicher Texte. Unpublished Habilitationsschrift, University of Heidelberg.

Talmy, L. (1983). How language structures space. In H. Pick & L. Acredolo (Eds.), *Spatial Orientation: Theory, Research and Application*. New York: Plenum Press.

Connectionist Syntax and Planning in the Production of Object Specifications*

Hans-Jürgen Eikmeyer, Ulrich Schade,
Marc Kupietz, Uwe Laubenstein

Universität Bielefeld, Fakultät für Linguistik und Literaturwissenschaft,
Postfach 10 01 31, D-33501 Bielefeld

Abstract

We present a connectionist model for the cognitive process of language production which, in contrast to other such models, describes the production of linguistic expression more complex than single words, namely referential noun phrases or, in psycholinguistic terms, object specifications. The model thus accepts the challenge to describe not only the process of lexical access, but also higher level cognitive planning of "what to say". Object specifications are a good starting point for such an enterprise because German language psychologists have extensively studied this topic during the last decades and therefore provide a lot of empirical data against which the model can be tested. The model is also suitable for other aspects of language production which we will shortly address.

1 Introduction

The model presented below shares many features and convictions with the models of Berg [1988], Dell [1986], [1988]; Dell & Reich [1981]; Harley [1984]; MacKay [1987]; Stemberger [1985]. Like these models, it is a local connectionist model, i.e. every linguistic unit is represented by a single node. Such nodes are collected in a subnetwork, the network of *content nodes*. It is hierarchically organised in layers such as the concept layer, the word layer, the syllable layer and the phoneme layer. The last layer determines the output of the production model in

*Work on this project has been funded by the Deutsche Forschungsgemeinschaft within the framework of the SFB 360 "Situierte Künstliche Kommunikatoren", Project B2 "Computersimulation von Prozessen der Objektbenennung".

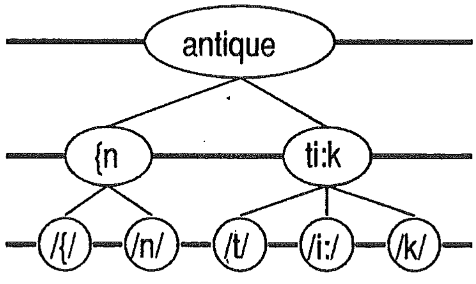


Figure 1: A part of the content network for the word, syllable and phoneme layer. Excitatory connections are depicted as thin lines and lateral inhibition is indicated by bold lines.

accordance with Dell [1986] who presents arguments in favour of such a design on the basis of speech error data. Dell also argues that the syllable level is cognitively relevant, because phonological speech errors respect syllable positions.

For these models language production thus means bringing about the sequence of phonemes of which a linguistic expression consists. The selection of the phoneme to be produced at a certain point in time is determined by the distribution of activation in the phoneme layer. This is done in such a way that the phoneme which has maximal activation is selected and counts as being produced by the model. The activation of phonemes is determined by the activation of syllables, which in turn is determined by the activation of words. Finally, the activation of words is determined by the activation of conceptual nodes. For this purpose there are bidirectional excitatory connections between a concept and all words which can be used to verbalise it, between a word and all its syllables, and between a syllable and all its phonemes (see Fig. 1)[1]. Since a word activates all its syllables and a syllable activates all its phonemes, it is obvious that a mechanism is needed which assigns maximal activation to a word's syllables and, in turn, a syllable's phonemes. There are several proposals available for such a *linearisation device* in the literature quoted above. We will present our solution to the linearisation problem in section 3. Our model differs from some of the aforementioned models in another respect also since it uses *lateral inhibition* on all layers. An argumentation in favour of this design can be found in Berg & Schade [1992]; Schade [1992]; Schade & Berg [1992], but see also Dell & O'Seaghdha [1994].

Since the model's architecture is based upon speech data it is quite naturally able to model *slips of the tongue* (cf. Schade [1992]) by adding noise to the system or by increasing the speech rate. *Aphasic speech* can be equally well modelled by manipulating the strength of excitatory and/or inhibitory connections (cf. Eikmeyer & Schade [1993]; Harley [1990], [1995]; Harley & MacAndrew [1992];

[1] We use the SAM Phonetic Alphabet in this figure. /{/ corresponds to /æ/.

Schade [1996]; Schwartz et al. [1994]). Another variant of the model has been created for Finnish word production (cf. Tikkala et al. [1995]; Tikkala [1995]). Finnish is a *morphologically complex language* where nouns can have suffixes for case and personal possessor and several enclitic suffixes. Moreover, morphological and syllabic structure are in some sense incongruent since e.g. case morphemes are spread over two adjacent syllables: the coda of the first syllable and the onset and nucleus of the second syllable are determined by the case morpheme. The variant of the model presented here describes linguistic expressions which are *syntactically more complex* than single words, namely referential noun phrases or object specifications.

2 Object Specifications

If a speaker wants her partner to identify an object, *the target object*, within a set of *context objects*, she produces an object specification, i.e. a referring noun phrase. The primary question to be asked is what properties of the target object are verbalised under what conditions. Obviously, it seems natural for a speaker to verbalise *discriminative* features of the target object, where a feature (property) is *discriminative* iff there is at least one context object which differs from the target object with respect to this feature. The initial observation within language psychology was, that the specification varies with the context objects (Olson [1970]). This is plausible since the discriminativity of a feature depends upon the context objects.

Research carried out by German language psychologists (cf. e.g. Herrmann & Deutsch [1976]; Herrmann & Grabowski [1994]; Mangold-Allwinn [1993]; Mangold-Allwinn et al. [1992]; Pechmann [1984], [1989], [1994]) led to further results which we will partially touch upon in the following. One important finding shall serve as a starting point for the exposition of our model: speakers do not only verbalise discriminative features, but also non-discriminative ones. When we call an object specification *minimal* iff it contains a minimal set of discriminative feature, we can rephrase this finding in such a way that speakers do not always produce minimal specifications. They frequently produce *overspecifications* containing features not necessarily needed for an identification. The colour feature e.g. is often verbalised by speakers although it is not necessary. We will start the presentation of our model from the perspective of minimal specifications and turn to overspecifications in a second step. As a prerequisite, however, we first have to explain the linearisation mechanism of the model in general.

3 Linearisation

An important aspect of psychological studies of actions relates to the problem of serial order, i.e. the problem of the correct sequence of the parts of an action (cf. Lashley [1951]; Bruce [1994]). This problem is equally relevant for linguistic actions as parts of the language–production process since linguistic utterances have a linear or sequential structure. The words of an utterance follow one another as do the syllables of a word or the phonemes of a syllable.

The majority of speech errors are errors in the linear order which proves that humans have problems in bringing about a desired order. These errors and their distribution can thus serve as starting point for cognitive models of language production in general and the linearisation device in particular. It has been shown that connectionist models, which are well suited for the description of speech errors, need a way to represent syntactic and morphological knowledge which can be exploited in the realisation of correct sequences of linguistic units (cf. Dell, Burger & Svec [1995]; Eikmeyer & Schade [1991]). Our model uses *networks of control nodes* for this purpose. Since linearisation is a problem on all linguistic levels, such networks are needed for all levels. The nodes in a control network, the *control nodes*, represent syntactic, morphological and phonological categories. These networks are hierarchically organised according to the linguistic level to which they correspond. The networks thus reflect dominance and linear precedence relations.

Linear constraints can either be strict or they can allow for a range of possibilities and the linearisation device presented can model both. Such control networks consist of a number of *control nodes*, and it is guaranteed than at any point in time exactly one control node is highly activated whereas all other nodes have a very low activation. After some time, activation is spread from such a highly activated control node to one of its successors. Spreading of activation is monitored by *gating nodes* which, so to speak, act as valves in the excitatory connection of one control node to its possible successor nodes. Since all gating nodes in connections leading to rivalling successors *mutually inhibit* each other, the next control node is determined by the gating node with maximal activation. Figure 2 depicts three control networks, one for the syntax of noun phrases (i.e. a "phrase network"), one for the arrangement of syllables in words (i.e. a "word network"), and one for the CV–structure of syllables (i.e. a "syllable network").

The phrase network allows for the production of noun phrases according to the rules specified in (1a) and (1b). The word network describes one and two syllable words according to (2) and the syllable network describes syllables according to rule (3).

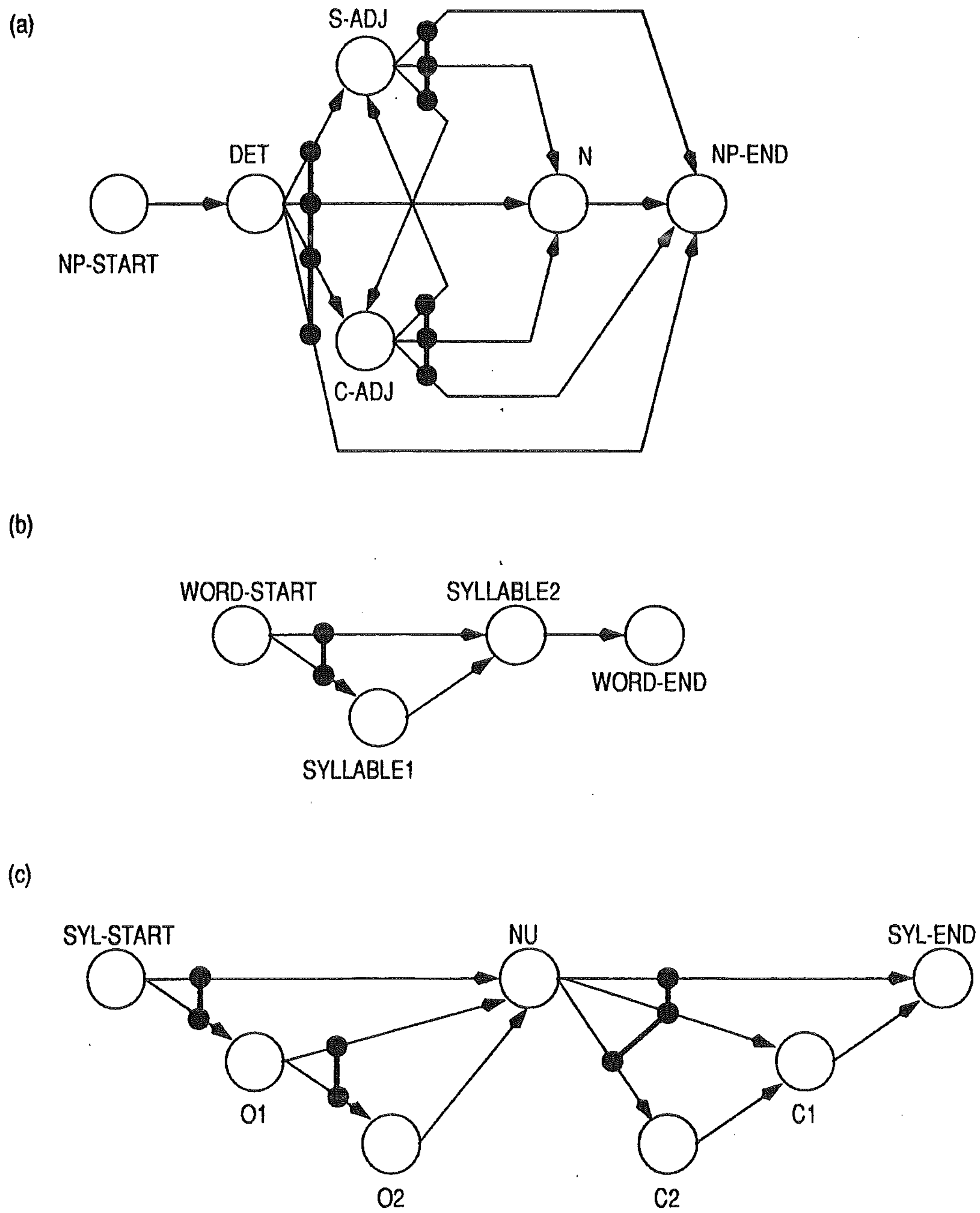


Figure 2: Control networks for three levels: (a) a phrase network (b) a word network and (c) a syllable network. Gating nodes are represented by small black dots, control nodes by the bigger circles. Excitatory connections are depicted as arrows and mutual inhibition is indicated by bold lines.

(1a) NP → DET (S-ADJ) (C-ADJ) (N)
(1b) NP → DET (C-ADJ) (S-ADJ) (N)
(2) WORD → (SYLLABLE1) SYLLABLE2
(3) SYLLABLE → (O1 (O2)) NU ((C2) C1)

Abbreviations used:
S-ADJ and C-ADJ for size and colour adjectives, respectively;
O1, O2 and C1, C2 for onset and coda consonants, respectively;
and NU for syllable nuclei.

Besides the activation flow within a control network described above, there is another activation flow between content nodes and control networks. The gating nodes receive activation from content nodes and the control nodes send activation to (other) content nodes. With respect to the syllable level, e.g. all syllables without an onset activate the gating node in the connection between the nodes SYL-START and NU, all other syllables activate the gating node between the nodes SYL-START and O1. All syllables with a single onset activate the gating node from O1 to NU, and all with a second onset consonant activate the gating node from O1 to O2. Analogous conditions hold for the coda nodes and for the gating nodes in all control networks. In general, content nodes of a level activate gating nodes in the corresponding control network.

Turning to the phrase network, we assume that its gating nodes are activated from a conceptual level. We furthermore take the perspective of a minimal specification. In this case all gating nodes in connections leading to the node C-ADJ are activated if the *colour feature* is discriminative. All gating nodes in connections leading to the node S-ADJ are activated if the *size feature* is discriminative. Finally, all gating nodes in connections leading to the N node are activated if the feature *object class* is discriminative. This description does not answer the question of how discriminativity of features is modelled, but we will address this point in the following section.

Control nodes not only activate other control nodes (their successors) but also content nodes. More specifically, a control node acitivates all linguistic units belonging to the category it represents. The node S-ADJ, e.g., is excitatorily connected to all size adjectives, the node C-ADJ to all colour adjectives, N to all nouns, etc. Thus, these connections legitimate the name assigned to the control nodes, which plays no role in the dynamic behaviour of the system. Node names are only used for purposes of reference and have been chosen on mnemotechnic grounds. The model thus uses a uniform linearisation mechanism for all levels, and it makes no distinction between phrase syntax, word syntax (= morphology) and syllable syntax (= morphophonology).

From the previous discussion it follows that content nodes may receive activation from two sources: they are activated directly from within the content network (see section 1) and by control nodes. Since there is exactly one highly active control node on every level, there is one class of linguistic entity on every level which is activated from the control networks. If, e.g. the nucleus node NU is active, all nucleus phonemes receive activation from here. Since within the network of content nodes all phonemes of the current syllable are activated from the syllable node, the nucleus of this syllable receives activation form two sources, while all other phonemes receive activation from only one source. Under normal, undisturbed conditions a node activated by two sources will be more highly activated than all other nodes equally activated by one of these sources and **not** activated by the other source. This is thus the basic principle of the linearisation device which is based on spreading of activation only.

4 Planning an object specification

The production of an object specification amounts to following a path in the phrase control network. Following a path in a control network means guaranteeing a subset of it's control nodes being active one after another. This, in turn, means that the gating nodes leading to these control nodes are the most active among their rivals. Finally, this means that the corresponding features have to be discriminative. If several paths can be taken in a control network, the feature which is more discriminative than another one determines the route chosen.

In our model the discriminativity of a feature is determined by special conceptual networks, one for each dimension (feature, property). With respect to colour and size these networks are shown in Fig. 3. They always contain pairs of nodes for each feature of a dimension. The target component of each pair is eventually activated by the target object, the context component is eventually activated by context objects. The connections between these nodes vary with the type of the feature. If the feature in question is (at least as a first approximation) rather a context-independent feature, as e.g. colour, shape or object class, we use a simple connection scheme with the target feature node inhibiting its paired context feature node. If the feature in question is a relative feature, as e.g. size, we need a more complicated scheme.

Let us assume that the speaker wants to refer to a long yellow bolt in the context of a shorter yellow bolt and an equally long blue bolt. In this case target YELLOW is activated by the target object, and context YELLOW and context BLUE are activated by the two context objects. Since target YELLOW inhibits its paired context node, only BLUE stays active among the colour context nodes. The fact that there is an active context colour node thus indicates that the colour feature is discriminative. As a consequence, all context colour nodes will activate gating nodes on connections leading to the C-ADJ control node in the phrase network. This control node, in turn, will activate all colour adjectives (cf. section 3). The target colour node YELLOW activates the word node `yellow`. Since this node is also a colour adjective node and thus receives activation from two sources, it will

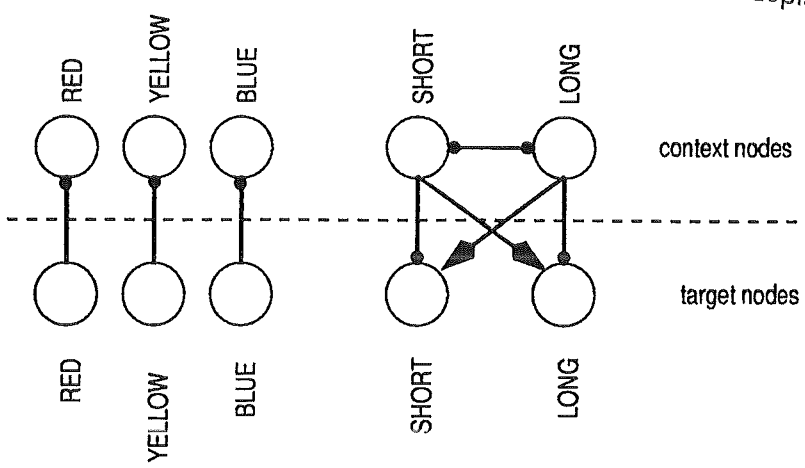


Figure 3: Context and target nodes for the dimensions colour and size

be produced in due time.

Size is a relative feature: an object is neither short nor long as such, but only in comparison to other objects. The target object can thus be qualified as long if there are shorter context objects around (and no longer ones). Consequently, a context object shorter than the target object, as given in the example presently discussed, will activate the context node SHORT. This node activates target LONG and inhibits the other two nodes of the size–dimension network. As in the case of colour we have an active context node, which indicates that the size feature is discriminative. Moreover, the target node LONG is active and it activates the adjective node `long`. Again, this adjective will be produced in due time.

The example discussed thus far was constructed in such a way, that there is a unique minimal set of features for the target object. With respect to this example it is an open question as to which of the two adjectives will be produced as the first one and which as the second in the corresponding noun phrase or specification (see the *problem of the order of adjectives* below). In a sligthly different example with the same long yellow bolt as target object and a shorter blue bolt as single context object, the model as described thus far would also specify both the colour and the size feature and not one of the two possible minimal specifications "`the yellow bolt`" or "`the long bolt`". But this would lead to over–specifications in general. As a remedy, the model has an inhibitory mechanism which inhibits all context objects (and their features) as soon as they are excluded by a verbalised feature. This inhibition is spread along connections running parallel to the excitatory connections from objects to features, but in the opposite direction. Depending on the strength of these connections it takes more or less time until the context objects are completely inhibited. Thus, the inhibition strength determines the degree of over–specification. This guarantees one of the two minimal specifications to be produced. However, it does not answer the question as to which one is actually produced by the model (see the *problem* *of multiple specifiability* below).

5 Interaction of dimensions

The first of the two questions addresses the *problem of the order of adjectives*, the second the *problem of multiple specifiability*. Our model gives a uniform answer to both questions at the same time. Since all operations in a connectionsit network depend upon the degree of activation of nodes, discriminativity is also a matter of degree. It is modelled by the activation value of a corresponding gating node. With respect to the order of adjectives our model predicts, that the adjective corresponding to the more discriminative feature will be produced first. With respect to multiple specifiability, the prediction is that only the more discriminative one will be produced (in case of a minimal specification). Before we discuss this further, however, we have to present some empirical facts relating to the two problems and other approaches addressing them.

Herrmann & Deutsch [1976] were the first to study multiple specifiabilty with respect to dimensions of spatial extension (breadth of objects vs. their height). They came to the conclusion that the dimension with the biggest difference between context and target object decides the choice of the dimension. However, preferences for one of these dimensions can override the principle of maximal difference. Pechmann ([1989], [1994]) studied adjective order with respect to colour and size. He manipulated the difference with respect to both dimension, but he found a constant ratio of 70-80% for the sequence "size before colour" and 20-30% for the sequence "colour before size". Pechmann ([1994]) explains his results with a strong syntactic preference for the more frequent sequence.

In an experiment we carried out in cooperation with Swedish colleagues[2] for German and Swedish we could replicate the 80% "size before colour" sequences, but half of the remaining specifications show the sequence "colour before size" and half of them were expansions as e.g. "`the yellow bolt, the long one`", i.e. with the colour adjective in front of the noun and the size adjective after it. This pattern had not been taken into account by Pechmann (personal communication). However, it nicely demonstrates the incremental character of the production of object specifications (see section 6).

Our model explains both problems by degrees of discriminativity. The more colour context nodes are activated, the higher the activation of the corresponding gating node is. If there are several context objects with the same colour, they do not act that much on the gating node due to a saturation of the colour context node. Things are similar with respect to size: The more context objects differ from the target object in size, the higher the activation of the context nodes and the gating node is. It is an open question as to whether the predictions of our

[2]This was a cooperation with the University of Gothenburg, Linguistics Department and our faculty. This cooperation was supported by the DAAD (German Academic Exchange Agency)

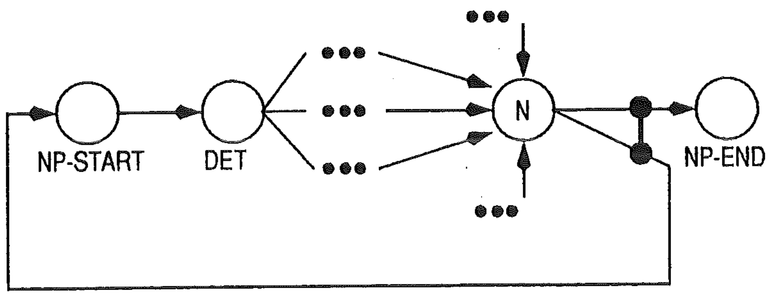


Figure 4: Phrase network for expansions

model are correct. If it turns out that Pechmann's assumption of a strong syntactic preference is correct, we can make the size dimension more discriminative from the start by assigning a higher resting value to the corresponding gating node.

6 Expanded Noun Phrases

Pechmann [1989] carried out some eye movement studies on subjects producing object specifications. He found that his subjects did not fixate all context objects before they start the production. Together with the expansion productions mentioned above this shows that not the set of context objects is not constant during the production of a specification, i.e. it may increase in size. In order to model the production of expanded noun phrases we modified the phrase network as shown in Fig. 4[3]. There is an additional connection from the node N leading back to the starting node. It has a gating node activated by **all** context dimension nodes while the gating node in the other connection has a small resting value.

If the context grows larger during the production of a specification, then features which had not been discriminative before may become so. This entails that gating nodes which had not been active from the start become activated. Depending upon which control node of the phrase network is active, two cases are possible. (i) If the newly activated gating node is located on a connection which leads to a successor of the currently active control node, then the newly discriminative feature will be verbalised without any special procedure. The timing may, however, influence e.g. the order of adjectives. (ii) If the system has passed all newly active gating nodes, the gating node leading back to the start will be active, and the verbalisation of the feature in question is ensured by passing the control network for a second time.

[3]Only the modification of the connections from N to NP-END are shown here. Corresponding modifications have to be added to all connections leading to the final node.

7 Conclusions

We presented a version of a connectionist language–production model which addresses syntax and linearisation on several linguistic levels. Although we restricted ourselves to relatively simple patterns for referring noun phrases, we think that the approach is easily extendible to larger fragments of syntax. However, from the point of view of planning, much has to be done in order to set up the corresponding conceptual networks and its connections into corresponding control networks. With respect to the limited domain of object specifications, however, our simulations can be connected to image analysis systems developed in the SFB 360. Thus, the system produces specifications driven by real world image data of collections of objects.

References

[1988] Berg, T.: Die Abbildung des Sprachproduktionsprozesses in einem Aktivationsflußmodell. Tübingen: Niemeyer.

[1992] Berg, T. & Schade, U.: The role of inhibition in a spreading–activation model of language production. Part 1: The psycholinguistic perspective. Journal of Psycholinguistic Research **22**, 405–434.

[1994] Bruce, D.: Lashley and the problem of serial order. American Psychologist **49**, 95–103.

[1986] Dell, G.S.: A spreading-activation theory of retrieval in sentence production. Psychological Review **93**, 283–321.

[1988] Dell, G.S.: The retrieval of phonological forms in production: Tests of predictions from a connectionist model. Journal of Memory and Language **27**, 124–142.

[1994] Dell, G.S. & O'Seaghdha, P.G.: Inhibition in interactive activation models of linguistic selection and sequencing. In: Dagenbach, D. & Carr, T.H. (eds.) Inhibitory Processes in Attention, Memory, and Language, (p. 409-453). San Diego: Academic Press.

[1995] Dell, G.S., Burger, L.K. & Svec, W.R.: Language production and serial order: A functional analysis and a model. Technical Report UIUC-BI-CS-95-04. Urbana, IL: The Beckman Institut, University of Illinois.

[1981] Dell, G.S. & Reich, P.A.: Stages in sentence production: An analysis of speech error data. Journal of Verbal Learning and Verbal Behaviour **20**, 611–629.

[1991] Eikmeyer, H.-J. & Schade, U.: Sequentialization in connectionist language–production models. Cognitive Systems **3**, 128–138.

[1993] Eikmeyer, H.-J. & Schade, U.: The Role of Computer Simulation in Neurolinguistics. Nordic Journal of Linguistics **16**, 153–169.

[1995] Eikmeyer, H.-J. Schade, U. & Kupietz, M.: Ein konnektionistisches Modell für die Produktion von Objektbenennungen. Kognitionswissenschaft **5**, 108–117.

[1984] Harley, T.A.: A critique or top-down independent levels models of speech production: Evidence from non-plan-internal speech errors. Cognitve Science **8**, 191–219.

[1990] Harley, T.A.: Paragrammatism: Syntactic disturbance or failure of control? Cognition **34**, 85-91.

[1995] Harley, T.A.: Connectionist models of aphasia: A comment on Nickels. Language and Cognitive Processes **10**, 47–58.

[1992] Harley, T.A. & MacAndrew, S.B.G.: Modelling paraphasia in normal and aphasic speech. Proceedings of the 14th Annual Conference of the Cognitive Science Society (S. 378-383). Hillsdale, NJ: Erlbaum.

[1976] Herrmann, T. & Deutsch, W.: Pychologie der Objektbenennung. (= Studien zur Sprachpsychologie, Bd. 5). Bern: Huber.

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

[1951] Lashley, K.S.: The problem of serial order in behavior. In: Jeffress, L.A. (Eds.), Cerebral Mechanisms in Behavior (pp. 112–136). New York, NY: Wiley.

[1987] MacKay, D.G.: The Organization of Perception and Action: A Theory for Language and other Cognitive Skills. New York, NY: Springer.

[1993] Mangold–Allwinn, R.: Flexible Konzepte: Experimente, Modelle, Simulationen. Frankfurt am Main: Lang.

[1992] Mangold–Allwinn, R. Stutterheim, C.v., Barattelli, S. Kohlmann, U. & Koelbing, H.G.: Objektbenennung im Diskurs: Eine interdisziplinäre Untersuchung. Kognitionswissenschaft **3**, 1–11.

[1970] Olson, D.R.: Language and thought: Aspects of a cognitive theory of semantics. Psychological Review **77**, 257–273.

[1984] Pechmann, T.: Überspezifizierung und Betonung in referentieller Kommunikation. Universität Mannheim: Dissertation.

[1989] Pechmann, T.: Incremental Speech Production and Referential Communication. Linguistics **27**, 89–110.

[1994] Pechmann, T.: Sprachproduktion: Zur Generierung komplexer Nominalphrasen. Opladen: Westdeutscher Verlag.

[1992] Schade, U.: Konnektionismus — Zur Modellierung der Sprachproduktion. Opladen: Westdeutscher Verlag.

[1996] Schade, U.: Die Klassifikation von Benennstörungen in einem konnektionistischen Produktionsmodell. In: Rickheit, G. (Hrsg.), Studien zur Klinischen Linguistik — Modelle, Methoden, Intervention. Opladen: Westdeutscher Verlag.

[1992] Schade, U. & Berg, T.: The role of inhibition in a spreading–activation model of language production. Part 2: The simulational perspective. Journal of Psycholinguistic Research **22**, 435–462.

[1988] Schriefers, H. & Pechmann, T.: Incremental production of referential noun–phrases by human speakers. In: Zock, M. & Sabah, G. (eds.), Advances in Natural Language Generation, (Vol. 1, p. 172-180). London: Pinter.

[1994] Schwartz, M.F., Saffran, E.M., Bloch, D.E. & Dell, G.S.: Disordered speech production in aphasic and normal speakers. Brain and Language **47**, 52-88.

[1985] Stemberger, J.P.: An interactive activation model of language production. In: Ellis, A.W. (ed.) Progress in the Psychology of Language. (Vol. 1, p. 143-186). London: Erlbaum.

[1995] Tikkala, A., Eikmeyer, H.-J., Niemi, J. & Laine, M.: On the Production of Finnish Nouns — A Psycholinguistically Motivated Connectionist Model. Submitted for publication.

[1995] Tikkala, A.: MORPHO: A Connectionist Language Production Tool for Finnish Nouns. Report A/1995/1. Department of Computer Science and Applied Mathematics. University of Kuopio.

Martin, N., Dell, G., Saffran, E.M. & Schwartz, M.F. (1994). Origins of paraphasias in deep dysphasia: Testing the consequences of a decay impairment to an interactive spreading activation model of lexical retrieval. *Brain and Language* **47**, 609-660.

Efficient Processing of Spatial Relations in General Object Localization Tasks

Klaus-Peter Gapp
Cognitive Science Program
Dept. of Computer Science
Universität des Saarlandes
D-66041 Saarbrücken, Germany
Email: gapp@cs.uni-sb.de

Abstract

The localization of an object within a certain environment requires the selection of an appropriate reference object and the establishment of adequate spatial relations between the objects. The decision, which reference object to choose, demands consideration of all applicable spatial relations, which therefore have to be computed before the final decision can be made. However, the calculation of all applicable spatial relations for all potential reference objects is a computationally intensive task. In order to increase performance, it is proposed that the set of spatial relations to be tested can be reduced in two respects without any loss of quality in the generated answer. The two ways in which reduction can be achieved are by taking into account the type of the localization — approximate or precise — and the type of the reference object.

1 Introduction

Object localization is a task in which the integration of language and vision is a required necessity per se. The localization of an object can be processed in two directions: from vision to language (cf. [André et al. 88, Gapp 94c]) and from language to vision (cf. [Olivier et al. 94, Schirra & Stopp 93]). In the former, the position of objects in a given spatial scene is described propositionally or in natural language (NL), in the latter a spatial layout is generated from a NL description. In this paper we focus on the generation of descriptions for answering queries about the spatial position of objects in general domains. A typical query is therefore *"Where is the position of object A?"*, with *object A* being a 2- or 3-dimensional object in the domain.

The task of object localization can be divided up into three major parts: first, the object of interest, the LO, is identified, then an object suitable for use as a referent, the RO, is selected, and finally a linguistic description of the relationship between the objects from some specific point of view is specified (cf. [Gapp 95c]). In this paper we will place emphasis on examining the processing of spatial relationships between objects as a whole, which is more than merely the computation of essential relations.

First, a short review about an universal model for the computation of distance (topological) and directional (projective) relations is given. The subset of directional relations is then extended to internal projective and cardinal relations. The extended model provides the basis for our further considerations of the processing of spatial relations in general object localization tasks.

Possible applicable relations in a particular spatial configuration can be manifold and therefore computationally intensive. It is shown that in object localization tasks the set of spatial relations to be considered is situational dependent and can therefore be restricted to specific subsets in advance.

2 A Computational Model of Spatial Relations

Increasingly sophisticated approaches to the computation of spatial relations have been developed in the last couple of decades (e.g. [André et al. 88, Egenhofer & Herring 90, Mukerjee & Joe 90, Abella & Kender 93, Rajagopalan 94, Schirra & Stopp 93, Olivier et al. 94]). In [Gapp 94a] the CSR-3D[1] system, a model for the computation of topological and projective relations (cf. [Talmy 83, Herskovits 86]) in 2D and 3D space, was presented. The model was enhanced to include composite spatial relations, such as *to the left of and behind* or *to the left of and near* which are very common used in German (cf. [Gapp 94b]), and proved for cognitive plausibility (cf. [Gapp 95a, Gapp 95b]).

The essential idea of the model is to measure the distance (topological relations) and the angle of deviation (projective relations) from the LO's center of gravity to the RO, in a space which is scaled by the extension of the reference object, and to account for the vagueness of spatial relations by mapping these values to specific evaluation functions (cf. Figure 1). The compositional relations are obtained by applying a weighted minimum function to the elementary relations.

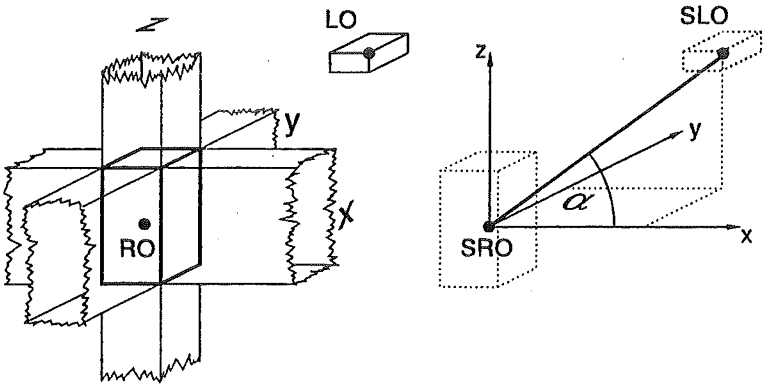


Figure 1: Distance and angle are measured in a space which is scaled by the RO's extension

To account for object localizations in large scale space, and for the interior of objects it is necessary to extend the set of directional relations to external and internal cardinal, e.g. *north* or *in the south*, and internal projective relations, such as *top* or *top left*.

2.1 Cardinal Relations

Cardinal relations, e.g. *south, south-west*, or *west-south-west*, are mostly used to ensure appropriate descriptions of the orientation of objects in large-scale geographic space. Cardinal relations are independent from a reference object's extension. The angle of deviation of the located object from the relation's canonical direction can therefore be measured absolutely. Three kinds of evaluation functions are used to model their accuracy (cf. Figure 2). The more specific the relation, the more restricted the evaluation function. The following ranges are defined:

[1]The CSR-3D system (Computation of Spatial Relations in 3D-Space) is part of the *VITRA* (VIsual TRAnslator) project which deals with the relationship between vision and natural language (cf. [André et al. 89]). The system starts from a geometric scene description, given by a vision system (cf. [Koller 92]), and proceeds to describe a recorded video sequence or to answer questions about the location of the objects involved in a cognitively plausible manner.

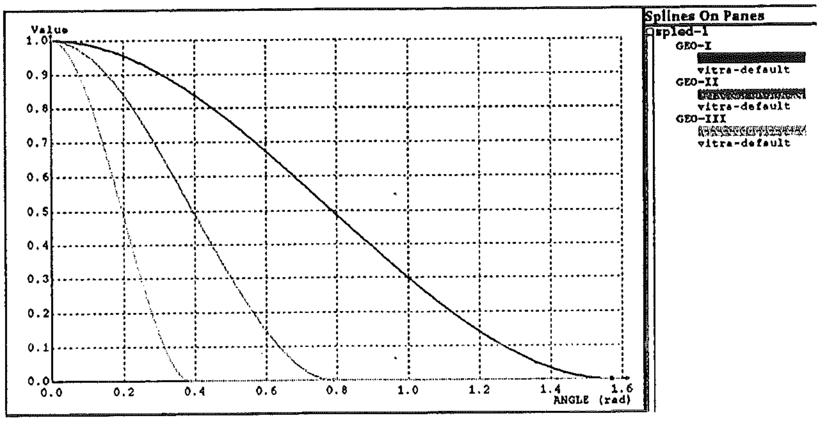


Figure 2: Definition of the evaluation functions for the three groups of cardinal relations

- 90° for the relations *north, east, south,* and *west* (GEO-I).
- 45° for the relations *north-east, south-east, south-west,* and *north-west* (GEO-II).
- 22.5° for the relations *north-north-east, east-north-east, east-south-east, south-south-east, south-south-west, west-south-west, west-north-west,* and *north-north-west* (GEO-III).

Figure 3 shows an example of how the different grades of evaluation functions affect the applicability structure of the relations in 2D space (3D space analogous). The darker the shading in the relation's region of applicability, the higher the degree of applicability.

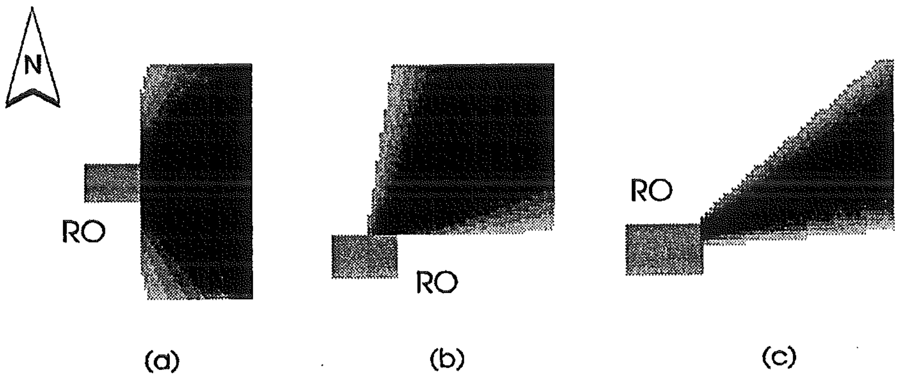


Figure 3: 2D applicability fields for the relations (a) *east* (b) *north-east* (c) *east-north-east*

2.2 Internal Directional Relations

In [Wazinski 92] a partitioning of the reference object into three parts for each dimension[2] was proposed in order to compute internal projective relations. The regions covered by an object were used to compute its relationship to the reference object.

[2] Only 2-dimensional space was considered.

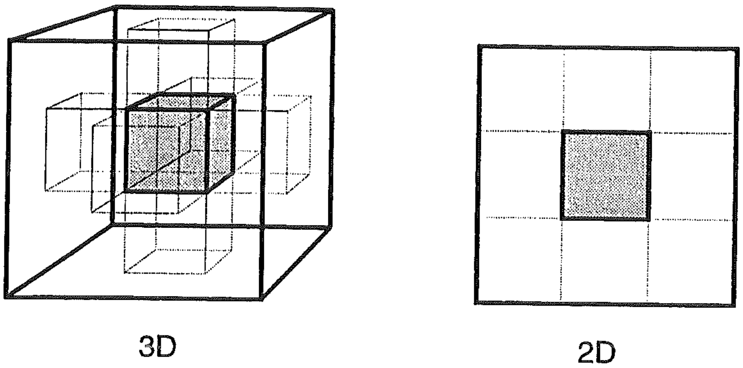


Figure 4: Partitioning of the reference object to compute interior projective relations

By dividing up the reference object into three regions for each dimension and using the center region as the reference, CSR-3D now accounts for relations such as *front, back, top, center, top left*, etc. (cf. Figure 4). In contrast to Wazinski's approach, our model can cope with internal projective relations in 3D as well as in 2D space and there are no sharp boundaries between neighboring relations. Also the applicability regions between internal *right/left* and external *to the right/left of* merge consistent into one another as shown in Figure 5.

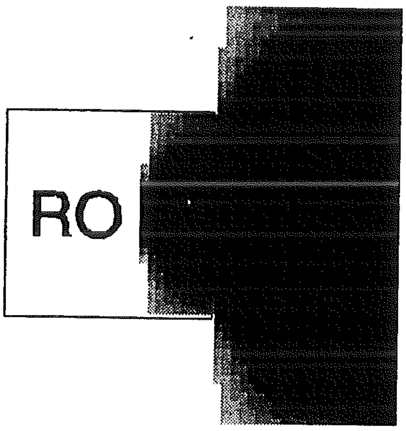


Figure 5: Applicability regions of internal *right* and external *to the right of* merge into one another

Combinations of external 3D and internal 2D computations can be used to obtain more detailed information about the location of an object. For example, in Figure 6 an appropriate answer to the query *"Where is the book located?"* would be *"The book is located on the right side of the desk"* which is a combination of the 3D relation *on* and the 2D relation *right*. The computation of the latter requires an abstraction from the original 3D shape of the desk and to use the derived 2D projection.

Cardinal relations can be used as well inside and outside an appropriate reference object, e.g. *in the south* vs. *south*. Although the actual extension of the reference object defines the distiction between internal and external cardinal relations, it doesn't influence the evaluation of the angle of deviation. Therefore the method for determining the degree of applicability of external relations can be used in exactly the same way for the internal counterpart. There is only one difference, because the class of GEO-III relations are usually not applied inside reference objects and are therefore excluded for

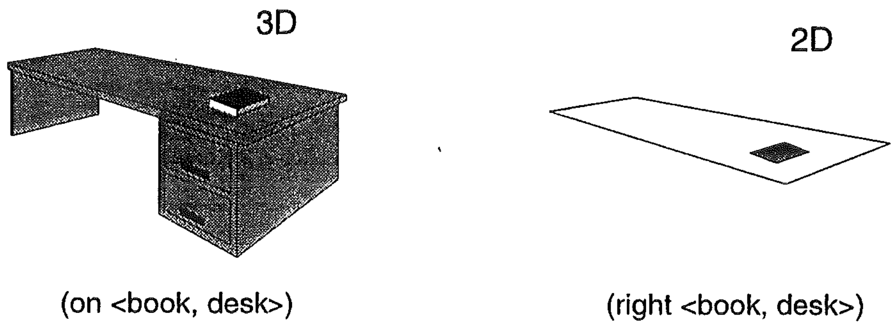


Figure 6: Combined computation in 2D and 3D space achieves more detailed information

internal use.

In total the CSR-3D system can now account for more than 70 spatial relations which can be directly linked to linguistic entities, such as spatial prepositions or adverbs.

3 Preprocessing

The process of object localization requires the focussing on potential reference objects proximate to the object to be localized. The size of the visual focus depends mainly on the extension of the reference object. Potential reference object candidates are required to be of a certain minimal size, which is again determined by the extension of the reference object. The number of such candidates is situation dependent, and can therefore vary from one to a considerable number of objects. The selection of the best available reference object in the focussed range requires the consideration of all applicable spatial relations to finally decide which object to choose. Figure 7 gives an example of why this might be necessary. In the figure the two reference objects are equal in size and visual salience, and both are the same distance from the located object. However, one would choose RO_1 as the appropriate reference object to describe the LO's position. This is because the spatial relation *below* is simpler and therefore easier to establish than the composite relation *to the right of and below.*[3]

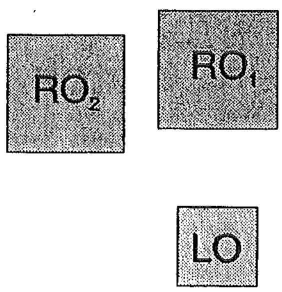


Figure 7: The spatial relationship between the LO and potential ROs can be a decisive criterion for the selection of the best reference object

Establishing all the applicable spatial relations of potential reference objects in advance is computationally intensive if one considers the large amount of possible relations to be tested. However, the set of relations which are of relevance within a certain context can be restricted in two respects: the type of the reference object and the type of the localization.

[3] Note, that it is necessary in the example for both relations to be of the same degree of applicability.

1.	2.	3.
distance		
projective	cardinal	cardinal projective

Table 1: Three categories of applicable relations depending on the type of the reference object

3.1 Type of the Reference Object

Dependent on the type of the potential reference object are the types of applicable spatial relations. The location of objects in *large-scale spaces* [Acredolo 81, Ittelson 73] for example are mainly described by using cardinal relations like *"Europe is north of Africa"*, whereas in *small-scale spaces* terms of projective relations are more appropriate: *"The car stands in front of the house"*. For objects in *medium-scale spaces* [Mandler 83], e.g. buildings, both projective and cardinal relations are permissible in certain situations. If the answer is generated using a deictic or intrinsic frame of reference, projective relations are used. However, it is nevertheless possible to use a cardinal relation, as in: *"The church is north-west of the City Hall"*. The categorization consists, therefore, of the three classes shown in Table 1: the first and the second class deny the use of cardinal respectively projective relations and the third class permits all subclasses.

Using this knowledge the number of spatial relations to be tested can be reduced before their applicability is computed.

3.2 Type of the Localization

The type of the localization depends on the context in which the question is raised. Two types can be observed: approximate and precise localizations. For example, in the question: *"Where is the post office?"* people at first search for approximate localizations using distance relations like: *"The post office is near the City Hall"*. If this is not sufficient exact for the addressee, projective or cardinal relations are used for a more precise localization. There are two reasons for this behavior: First, distance relations are easier to apply, e.g. no frame of reference has to be established, and second, the approximate localization of a salient reference object is used to determine if this object is known to the addressee. If the object is known then a more detailed description can be added, if necessary.

The class of relations used for approximate localizations are distance, like *at, near, in*, etc. Relations which contain orientational information, like projective or cardinal relations, are more precise. However, in an answer to a spatial query about the location of a certain object using the latter relations, the information of proximity is nevertheless included, because a chosen reference object is required to be close to the located object.

In summary, the type of a particular object localization can be classified in an approximate or a precise localization in advance, which yields a further restriction of the possible set of applicable spatial relations.

4 Evaluation of the Restrictions

In this section we will evaluate how preprocessing affects the performance of the computation of spatial relations in object localization tasks. The original German relations accounted for by the CSR-3D system are listed in Table 2.

Hence the number of spatial relations for each group in the CSR-3D system is as follows:

class	external	internal
distance	*an*, *bei*, *nahe*, *fern*, *außerhalb*	*zentral*, *inmitten*, *in*, *innerhalb*
projectiv	*links*, *rechts*, *vor*, *hinter*, *über*, *unter*, *oberhalb*, *unterhalb*	*links*, *rechts*, *vorne*, *hinten*, *oben*, *unten*
cardinal	*Nord*, *Ost*, *Süd*, *West*, *Nordost*, *Südost*, *Südwest*, *Nordwest*. *Nordnordost*, *Ostnordost*, *Ostsüdost*, *Südsüdost*, *Südsüdwest*, *Westsüdwest*, *Westnordwest*, *Nordnordwest*	*im Norden*, *im Osten*, *im Süden*, *im Westen*, *im Nordosten*, *im Südosten*, *im Südwesten*, *im Nordwesten*
projectiv & projectiv	*links-vor*, *links-hinter*, *links-über*, *links-unter*, *rechts-vor*, *rechts-hinter*, *rechts-über*, *rechts-unter*, *rechts-oberhalb*, *rechts-unterhalb*, *links-oberhalb*, *links-unterhalb*	*links-vorne*, *links-hinten*, *links-oben*, *links-unten*, *rechts-vorne*, *rechts-hinten*, *rechts-oben*, *rechts-unten*
projectiv & distance	*links-an*, *links-bei*, *links-nahe*, *rechts-an*, *rechts-bei*, *rechts-nahe*	
other	*neben*, *zwischen*, *auf*	

Table 2: Classification of the German spatial relations used in CSR-3D

- 8 distance relations
- 15 (12 + 3) projective relations, incl. the three special relations *beside*, *between* and *on*
- 24 (16 + 8) cardinal relations
- 26 composite relations

Therefore, without any restrictions, 73 spatial relations have to be tested in a general localization task. Depending on the context, in a localization either one, two, or none of the preprocessing factors discussed above can be used to decrease the number of relations to be tested. The following examples describe each of the six possible situations.

```
USER:1> Where is the phone book?
SYSTEM:1> It is on the desk.

USER:2> Where exactly is the phone book?
SYSTEM:2> It is on the left side of the desk.

USER:3> Where is Bavaria located?
SYSTEM:3> Bavaria is located in Germany.

USER:4> Where exactly is Bavaria located?
SYSTEM:4> Bavaria is located in the south-east of Germany.

USER:5> Where is the city hall located?
SYSTEM:5> The city hall is located behind the church.

USER:6> Where is the city hall located?
SYSTEM:6> The city hall is located south of the church.
```

case	total no pp	type of RO	type of local.	total with pp	%
1	73	49 (8 + 15 + 26)	8	8	11%
2	73	49 (8 + 15 + 26)	65 (15 + 24 + 26)	31	42%
3	73	32 (8 + 24)	8	8	11%
4	73	32 (8 + 24)	65 (15 + 24 + 26)	24	33%
5	73	73 (8 + 15 + 24 + 26)	8	8	11%
6	73	73 (8 + 15 + 24 + 26)	65 (15 + 24 + 26)	65	89%

Table 3: Relations to test in the examples with and without preprocessing

In Table 3 the number of spatial relations to be tested are listed with respect to no preprocessing (pp), to the type of RO preprocessing, to the type of localization preprocessing, and to total preprocessing. The example shows a decrease to 89% of the original set of relations in the worst case, a decrease to 11% in the best case, and an average improvement down to 33% which is less than a third of the original number of relations.

5 Conclusion

After identifying the object of interest, the task of object localization can be divided up into two major parts: the selection of an adequate reference object and the processing of spatial relations between the objects. The number of potential reference object candidates can be large, as can the number of spatial relations to be tested. After extending the computational model for spatial relations, presented in [Gapp 94a, Gapp 94b], to internal projective and cardinal relations, it was shown that the set of relations to be tested can be reduced in two respects: the type of the reference object and the type of the localization. The reduction yields a minimal set of relations to be tested and therefore increases the overall performance of object localization processing substantially.

References

[Abella & Kender 93] A. **Abella** and J. R. **Kender**. *Qualitatively Describing Objects Using Spatial Prepositions.* In: Proc. of AAAI-93, pp. 536–540, Washington, DC, 1993.

[Acredolo 81] L. P. **Acredolo**. *Small- and large-scale spatial concepts in infancy and childhood.* In: L.S. Liben, A.H. Patterson, and N. Newcombe (eds.), Spatial representation and behaviour across the life span, pp. 63–81. New York: Academic Press, 1981.

[André et al. 88] E. **André**, G. **Herzog**, and T. **Rist**. *On the Simultaneous Interpretation of Real World Image Sequences and their Natural Language Description: The System SOCCER.* In: Proc. of the 8 th ECAI, pp. 449-454, Munich, 1988.

[André et al. 89] E. **André**, G. **Herzog**, and T. **Rist**. *Natural Language Access to Visual Data: Dealing with Space and Movement.* In: F. Nef and M. Borillo (eds.), Proc. of the 1^{st} Workshop of Logical Semantics of Time, Space and Movement in Natural Language. Hermès, 1989.

[Egenhofer & Herring 90] M. J. **Egenhofer** and J. R. **Herring**. *A Mathematical Framework for the Definition of Topological Relationships.* In: K. Brassel and H. Kishimoto (eds.), Proc. of the 4th International Symposium on Spatial Data Handling, pp. 803–813. Zürich, 1990.

[Gapp 94a] K.-P. **Gapp**. *Basic Meanings of Spatial Relations: Computation and Evaluation in 3D Space.* In: Proc. of AAAI-94, pp. 1393–1398, Seattle, WA, 1994.

[Gapp 94b] K.-P. **Gapp**. *A Computational Model of the Basic Meanings of Graded Composite Spatial Relations in 3D Space.* In: M. Molenaar and S. De Hoop (eds.), Proc. of the AGDM'94 Workshop, pp. 66–79, Delft, The Netherlands, 1994. Publications on Geodesy, Netherlands Geodetic Commission.

[Gapp 94c] K.-P. **Gapp**. *From Vision to Language: A Cognitive Approach to the Computation of Spatial Relations in 3D Space.* In: Proc. of the 1 st European Conf. on Cognitive Science in Industry, pp. 339–358, Luxembourg, 1994.

[Gapp 95a] K.-P. **Gapp**. *Angle, Distance, Shape, and their Relationship to Projective Relations.* In: J. D. Moore and J. F. Lehman (eds.), Proc. of the 17 th Annual Conference of the Cognitive Science Society, pp. 112–117, Mahwah, NJ, 1995. Cognitive Science Society, Lawrence Erlbaum Associates, Inc.

[Gapp 95b] K.-P. **Gapp**. *An Empirically Validated Model for Computing Spatial Relations.* In: I. Wachsmuth, C.-R. Rollinger, and W. Brauer (eds.), KI-95: Advances in Artificial Intelligence. Proceedings of the 19th Annual German Conference on Artificial Intelligence, pp. 245–256, Berlin, Heidelberg, 1995. Springer.

[Gapp 95c] K.-P. **Gapp**. *Object Localization: Selection of Optimal Reference Objects.* In: A. U. Frank and W. Kuhn (eds.), Spatial Information Theory: A Theoretical Basis for GIS. Proceedings. Lecture Notes in Computer Science, **988**, pp. 519–536, Berlin, 1995. Springer.

[Herskovits 86] A. **Herskovits**. *Language and Spatial Cognition. An Interdisciplinary Study of the Prepositions in English.* Cambridge, London: Cambridge University Press, 1986.

[Ittelson 73] W. H. **Ittelson**. *Environment perception and contemporary perceptual theory.* Environment and cognition. Seminar, pp. 1–19, 1973.

[Koller 92] D. **Koller**. *Detektion, Verfolgung und Klassifikation bewegter Objekte in monokularen Bildfolgen am Beispiel von Straßenverkehrsszenen.* St. Augustin: Infix, 1992.

[Mandler 83] J. M. **Mandler**. *Representation.* In: P. Mussen (ed.), Handbook of child psychology, chapter Vol. III, pp. 420–494. New York: John Wiley & Sons, 1983.

[Mukerjee & Joe 90] A. **Mukerjee** and G. **Joe**. *A Qualitative Model for Space.* In: Proc. of AAAI-90, pp. 721–727, Boston, MA, 1990.

[Olivier et al. 94] P. **Olivier**, T. **Maeda**, and J. **Tsuji**. *Automatic Depiction of Spatial Descriptions.* In: Proc. of AAAI-94, pp. 1405–1410, Seattle, WA, 1994.

[Rajagopalan 94] R. **Rajagopalan**. *A Model for Integrated Qualitative Spatial and Dynamic Reasoning about Physical Systems.* In: Proc. of AAAI-94, pp. 1411–1417, Seattle, WA, 1994.

[Schirra & Stopp 93] J. R. J. **Schirra** and Eva **Stopp**. *ANTLIMA – A Listener Model with Mental Images.* In: Proc. of the 13 th IJCAI, pp. 175–180, Chambery, France, 1993.

[Talmy 83] L. **Talmy**. *How Language Structures Space.* In: H. Pick and L. Acredolo (eds.), Spatial Orientation: Theory, Research and Application, pp. 225–282. New York, London: Plenum, 1983.

[Wazinski 92] P. **Wazinski**. *Generating Spatial Descriptions for Cross-modal References.* In: 3rd Conference on Applied Natural Language Processing, Trento, Italy, pp. 56–63. Association for Computer Linguistics, 1992.

Some Issues in Generating Preverbal Messages: the Case of Spatial Descriptions

Daniel Glatz, Ralf Meyer-Klabunde, Robert Porzel*
Centre for Computational Linguistics, University of Heidelberg
glatz@jerry.iued.uni-heidelberg.de
meyer@novell1.gs.uni-heidelberg.de
porzel@jerry.iued.uni-heidelberg.de

Abstract

In order to be communicatively successful while describing a spatial scenario, a speaker should satisfy several constraints that are related to certain properties of the make-up of the spatial representation, relevant properties of the partner contained in the speaker's partner model, and properties of the communicative situation. Based on empirical studies on structures of spatial representations and strategies for their verbalization, we present a computational model of the conceptual processes underlying the production of spatial descriptions.

1 Preverbal Messages

The term 'preverbal message' is taken from Levelt's (1989) language production model and denotes the propositional conceptual structure that shall be verbalized by means of grammatical and phonological processes. Generally, preverbal messages as understood by Levelt correspond to other theoretical entities as e.g. the *encoding input* in the model of Herrmann & Grabowski (1994), with the *message* in Bock (1995) or simply the *conceptual structure* in De Smedt & Kempen (1987).

Although the respective expressibility of individual languages imposes their own requirements on preverbal messages, they are assumed to be language-independent representations. Levelt's (1989:103ff.) example for taking language-specific requirements into account in the preverbal message is the deictic system of English and Spanish: whereas English uses a bipartite deictic system that distingushes between a proximal and a distal distance to the speaker (*here* – *there*), Spanish speakers have access to a tripartite system that distinguishes between proximal, medial and distal distances. It is encoded in

*This research is being supported by the German Science Foundation (DFG) in its Special Collaborative Research Program 'Language and Situation' (SFB 245), project A7 ('Implicit Information').

the lexicon by the adverbials (*aquí - ahí -allí*). Such language-specific requirements must be encoded in the preverbal message, even if they don't apply for the current situation. Otherwise the appropriate lexical item cannot be retrieved in the formulator.

To arrive at a preverbal message that provides the information that is instrumental in realizing the speaker's goal in describing a spatial scenario, the speakers have to prepare, select, and linearize various kinds of conceptual information, viz. the spatial representation, the discourse protocol which keeps track of what has been said by whom, general conceptual knowledge, and a model of the current addressee.

Most often speaking is a means for communicating with somebody else and what speakers want to say and how they say it is strongly influenced by the speaker's assumptions of the informational demand of the addressee, his skills, social status and many other features. In other words: the information the speaker wants to convey is tailored to the speaker's internal model of the actual addressee.[1] The speaker's estimation of the informational demand of the addressee and his cognitive and linguistic skills influence the selection, preparation, and linearization of information. An obvious example would be talking to a child vs. an adult. However, this tailoring of the message to the addressee is not only reflected in conceptual processes; rather it will be handed down to the verbalized message by means of lexical, grammatical and phonological structures and features and is reflected in the *information structure* of the utterance.

Although it is reasonable to assume a close relation between message structure and linguistic form, most psycholinguists and (computational) linguists put their emphasis on the structure of and processes for grammatical and phonological encoding. Descriptions of the conceptualizer are rather sketchy because a theory of message structure is not in the offing. Analogously, a detailed theory of message generation or conceptualizing has not been presented as a whole (see Levelt (1989:70ff.) on this topic and Ward (1989)). One psycholinguistic model that stresses the relevance of conceptual processes for the linguistic form is the model of Herrmann & Grabowski (1994). Herrmann & Grabowski emphasize the situation-relatedness and, therefore, the addressee-relatedness of speaking, whereby the actual addressee is regarded as a component of the current situation the speaker perceives and must cope with.

In order to outline our assumptions concerning the relation between preverbal message and linguistic form, let us have a look at the following example:

(1) Ihr Auto steht noch bei der KNEIPE.
$\text{your}_{politeform}$ car stands still at the PUB.
$\text{Your}_{politeform}$ car is still standing at the PUB.

The capitals represent the accented focus exponent. It is prosodically prominent. Let us

[1]The relationships between the speaker's model of her listener/addressee and the speaker's own beliefs are quite complex, as Dell & Brown (1991) point out. Speakers do not have one model for specific addressees, but they possess representations of various generic partner models with models of current addressees being instances of them. A very interesting question is how speakers construct generic and specific partner models by means of beliefs, communication and social interaction, but we will not touch this field.

look at the details of this sentence and the conceptual conditions that must be met while planning this sentence.

First, the subject NP is a possessive phrase without an explicit relation between holder and property. The preverbal message must provide sufficient information for generating the phrase *Ihr Auto* and not a phrase that explicates the relation between addressee and the car, as e.g. *Das Auto, das Ihnen gehört/ das Sie gefahren haben/ etc.* ('the car that belongs to you/ you drove'/ etc.) or *Das Auto von Ihnen* ('the car of you'). It is therefore reasonable to assume that the preverbal messages underlying possessive phrases contain a slot for the relation between owner and property, but that relation does not necessarily have to be explicated. If the relation has not been determined, the possessive pronoun will be shifted into the determiner position. Furthermore, the phrase *Ihr Auto* indicates that the speaker has a certain car in mind and that she is certain that the addressee is able to identify that car as well: this phrase can only be used with *Auto* being definite.[2] The other constructions would also allow an indefinite NP, as in *Ein Auto, das Ihnen gehört.* The retrieval of the politeness form *Ihr* instead of *dein* finally requires information concerning the social distance between speaker and addressee.

Second, the prepositional phrase *bei der Kneipe* ('at the pub') should only be uttered if the speaker beliefs that the addressee is able to identify the intended pub. The speaker should also be certain that the location of the car in the proximal region of the pub by means of the vague preposition *bei* ('at') is sufficiently specific for the addressee.

Third, the preverbal message must contain information concerning time, modality, and the static localization to be expressed. The use of the particle 'noch' (roughly: 'still') presupposes also that the speaker believes of an early change of the expressed state.[3]

Fourth, the preverbal message must additionally constrain the syntactic and phonological realization. If the message would only contain information on the localization of the specific car with respect to the proximal region around the specific pub, this information could also be expressed as in:

(2) Ihr Auto, das steht noch bei der KNEIPE.

(3) Was Ihr Auto betrifft, das steht noch bei der KNEIPE.

(4) Ihr Auto steht NOCH bei der Kneipe.

(5) Ihr Auto, das steht NOCH bei der Kneipe.

(6) Was Ihr Auto betrifft, das steht NOCH bei der Kneipe.

and so on. These utterances differ from the first one only in the syntactic construction used for expressing the same conceptual relation between the car and the pub, and/or the place of the accent. In other words: these sentences differ in their information structure, i.e. their topic/comment and focus/background structure. Of course, these sentences convey different information because they aren't mutually interchangeable in the contexts of their appearance.

---

[2]Or generic, as in *Ihr Auto sollte ein Lenkrad besitzen.* ('Your car should have a steering wheel')

[3]See de Mey (1994) for a comparison of different formal analyses of German *noch* and its Dutch pendant *nog.*

According to Jacobs (1984:46), the *topic* could be defined as that element which provides the frame for the interpretation of the additional sentential material, i.e. the comment. It serves as the anchor for the information provided by the sentence and, for being an 'anchor', the addressee must be able to allocate the topic element in his discourse model. For this reason, brand-new topic elements serve only very rarely as topics; they cannot be appropriately incorporated into the discourse model (see Lambrecht 1994:165ff.).

Probably for cognitive reasons the topic is preferably placed towards the beginning of a sentence,[4] because in order to update his discourse model efficiently while comprehending a new sentence, the addressee first should find the appropriate 'register' for incorporating the new information. Besides being anchorable, topic elements possess an additional function: the different syntactic constructions used indicate for the addressee also the status of the topic element. In example (1), the topic is realized as subject, in (2) by left dislocation, and in (3) as a so-called hanging topic. These three different syntactic constructions trigger different foregrounding processes with respect to the topic element, as the following example with hanging topic demonstrates:

(7) Peter küsst Maria. ??Was Peter betrifft, der hatte das schon lange vor.

The second sentence is odd because the discourse referent for Peter has been introduced in the first sentence and is still in the foreground. The use of a hanging topic, however, makes only sense if there is an already introduced discourse referent that was shifted into the background and shall be foregrounded again.[5]

Next to the *focus*, the focussed elements carry information that is assumed to be 'new' or 'relevant'[6] for the addressee. The speaker's problem with respect to the focus is to put the focus accent onto the appropriate element. The speaker assumes which part of a sentence is the focused one and which belongs to the background, but the actual problem of identifying the whole focused phrase is a task for the listener.

The brief characterization of the bipartite information structure indicates that the propositional content of the exemplary sentence is also organized with respect to the speaker's model of the addressee's discourse model. Foregrounding a certain discourse referent makes only sense if the speaker believes that the addressee has his attention on a different referent and his attention should be directed to that older discourse referent again.

Where does the decision to generate a sentence with the first NP realized as subject and accent on the last word come from? In our opinion this decision is already encoded in the preverbal message. The discourse referent for the specific car must be coded as topic candidate and, additionally, its specific linkage with the addressee's discourse model[7] is marked as well. The referent for the intended pub must be marked as focus exponent.

[4]From a typological point of view this statement might be too strong. See e.g. Tomlin & Rhodes (1992).

[5]Of course, the notions of 'salience', 'shifting to the background', and 'foregrounding' are still pre-theoretic, metaphoric ones and must be defined in precise terms. See Lambrecht (1994) for a first account.

[6]Also in a sense to be explicated.

[7]Of course, only as assumed by the speaker.

This information on topichood and focus constrains the syntactic and prosodic realization because at our disposal there are preferred linguistic realizations for certain topic– and focus–feature configurations. Example (1) with subject NP is the preferred realization for a highly salient topic candidate, whereas example (2) with a left dislocated topic is preferred in case of a weaker salient topic candidate.

To summarize, if the mentioned requirements for the underlying preverbal message would be modeled by means of a feature-value structure, the preverbal message underlying example (1) might be represented as in figure 1. In order to stress the non-linguistic nature of preverbal messages, we borrow some terminology from the knowledge representation community. Relationships between individuals will be called roles, terms to describe classes of individuals will be called concepts, and placeholders for objects that belong to the extensions of concepts will be called instances instead of discourse referents.[8]

```
[ role:            stand
                   [            [ concept:        car ]                              ]
                   [  main:[1]  [ instance:       x   ]                              ]
                   [            [ number:         one ]                              ]
                   [            [ identifiable:   yes ]                              ]
  agent:           [            [ role:           --                               ] ]
                   [            [                 [ concept:       listener ]      ] ]
                   [  specify:  [ possessor:      [ instance :     y        ]      ] ]
                   [            [                 [ social_dist:   distant  ]      ] ]
                   [            [ possessing:     [1]                              ] ]
                   [ concept:        pub      ]
                   [ instance:       u[2]     ]
  place:           [ number:         one      ]
                   [ identifiable:   yes      ]
                   [ loc:            proximal ]
  time:            continual
  modality:        fact
  topic:           [ candidate:  [1]        ]
                   [ status:     accessible ]
  focus_exponent:  [2]                                                                 ]
```

Figure 1: Preverbal message underlying German equivalent of 'Your car is still standing at the pub'

After this brief characterization of the conditions for verbalizing the single exemplary sentence the question might come up why we are so interested in the preverbal message only

[8] A further name could be 'reference object' ('refo') from the theory of referential networks described in Habel (1986).

and not in the entire production process. It seems as if we would pay our attention only to the first half of the whole story. We hope that the example given above demonstrates the relevance of the structure of preverbal messages for the linguistic realization. Preverbal messages do not only have to *meet* language-specific requirements, they also *determine*, at least partially, the linguistic form of the single clauses. The information structure of sentences is tracable back to the preverbal message.

Additionally, if we leave the level of sentence production and have a look at whole discourses, preverbal messages should also provide information that is mirrored in the overall verbalization strategy which has to cope with the entire utterance. Furthermore, preceeding clauses function as part of the background against which the discourse referents for the current clause are being weighted.

1.1 Preverbal Messages for Spatial Descriptions

Besides general information concerning the information structure, time, and modality, preverbal messages for spatial descriptions must provide spatial relations between discourse referents for objects or object parts and information concerning the overall strategy used for the description. An appropriate spatial relation depends on the point of view the speaker assumes for the conceptualization, and the appropriate point of view depends on certain estimated features of the addressee.

The overall strategy depends partially on the linearization of the involved objects and the linearization depends on the speaker's addressee-model again. Hence, the basic questions are: how is an appropriate spatial relation between objects determined that will be inserted into the preverbal message for the retrieval of appropriate lexical items? And how can conditions for the linearization strategy be derived from the preverbal message? Answers to both questions obviously provide also statements on addressee-relatedness.

A spatial description that allows the listener to construct an adequate mental model of this spatial object constellation depends on the one hand on general properties of the chosen linearization,[9] and on the other hand, on the speaker's estimation of cognitive features of the addressee, allowing the speaker to describe the spatial scenario in a way that is appropriate to the hearer's skills. The basic problems in spatial selection and linearization tasks are:

1. Which point of view will be chosen?

2. Which objects may serve as reference objects/ primary objects?

3. What is the very first reference object, i.e. the starting point of the overall localization?

4. Which linearization strategy will be employed?

[9]Some general, addressee-independent, economic principles guiding certain linearizations are presented in Levelt (1982).

5. What is the best linearization?

The decisions with respect to each of these problems determine the domain-specific components of the preverbal message for the overall decription. Spatial descriptions consist of sequences of localizations of some primary objects – usually exactly one – in relation to one or more reference objects, but descriptions of one and the same spatial object constellation differ considerably in their linguistic form. Possible sources of the variety of forms are the use of different viewpoints from which the objects have been localized, the different determinations of objects as reference or primary objects, and the overall structure that is employed for that determination and ordering.

To demonstrate the various influences of preverbal messages onto linguistic forms, let us assume that the following spatial scenario is given: a church, a bus stop, a town hall, a fire station, and a speaker and an addressee are given as in figure 2. Their intrinsic fronts have been marked by an arrow.

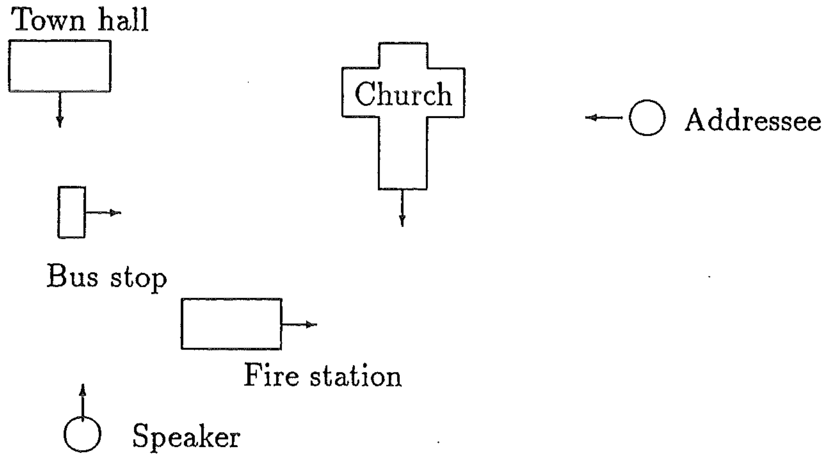


Figure 2: An exemplary spatial scenario

In order to describe this object constellation, the speaker might utter:

There is a church in front of you. Right of the church is the town hall. In front of the town hall is a bus stop. Diagonally to the right of the bus stop is a fire station.

Another correct description would be:

If you go straight, to your right you will see a fire station. If you continue going straight, you will arrive at a bus stop. Beyond the bus stop you will see the town hall. To your right, you will see a church.

The former description starts with the addressee as first reference object and uses solely intrinsic orientations for the determination of spatial relations. The overall structure of that description is characterized by a point-by-point strategy, and the underlying linearization is: ⟨addressee – church⟩, ⟨church – town hall⟩, ⟨town hall – bus stop⟩, ⟨bus stop – fire station⟩.

The second description is given as an imaginary tour. The fictive observer ('you' does not refer to the hearer) starts at the speaker's position and the spatial relations are determined by the wanderer's observer axis. The underlying linearization is: ⟨fictive observer – fire station⟩, ⟨fictive observer – bus stop⟩, ⟨fictive observer – town hall⟩, ⟨fictive observer – church⟩. Moreover, on the linguistic level the example demonstrates the frequent use of conditionals in imaginary tours.

Both descriptions are, inter alia, so fundamentally different because the decisions of a speaker for a certain perspective, linearization, and strategy are also results of her estimation of the addressee's skills to comprehend the corresponding description.

1.2 Addressee-relatedness of Spatial Descriptions: Some Empirical Results

In general, a speaker talking about spatial configurations is either able to see the objects at the same time, or she is imagining a certain object configuration, or she is talking about the mental representation of an object constellation which she had been obtained by spatial perception some time ago, and she must retrieve from memory for the production process (Buhl, 1995). We will confine ourselves to the last case. We leave out any questions concerning image-like representations and hybrid approaches (Kosslyn, 1994; Habel et al., 1995), and use propositional representations only, which, however, is not intended to imply the irrelevance of analogue representations. The main reason for leaving out analogue representations and reasoning processes upon them is simply that we don't need them. Modelling the pragmatic processes concerning the preparation and linearization of spatial representations does not necessarily comprise processes related to image-generation.

There are a number of empirical studies concerning the first question from the list of problems mentioned above, but the other problems have not attracted much attention yet, apart from some studies on the establishment of spatial relations between objects in 2D or 3D space (e.g. Gapp 1995). Based on our examination of the features of spatial descriptions in the light of village descriptions, we are going into these questions in more detail.

Point of View: Only if the point of view has been determined can a corresponding spatial relation between the respective reference and primary objects be determined. The default point of view chosen in conversations with present partners is the speaker–oriented one. Probably this is due to the fact that the mental representation of a spatial configuration does not provide viewpoint-free information so that the position of every object can be objectively determined in relation to any other object. It is rather bound to the point of view the speaker held during the acquisition of that representation. However, the viewpoint in the acquisition phase does not need to be the appropriate viewpoint in the communication situation where the speaker is transforming the representation into a description; a point of view difference might occur (Shober 1993). Additional points of view that might become relevant for the linearization might include the point of view of a fictive observer (when an imaginary tour is appropriate as linearization strategy) or the point of view of the listener (when it is appropriate to localize from the listener's point of view).

However, these different viewpoints are not of equal value with respect to their availability. Buhl (1995) shows in her empirical study the dominant status of the viewpoint from the acquisition phase. Subjects learned the localization of the buildings of a toy village by means of a movie showing a specific route through the village along two target objects, a church and a fountain. When the communicative task was to describe the route along the church and fountain from the opposite direction than that given in the movie, an overwhelming number of subjects produced an imaginary tour containing a localization of the fountain with respect to the church that was given from the viewpoint established during the acquisition phase. For this the subjects guided the fictive observer in such a way through the village that the point of view of the observer and their own point of view during the acquisition phase coincided, although this resulted in a longer tour through the village. Buhl concludes that the mental spatial representation also contains the point of view from the acquisition phase. This observation is in agreement with Pinker's (1988) assignment of a special status to the viewpoint during the acquisition of the spatial representation (among all possible points of view).

Although the point of view from the acquisition phase is the dominant one, under certain conditions there are at least two reasons for the speaker to rotate the spatial representation into the point of view of the addressee in order to reduce the burden of perspective-taking for him. In the first case, the speaker may not be sure that the addressee will be able to perform this mental rotation by himself due to his lack of cognitive competence. The second case is related to politeness. If the addressee occupies a higher social position than the speaker, the act of being polite can be enhanced by adopting the other's point of view and its explicit verbalization. Empirical studies with German and Japanese subjects[10] show the relevance of these factors for the production process (see Herrmann and Grabowski, 1994:125ff.; Shober, 1993:4). To summarize, the speaker's estimation of the addressee's skills and his social status affect the choice of an appropriate point of view.

We should finally mention that these studies are dealing with rotations of reference systems, viz. the rotation of the speaker's orientation into the hearer's orientation, not of objects as in Shepard and Metzler's (1971) famous model. The orientations of the objects kept constant. However, both kinds of rotation have in common that they cause cognitive load and, therefore, the reaction time – by and large – increases with a growing rotation angle.

[10]The comparative study was undertaken in Japan because Japanese society, roughly speaking, stands out as having a more distinct consciousness for social hierarchies. As a matter of fact, the comparative study was shown to clearly have a stronger effect of the social status of the addressee.

The Problems Concerning Reference Objects: Objects that serve as reference objects for primary objects must be identifiable for the addressee, either by perception, by inference, or by knowledge of the preceding discourse. This holds especially true for the starting point of a description, i.e. the very first reference object, which can also be the speaker or the listener. In other words: reference objects may be linguistically introduced via definite or indefinite noun phrases, but they always constitute that part of the information a proposition carries that serves as an anchor for the other information to be conveyed. It has, therefore, to be information the speaker believes to be part of the listener's domain or general knowledge, inferable as a result of the listener's knowledge, or accessible through other cognitive means.

If the reference objects serve as an anchor due to their identifiability, they are also used as the topic of that clause, because anchoring was the key notion in defining topics. Therefore a sentence as *Ein Haus befindet sich hinter der Kirche* ('a house is behind the church') with the primary object as topic is rather seldom. Instead first the house should explicitly be introduced as new referent and afterwards it should be located with respect to the church, as in: *Dann ist da noch ein Haus. Das befindet sich hinter der Kirche.*

The explicit naming of the speaker or listener as the first reference object can also indicate to the addressee whose point of view has been chosen, as in: *From your point of view,* .... Usually this will be done if the speaker cannot estimate whether the addressee is able to determine the point of view. However, Shober's (1993) data show that in conversations "most pairs never discussed explicitly whose spatial point of view they should use" (p. 13).

Other than being identifiable, a reference object must also be sufficiently salient for the primary object. The possible localization ??*The church behind the bicycle* would violate this constraint.

Linearization Strategies: Finally, the appropiate linearization strategy depends on the kind of first encounter with the spatial object constellations and the overall placement of the objects. A first encounter by means of a tour through an object constellation usually also results in a description by means of an imaginary tour. In general, a speaker can use three linearization strategies: a global one, a point-by-point strategy and the use of an imaginary tour (Carroll, 1993). Global views employ a strategy that superimposes a spatial structure, e.g., a coordinate system that encompasses the 'entity' under a description as a whole and the individual objects are localized in reference to that structure. A point-by-point strategy employs spatial structures that are anchored at the individual objects which make up the complete constellation being described, and the individual objects are localized in reference to the structures and spatial properties of the other objects. Another strategy can be established by means of an imaginary tour employing a fictive observer that is placed and moved within the scenario, and the individual objects are localized in reference to the fictive observer.

An appropriate linearization strategy should correspond to an appropriate linearization of the objects. For example, an imaginary tour should not be based on a linearization with one object that will be used at different times as a reference object, because this would imply a tour with redundant paths where the fictive observer has to return to the reference object several times. A suitable linearization strategy based on such a linearization is better provided by a point-by-point strategy.

Similar to Logan (1995), we assume that linearization is necessarily linked with the direction of attention. Logan views reference frames as mechanisms of attention. Adopted to our linearization problem, the question arises how attention is directed from one object to another one, from that one to the next one, and so on? Why does one object stand out between all other ones? In our computational model, this attention–shift has been modelled as the dynamic change of a salience value each object is annotated with (see below).

The 'Best' Linearization: Although a spatial object configuration can be linearized in many ways, we assume that there are appropriate and inappropriate ones. Appropriateness, however, is not a binary notion because there are gradiations between being communicatively successful and being unsuccessful. In optimality-theoretic terms (Prince and Smolensky, 1993), we could argue for the existence of a 'best' description and additional, still good, ones for one and the same situation. The best spatial description is one that is easy to follow for the actual addressee and is, therefore, tailored appropriately to that addressee's capabilities.[11]

A linearization strategy makes sense only if it matches with the linearized sequence of objects that has been tailored to the addressee's capabilities. The different constraints hold for different candidates.[12] For example, the identifiability-constraint is applicable to object referents only; salience, however, is based on a relation between two referents.

1.3 Handling Aspects of Speaker Intention

The previous description touches the domain–specific problems in the constructions of preverbal messages for spatial descriptions. But how can we handle the appropriate temporal and modal specification of the entire discourse and the construction of the preverbal message within this framework?

Certain 'pre–adjustments' for the entire discourse can be explained if the discourse would be taken as an answer to an implicitly or explicitly given question. This is the basic idea of the text–linguistic quaestio–approach (Klein & von Stutterheim, 1987; von Stutterheim, 1995). The question or quaestio mirrors aspects of the informational demand of the addressee and, therefore, it might be considered as part of the speaker's intention.

A quaestio establishes a number of constraints concerning time, modality, the set of possible topic referents, and focused elements. For example, a quaestio verbalized as *Where are the objects?*, requires present tense and factive modality. Additionally, the set of objects in the domain provide the set of possible topic referents and the spatial relations

[11]The link to optimality theory has only been given because we are using a constraint-based approach to the generation of preverbal messages and constrained-ranking is the key notion of optimality theory.

[12]This is the reason why we cannot present a simple constraint tableau as they are known from optimality-theoretic papers.

are focussed and moved to the foreground (It is the spatial relations that are asked for). The preverbal message underlying spatial descriptions must also meet these more general constraints given by the quaestio.

2 Implementation: the System PAROLE

The aforementioned theoretical models and empirical studies provide the basis for a computational model of the generation of spatial descriptions. We are going to explain our generation system PAROLE[13] by means of the artifical village from figure 2. The speaker (i.e. the system) has a certain view of the village and knows the listener's position and estimates his cognitive competence. In case of low competence and a present listener, the speaker may rotate his spatial view to the view from the listener's position, but the egocentric perspective functions as a 'brake' for possible rotations.

2.1 The Spatial Representation

The spatial representation resembles in some aspects the model proposed in Hernández (1994). Topologically, our model is confined to disjoint objects; overlapping or tangent objects have been left out. Furthermore, the objects are treated as having no extensions; they are points. Information concerning shape and size is only implicitly represented, viz. by means of a basic salience value that represents the salience of the objects with respect to all other objects. The basic salience is determined by the object's function, its size, shape etc., but how such a basic value is determined by perceptual and other mechanisms does not concern us here.

The spatial representation consists of two levels: one level states the set of possible places objects can occupy by means of a directed acyclic graph and the second level represents the actual objects that have been positioned on the nodes of the former level, together with a possibly intrinsic orientation if they possess intrinsic fronts and sides. The distinction between possible places and actual object places enables us to perform a rotation from one point of view into another quite easily.

Furthermore, the speaker/system position is always included in the spatial representation. The speaker is put onto a node with her intrinsic orientation and, by that, the orientation of the objects to each other with respect to the point of view of the speaker can be computed.

If an object gets a certain position, its surrounding region is separated into eight sectors by means of eight axes. The eight axes are given in figure 3.

The idea is that every axis determines a region for relating another object to the currrent object, but the axes are not of equal value. The front, back, right, and left axes get priority over the left-front/right-front and left-back/right-back axes because they match with the corresponding spatial prepositions. The hypothesis is that speakers tend to locate as simple as possible. Hence, if one object can be related to another one by means of the spatial

[13] PAROLE stands for 'partner-oriented linguistic explication'.

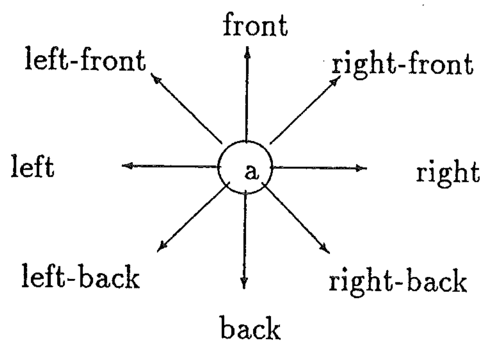


Figure 3: Eight Axes Structuring the Surrounding Region of an Object

'right' relation and there is another object, standing in a left–back relation to the same object, the object with simpler relation will be preferred in the linearization.

2.2 Linearization and the Determination of Point of View and Strategy

The empirical studies suggest to localize with egocentric point of view as long as possible. Only if the speaker estimates the addressee's competence as rather low, a rotation into the addressee's point of view might occur. The intended linearization depends on the one hand on the point of view and on the other hand on the salience of the objects and the simplicity of the relations between them. Finally, the linearization strategy (whether a point-by-point description will be given or an imaginary tour) depends on the linearization. An imaginary tour should not be used as strategy if one object should several times be used as reference object because the tour would in this case require several redundant paths. A point by point strategy should not be used when the resulting spatial relations are too complex for the localizations.

2.2.1 Finding the Optimal Linearization

This is how the optimal linearization is determined:

- Depending on the estimated listener's competence, either the speaker or the listener or a salient object within the spatial representation are determined as the very first reference object. Which one will be taken depends on whether the listener is also represented in the spatial representation and secondly, on the estimated cognitive competence and social position of the listener. For example, in case of a listener with estimated low social psition and low competence, the listener becomes the first reference object and an intrinsic point of view will be used. If the listener's social position is low and he possesses a high competence, a speaker-oriented extrinsic linearization will be employed, and so on.

- All reference objects must be identifiable for the addressee, either by perception, by inference, or by knowledge of the preceding discourse. In particular, objects that have been introduced as primary objects in the previously built-up discourse protocol may be used as reference objects.
- For every reference object, only the directly neighboured objects may become primary objects. Since the region around each object is divided into eight sections by means of the four primary and four secondary axes (left, right, front, back and left–front, right–front, left–back, right–back, respectively), maximally eight possible primary objects can exist. Objects standing in one of the four primary relations to the primary object will be preferred.
- Starting with the first reference object, i.e. the speaker, the listener, or a salient object, the 'best' linearization is being determined by the computation of a sequence of ⟨reference object, primary object⟩ pairs. For that, five basic factors are considered, viz.
  - the basic salience value of each object, i.e. how an object stands out between all other ones,
  - the quality of the respective neighbored objects with respect to their basic salience compared with the basic salience of all other neighbored objects,
  - the kind of relation (primary or secondary ones) between the reference object and the other objects ,
  - the possibly redundant use of reference objects for different primary objects
  - the estimation of the next possible localization; i.e. how the system can continue from the chosen localization unit.

  The movement from a reference to a primary object implies a direction of attention and this attention shift is intented to be modelled by means of these factors.

2.3 From Preverbal Messages to the Description

Currently, the verbalization of the preverbal message is pie in the sky, but nevertheless we can give an outline of our basic idea. A lemma selector maps the spatial relations onto prepositions and the object names onto nouns, whereby the identifiability feature triggers the appropriate determiner. Since no verb exists, a copula will be retrieved in order to generate a complete sentence. By means of the argument structure of the preposition and the copula, and the information–structural features, the syntactic structures of the single sentences are constructed.

2.4 An Example

Let us take the artificial village in figure 2 as an example. Suppose the listener's competence and social position are both estimated as low so that an addressee-oriented linearization shall be employed. There are only two neighbored objects, viz. the church and the fire station, but the church is closer to the listener than the fire station. Additionally the fire station stands in a secondary relation to the listener, and the church possesses a higher basic salience than the fire station (due to its size) so that the church becomes the better primary object.

Now that the church has been introduced, both the listener and the church may be used as reference objects for other primary objects. If the listener would still be the reference object, only the fire station could become the next primary object, but it would violate the constraint of being in a simple relation. The church's neighbored objects are the town hall, the bus stop, the fire station, and the addressee. The last one has already been used as reference object and will therefore be deleted from the list of potential primary objects. The town hall will be evaluated as best candidate for the primary object because it stands in a primary relation to the church – as opposed to the fire station and the bus stop – and it possesses among all other neighbored objects the highest basic salience.

Now the listener, the church, and the town hall habe been mentioned and may be used as reference objects for further primary objects. If there is an object among the set of neighbored objects with respect to the town hall that can be evaluated higher than the remaining neighboring objects with respect to the church (i.e. the bus stop and the fire station), the town hall will become the next reference object. Otherwise it will be the church again. Directing the attention from the church to the bus stop or to the fire station would violate the simplicity-constraint, but using the bus stop as primary object with respect to the town hall does not violate any constraint. Therefore it is viewed as the 'better' linearization unit.

To shorten the whole process, going from the bus stop to the fire station is better than going again from the town hall or the church to the fire station due to the resulting redundancy. Hence, the final sequence of ⟨reference object – primary object⟩ pairs is: ⟨⟨addressee – church⟩ ⟨church – town hall⟩ ⟨town hall – bus stop⟩ ⟨bus stop – fire station⟩⟩. This overall linearization is redundancy-free. Therefore an imaginary tour might be appropriate as linearization strategy. Thus, a possible verbalization for a listener with low competence would be:

> *In front of you is a church. If you go straight, you will see a town hall. If you go right, there is a bus stop. If you go diagonally to the right, there is a fire station.*

2.5 The Next Steps

The implementation so far covers only parts of our overall plan. First of all, the preverbal message must be refined, especially for the employments of imaginary tours and for the realization of the information structure by means of word order and prosody. Nevertheless, we hope it has becomes clear that a study in language production should start with prelinguistic processes in the construction of preverbal messages as those mentioned in this paper.

References

[1] Bock, K. (1995) Sentence Production: From Mind to Mouth. In: J.L. Miller & P.D. Eimas (eds.) *Speech, Language, and Communication* (Handbook of Perception and Cognition). San Diego: Academic Press; 181–216.

[2] Buhl, H.M. (1995) Erwerbssituation, mentale Repräsentation und sprachliche Lokalisationen – Blickpunktinformation als Bestandteil der Raumrepräsentation. Submitted.

[3] Carroll, M. (1993) Deictic and Intrinsic Orientation in Spatial Descriptions: A Comparison between English and Dutch. In: Altarriba, J. (ed.) *Cognition and Culture: A Cross-Cultural Approach to Psychology.* Amsterdam: Elsevier; 23–44.

[4] Dell, G.S. & Brown, P.M. (1991) Mechanisms for Listener–Adaptation in Language Production: Limiting the Role of the "Model of the Listener". In: D.J. Napoli & J.A. Kegl (eds.) *Bridges Between Psychology and Linguistics: A Swarthmore Festschrift for Lila Gleitman.* Hillsdale, N.J.: Lawrence Erlbaum; 105–129.

[5] de Mey, S. (1994) 'Al' en 'nog' als temporele operatoren. *TABU 24 (4). Bulletin voor Taalwetenschap*; 169–186.

[6] De Smedt, K. & Kempen, G. (1987) Incremental Sentence Production, Self-Correction and Coordination. In: G. Kempen (ed.) *Natural Language Generation. New Results in Artificial Intelligence, Psychology and Linguistics.* Amsterdam: Martinus Nijhoff; 365–376.

[7] Gapp, K.-P. (1995) An Empirically Validated Model for Computing Spatial Relations. In: I. Wachsmuth et al. (eds.) *KI-95: Advances in Artificial Intelligence.* Berlin: Springer; 245–256.

[8] Habel, C. (1986) *Prinzipien der Referentialität.* Berlin: Springer.

[9] Habel, C.; Pribbenow, S. & Simmons, G. (1995) Partonomies and Depictions: A Hybrid Approach. In: J. Glasgow; H. Narayanan & B. Chandrasekaran (eds.) *Diagrammatic Reasoning. Computational and Cognitive Perspectives.* Cambridge, Ma.: AAAI/MIT Press.

[10] Hernández, D. (1994) *Qualitative Representation of Spatial Knowledge.* Berlin: Springer.

[11] Herrmann, T. & Grabowski, J. (1994) *Sprechen. Psychologie der Sprachproduktion.* Heidelberg: Spektrum.

[12] Jacobs, J. (1984) Funktionale Satzperspektive und Illokutionssemantik. *Linguistische Berichte, 91*; 25–58.

[13] Klein, W. & Stutterheim, Ch. von (1987) Quaestio und referentielle Bewegung in Erzählungen. *Linguistische Berichte, 109*; 163–183.

[14] Kosslyn, S.M. (1994) *Image and Brain. The Resolution of the Imagery Debate.* Cambridge, Ma.: MIT Press.

[15] Lambrecht, K. (1994) *Information Structure and Sentence Form.* Cambridge: Cambridge University Press.

[16] Levelt, W.J.M. (1982) Linearization in describing spatial networks. In: S. Peters & E. Saarinen (eds.) *Processes, Beliefs and Questions.* Dordrecht: Reidel; 199–220.

[17] Levelt, W.J.M. (1989) *Speaking. From Intention to Articulation.* Cambridge, Ma.: MIT Press.

[18] Logan, G.D. (1995) Linguistic and Conceptual Control of Visual Spatial Attention. *Cognitive Psychology 28*; 103–174.

[19] Prince, A.; and Smolensky, P. (1993) *Optimality Theory. Constraint Interaction in Generative Grammar.* Technical Report 2, Rutgers University Center for Cognitive Science, Rutgers University.

[20] Shepard, R.N. and Metzler, J. (1971) Mental rotation of three-dimensional objects. *Science, 171*; 701–703.

[21] Shober, M.F. (1993) Spatial perspective-taking in conversation. *Cognition, 47*; 1–24.

[22] Stutterheim, Ch. von (1995) *Prinzipien des Textaufbaus.* Habilitationsschrift, Universität Heidelberg.

[23] Tomlin, R.S. & Rhodes, R. (1992) Information distribution in Ojibwa. In: D.L. Payne (ed.) *Pragmatics of Word Order Flexibility.* Amsterdam: Benjamins; 117–135.

[24] Ward, N. (1989) Capturing Intuitions About Human Language Production. *Proceedings of the 11th Annual Conference of the Cognitive Science Society.* Hillsdale, N.J.: Lawrence Erlbaum; 956–963.

Räumliche Repräsentationsformate für die Sprachproduktion – eine Fallstudie*

Christopher Habel
Universität Hamburg
Fachbereich Informatik & Graduiertenkolleg Kognitionswissenschaft
Vogt-Kölln-Str. 30, D-22527 Hamburg
habel @informatik.uni-hamburg.de

1 Sprachverstehen, Sprachproduktion und interne Repräsentationen

Sprachverstehen und Sprachproduktion, die beiden zentralen Prozesse der Sprachverarbeitung, können aus kognitionswissenschaftlicher Perspektive als Prozesse angesehen werden, bei denen eine Zuordnung zwischen sprachlichen Entitäten, etwa Sätzen, und internen Repräsentationen hergestellt wird. Diese Repräsentationen, die zum Konzept der Bedeutungen – wie es in der theoretischen Linguistik und Sprachphilosophie verwendet wird – korrespondieren, werden von der Psycholinguistik auch als *mentale Repräsentationen* angesehen. Theoriespezifische Ausprägungen sind etwa „propositionale Repräsentationen" (McKoon und Ratcliff, 1992), „mentale Modelle" (Johnson-Laird, 1983; Glenberg, Meyer & Lindem, 1987) oder „Situationsmodelle" (Perrig & Kintsch, 1985; Morrow, Greenspan & Bower, 1987); auch in der Künstlichen Intelligenz und Computerlinguistik findet sich eine entsprechende Sichtweise: Verstehen wird meist als die Berechnung einer semantischen Repräsentation (häufig im Format logischer Repräsentationssysteme) und Generieren als den Aufbau sprachlicher Ausdrücke aus Wissens- und Diskursstrukturen (vgl. Palmer et al., 1993; Hovy 1993).

Was Textverstehen ausmacht, ist somit abhängig davon, was als *mentale Repräsentation* angenommen wird. Unter der Perspektive der mentalen Modelle (und Situationsmodelle) wird davon ausgegangen, dass nicht nur eine Repräsentation des Textinhalts aufgebaut wird, sondern insbesondere eine mentale Repräsentation eines Sachverhalts, der der Textinformation entspricht, gebildet wird (vgl. Abb. 1). Für derartige Repräsentationen wird

* Die im vorliegenden Aufsatz dargestellten Überlegungen zu räumlichen Repräsentationen, Gruppierungs- und Abgrenzungsprozessen sowie zu Prozessen der Sprachproduktion und des Sprachverstehens sind durch die DFG (Projekte: Ha 1237/2, Ha 1237/3 und Ha 1237/5) gefördert worden.
Besonders möchte ich meinen Kolleginnen Carola Eschenbach, Stephanie Kelter und Barbara Kaup danken, die Vorfassungen dieses Aufsatzes kritisch gelesen und kommentiert haben. Ich möchte mich bei ihnen und den LeserInnen dafür entschuldigen, dass ich einige der Vorschläge nicht berücksichtigt habe.
Diese Arbeit wird in einer leicht modifizierten, französisch-sprachigen Fassung in M. Denis (Ed.). *Langage et cognition spatiale*. Paris. Masson, erscheinen.

postuliert, dass sie von derselben Art sind wie diejenigen die bei der Wahrnehmung, Erinnerung oder beim Denken verwendet werden (vgl. Johnson-Laird, 1983 S. 397ff).[1] Entsprechend zu Verstehensprozessen kann auch für die Textproduktion angenommen werden, dass der Aufbau sprachlicher Strukturen – über die Zwischenstufe „propositionaler Repräsentationen“ – aus Sachverhaltsrepräsentationen erfolgt (Levelt, 1989 S 92ff).

Art der Repräsentation	Struktur der Repräsentation entspricht der
Textinhaltsrepräsentation	propositionalen Struktur
Verstehen ↓↑ Produzieren	
Sachverhaltsrepräsentation	Struktur des repräsentierten Sachverhalts

Abb. 1: Repräsentationsebenen in Sprachverarbeitungsprozessen

Zentral für die Textverarbeitungsansätze, die über propositionale Textinhaltsrepräsentationen hinaus Sachverhaltsrepräsentationen annehmen, ist das Kriterium der Strukturähnlichkeit zwischen Repräsentation und der repräsentierten Situation. Johnson-Laird formuliert – für Mentale Modelle unter der Perspektive des Textverstehens – sogar das strengere

> „Principle of structural identity: The structures of mental models are identical to the structures of the states of affairs, whether perceived or conceived, that the models represent.” (Johnson-Laird, 1983, S 419)

Offen bleibt jedoch meist – und dies gilt insbesondere auch für Johnson-Laird – was unter „Ähnlichkeit“ oder „Identität“ der Strukturen verstanden werden soll. Eine Klärung dieser Frage ist jedoch die unabdingbare Voraussetzung für jede Untersuchung in diesem Bereich. Im vorliegenden Aufsatz wird dieser Problemkreis auf räumliche Aspekte von Sachverhaltsrepräsentationen eingeschränkt: Wie gehen räumliche Eigenschaften der Repräsentationen in sprachliche Prozesse ein und welche Eigenschaften sind überhaupt als räumliche Eigenschaften zu bezeichnen?

Um die beiden – für die weiteren Überlegungen zentralen – Begriffe der „Strukturähnlichkeit“ und der „Räumlichkeit“ zu präzisieren, werde ich dem auf Palmer (1978) zurückgehenden

[1]Die hier skizzierte Auffassung ist nicht unumstritten: Im konkurrierenden „minimalistischen“ Theorieansatz von McKoon und Ratcliff (1992) etwa wird die Bildung einer Sachverhaltsrepräsentation nicht als integraler Bestandteil des Textverstehens angesehen, sondern nur als Produkt optionaler Verarbeitungstrategien. Zu Mentalen Modellen und zur Debatte zwischen Vertretern der Mentalen Modelle einerseits und des minimalistischen Ansatzes andererseits siehe auch Kelter, Kaup und Habel. (i. Vorb.).

Ansatz einer „allgemeinen Repräsentationstheorie" folgen (vgl. Rehkämper, 1991; Habel, 1994). Eine *Repräsentationssituation* ist gegeben durch eine repräsentierte Welt ($\mathcal{W}_1$), eine repräsentierende Welt ($\mathcal{W}_2$)[2] und eine Repräsentationsabbildung (ρ) von der repräsentierten Welt in die repräsentierende Welt

$$\rho: \mathcal{W}_1 \rightarrow \mathcal{W}_2 .$$

Repräsentationssituation können also durch Tripel $< \mathcal{W}_1, \mathcal{W}_2, \rho >$ charakterisiert werden. Die Komponenten dieses Tripels müssen, um überhaupt in adäquater Weise über Repräsentationsbeziehungen sprechen zu können, weiter spezifiziert werden:

- Welten $\mathcal{W}_i$ werden durch relationale Systeme $< D_i, Pr_i, Rel_i >$ expliziert, bestehend aus einem Diskurs-Universum D_i, d.h. einer Menge von Entitäten, einer Menge von Eigenschaften Pr_i und einer Menge von Relationen Rel_i über diesem Universum. (Diese Notation ist an die Modelltheorie angelehnt.)
- ρ ist eine partielle Abbildung, die einige der Eigenschaften und Relationen „respektiert"; die intuitive Idee des „Respektierens" wird dadurch expliziert, dass ρ als Homomorphismus (strukturerhaltende Abbildung) bezüglich ausgewählter Eigenschaften und Relationen angenommen wird.[3]

Verschiedene repräsentierende Welten, etwa $\mathcal{W}_2$ und $\mathcal{W}_3$, derselben repräsentierten Welt $\mathcal{W}_1$ unterscheiden sich insbesondere dadurch, dass gewisse Eigenschaften und Beziehungen von $\mathcal{W}_1$ durch

$$\rho_2: \mathcal{W}_1 \rightarrow \mathcal{W}_2 \quad \text{bzw.} \quad \rho_3: \mathcal{W}_1 \rightarrow \mathcal{W}_3$$

in unterschiedlicher Weise abgebildet bzw. respektiert oder nicht respektiert werden. So können etwa sowohl ein Globus ($\mathcal{W}_2$) als auch eine Weltkarte ($\mathcal{W}_3$) als Repräsentationen der

[2] Die repräsentierende Welt $\mathcal{W}_2 = \rho(\mathcal{W}_1)$ wird häufig als „Repräsentation von $\mathcal{W}_1$" bezeichnet. Die Palmersche Redeweise, meist von ***Welten*** statt von *Repräsentationen* zu sprechen, ist darin begründet, dass auch Repräsentation wieder Ausgangspunkt weiterer Repräsentationsprozesse sein können, also Repräsentationen von Repräsentationen betrachtet werden können. (So repräsentieren Texte Situationen der realen Welt und beim Textverstehen werden Repräsentationen von Texten und Textinhalten aufgebaut; vgl. Abb. 2.)

[3] In der kognitionswissenschaftlichen Diskussion – vgl. Palmer (1978) – wird häufig eine Repräsentation $\mathcal{R}$ als eine *analoge Repräsentation* einer Welt $\mathcal{W}$ bezeichnet, wenn die Repräsentationsabbildung $\rho: \mathcal{W} \rightarrow \mathcal{R}$ strukturerhaltend ist. Eine derartige Charakterisierung kann nur dann erfolgreich sein, wenn Welt und Repräsentation – als relationale Systeme $\mathcal{W}_i$ – sowie die Repräsentationsabbildung ρ voll spezifiziert sind und hierbei insbesondere explizit festgelegt ist, bzgl. welcher Eigenschaften und Relationen Strukturerhaltung gefordert ist. Mit anderen Worten: Die zweistellige Beziehung „...ist analoge Repräsentation zu..." erfordert die Ergänzung um eine dritte Stelle „...bzgl. der Aspekte ...", die über die Repräsentationsabbildung ρ festgelegt werden.

Erde ($\mathcal{W}_1$) aufgefasst werden; während in $\mathcal{W}_2$ Nachbarschaften respektiert werden, ist es aufgrund der zwei-dimensionalen Darstellungsweise von $\mathcal{W}_3$ nicht möglich, alle Nachbarschaften explizit zu repräsentieren. Erst unser Wissen über Weltkarten, genauer die Konvention, dass linker und rechter Rand der Karte zu identifizieren sind, ermöglicht es uns, die Nachbarschaftssituation auf der Erde auch auf einer Weltkarte nachzuvollziehen.[4] Grössenverhältnisse, wie Distanzen und Winkel, sind auf Globen normalerweise isomorph repräsentiert, auf zwei-dimensionalen Karten hängt die Ähnlichkeit zwischen $\mathcal{W}_1$ und $\mathcal{W}_3$von der gewählten Projektionsmethode ab. Vertikale Distanzen, d.h. Höhen- und Tiefeninformation, sind in beiden Repräsentationen üblicherweise durch Farbkodierungen dargestellt. Durch eine adäquate Farbwahl kann auch hier eine Strukturähnlichkeit erreicht werden, dann nämlich, wenn die *zwischen*-Struktur des Farbraumes und die der vertikalen Dimension der Erdkugel homomorph in Beziehung gesetzt werden, also die folgende Bedingung erfüllt wird:[5]

$$\text{blauton}(x) < \text{blauton}(y) < \text{blauton}(z) \Leftrightarrow \text{meerestiefe}(x) < \text{meerestiefe}(y) < \text{meerestiefe}(z)$$

Nachdem die zentralen Ideen einer allgemeinen Repräsentationstheorie – in ihren Grundzügen – eingeführt sind, kann ich auf die spezifischen Problemstellungen im Hinblick auf Sachverhaltsrepräsentationen zurückkommen. Diese stehen in repräsentationellen Beziehungen zu sprachlichen Entitäten einerseits und der realen Welt andererseits (vgl. Abb. 2), wobei von den Proponenten der der Konzeption Mentaler Modelle verpflichteten Textverarbeitungsansätze angenommen wird, dass zwischen der Sachverhaltsrepräsentation und der durch sie repräsentierten Situationen in der realen Welt eine interessante, d.h. nichttriviale, Strukturähnlichkeit besteht. Im Zentrum zahlreicher Untersuchungen und auch der weiteren Überlegungen liegt die Frage, wie einerseits die Strukturähnlichkeit – insbesondere im Hinblick auf räumliche Eigenschaften – zwischen der realen Welt und Sachverhaltsrepräsentationen in Einklang gebracht werden können mit der Rolle der Sachverhaltsrepräsentationen beim Verstehen und Produzieren natürlicher Sprache; das hier angesprochene Linearisierungsproblem wird in Abschnitt 3.1 diskutiert.

---

[4] Genauer: Die Identifizierung / Verklebung von linkem und rechtem Rand macht ein Rechteck homeomorph zu einer Zylinderoberfläche. Erst die zusätzliche Identifizierung aller oberen Randpunkte zu einem Punkt (entspricht dem Nordpol) und ebenso aller unteren Randpunkte (Südpol) stellt einen Homeomorphismus zu einer Kugel her. (Vorsicht: „homomorph" bezieht sich auf Abbildungen zwischen (fast) beliebigen Strukturen, „homeomorph" auf Abbildungen zwischen topologischen Räumen.

[5] Gärdenfors (1995) erläutert die konvexe Struktur des kognitiven Farbraums; Habel und Eschenbach (1995) stellen die Relevanz von *zwischen*-Strukturen für abstrakte und konkrete Räumlichkeit dar.

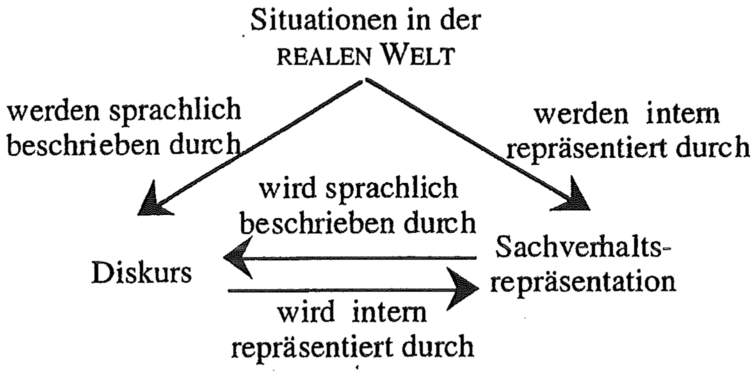


Abb. 2: Repräsentationsebenen in Sprachverarbeitungsprozessen

Wenn – wie in Abb. 1 dargestellt – innerhalb von Sprachverarbeitungsprozessen sowohl eine Ebene propositionaler Repräsentationen als auch eine Ebene von zur realen Welt strukturähnlichen, „analogen" Sachverhaltsrepräsentationen angenommen wird, ist zu klären, in welcher Weise die Verarbeitungsmechanismen von den beiden verschiedenen Repräsentationsebenen bzw. Repräsentationsformaten Gebrauch machen. Hier bieten sich im Wesentlichen zwei Architekturalternativen an: zum einen die auf Übersetzungsprozessen zwischen den Repräsentationsebenen basiert, zum anderen eine *hybride Architektur*, bei der einige der Komponenten auf beide Repräsentationsebenen zugreifen, d.h. beide Formate verwenden.[6] Die in Abschnitt 3.2 vorgestellten Prozesse der Sprachproduktion gehen von einem hybriden Repräsentationssystem, in dem propositionale und bildhaft-räumliche Repräsentationen integriert sind, aus.[7]

[6] Abb. 1 ist gegenüber diesen Architekturalternativen neutral; die propositionale Ebene ist gegenüber der Sachverhaltsebene dahingehend ausgezeichnet, dass sie den Übergang zu anderen sprachlichen Ebenen, wie der syntaktischen, ermöglicht.

[7] Auch im Rahmen des Ansatzes mentaler Modelle werden „hybride Repräsentationen" – ohne dass diese explizit so benannt würden – verwendet: Johnson-Laird (1995, S. 1007) schlägt den Einsatz *propositionaler Annotationen* u.a. für Alternativenbehandlung und Negation vor. Die Interaktion von propositionalen Repräsentationen und mentalen Modellen, über deren räumlichen Eigenschaften – im Sinne des Abschnittes 2 – bei Johnson-Laird keinerlei befriedigenden Aussagen gemacht werden, spielen in den Ansätzen der mental-model Theorie bestenfalls eine periphere Rolle.

2 Räumliche Repräsentationen und Repräsentationen des Raumes

Was macht nun eine Repräsentation zu einer *räumlichen Repräsentation* ? Dies zu klären ist mir deswegen wichtig, da der Unterschied zwischen „Repräsentation räumlicher Konstellationen" bzw. „Repräsentation des Raumes" einerseits und „räumlicher Repräsentation" bzw. „Repräsentation mit räumlichen Eigenschaften" andererseits häufig nicht mit hinreichender Klarheit gemacht wird; während der erste dieser Gesichtspunkte Eigenschaften der repräsentierten Welt betrifft, sind beim zweiten Eigenschaften der repräsentierenden Welt, bzw. des Mediums, in dem die Repräsentation und Verarbeitung erfolgt, angesprochen.

Die Frage nach der *Räumlichkeit interner Repräsentationen* macht den Kern der – besonders intensiv in den 70er und frühen 80er Jahren ausgetragenen – Kontroverse, die als „imagery debate" bezeichnet wird, aus:[8] Welche Art von Repräsentationen liegen Imaginationsprozessen bzw. „mentalen Bildern" zugrunde? Sind es Repräsentationen, die einen bildhaft-räumlichen Charakter aufweisen, wie es etwa Kosslyn (1980) annimmt, oder propositionale Repräsentationen, mit denen zwar räumliche Konstellationen dargestellt werden, die jedoch keine Strukturähnlichkeit zum physikalischen Raum aufweisen (vgl. Pylyshyn, 1981)?

Als Evidenz für die Existenz räumlicher Repräsentationen wurden beispielsweise die Resultate zahlreicher Untersuchungen zur *mentalen Rotation* (z.B. Shepard & Cooper, 1982) und die Befunde beim Absuchen („*scanning*") von mentalen Vorstellungen (z.B. Kosslyn, Ball & Reiser, 1978) gesehen.[9] Die Grundidee des experimentellen Paradigmas der mentalen Rotation besteht darin, Versuchspersonen zwei Stimuli in unterschiedlichen Orientierungen zu präsentieren und auf Formgleichheit beurteilen zu lassen. Die abhängige Variable ist hierbei die Verarbeitungszeit. In zahlreichen Studien ergab sich mit zunehmender Winkeldifferenz zwischen den dargebotenen Reizen eine monoton steigende Reaktionszeit; Metzler und Shepard (1975) etwa fanden eine lineare Abhängigkeit zwischen Winkeldifferenz und Reaktionszeit. Dies legt die Schlussfolgerung nahe, dass bei der Prüfung auf Gleichheit der beiden Stimuli eine mentale Drehung der mentalen Repräsentation des Testreizes stattfindet und somit die mentale Rotation als Analogon zu einer realen Drehbewegung, deren Zeitbedarf

---

[8] Vgl. hierzu die in Block (1981) publizierten Arbeiten, die für den Höhepunkt der Kontroverse Anfang der 80er Jahre charakteristisch sind, sowie die zusammenfassenden Darstellungen von Denis (1991, Kap. 2) und Kosslyn (1994, Kap. 1). Die Bedeutung dieser Kontroverse für die Sprachverarbeitung habe ich in Habel (1987) dargestellt.

[9] Eine Diskussion alternativer Erklärungsversuche, d.h. solcher, die nicht auf bildhaften bzw. räumlichen Repräsentationen basieren, gibt Farah (1988); vgl. auch Kosslyn (1994).

von der Grösse des Drehwinkels abhängt, angesehen werden kann. Kosslyn et al. fanden weiterhin eine lineare Abhängigkeit zwischen relativem euklidischem Abstand auf einer Landkarte und den Reaktionszeiten für das Absuchen der mentalen Vorstellung.

In den hier vorgestellten Experimenten liegt – aus Sichtweise der oben skizzierten Repräsentationstheorie – eine Repräsentationssituation zwischen der externen Welt der Stimuli, $\mathcal{W}_{stim}$, und der Welt $\mathcal{W}_{image}$, der internen Repräsentationen der Ausprägung mentaler Bilder, vor. Die genannten Befunde sprechen dafür, dass die Repräsentationskorrespondenz ein Homomorphismus ρ: $\mathcal{W}_{stim} \rightarrow \mathcal{W}_{image}$ ist, der *Winkel* und *Abstand* respektiert, wobei *Winkel* und *Abstand* jeweils in $< D_{stim}, Pr_{stim}, Rel_{stim} >$ und $< D_{image}, Pr_{image}, Rel_{image} >$ expliziert sind. Der Repräsentation $\mathcal{W}_{image}$ das Attribut *räumlich* zuzusprechen, scheint deswegen angemessen, weil Winkel bzw. Abstand, die für die Repräsentationsabbildung als relevant angesehenen Aspekte, in die Klasse der zweifelsfreien räumlichen Konzepte fallen (s.u.).

Eine Repräsentation als räumlich zu charakterisieren, wenn sie zum physikalischen Raum korrespondiert, ist auch die Grundlage für Finkes Erläuterung der räumlichen Eigenschaften mentaler Vorstellungen, die er auf dem „Prinzip räumlicher Äquivalenz“ aufbaut:

> „The spatial arrangement of the elements of a mental image corresponds to the way objects or their parts are arranged on actual physical surfaces or in actual physical space.“ (Finke, 1989, p.61)

Unbefriedigend ist hierbei jedoch, dass keine Explikation von „spatial arrangement“ vorgenommen wird, und dass somit über die Art der Korrespondenz zwischen Repräsentation und physikalischer Realität eigentlich alles offen bleibt.

Um dieses Desiderat zu überwinden, werde ich im Weiteren – auf Habel und Eschenbach (1995a, 1995b) aufbauend – *mathematische Raumkonzepte* als „Referenzkonzepte“ verwenden. Mit anderen Worten: Eine mentale Repräsentation $\mathcal{W}_{mental}$, in der Repräsentationssituation

$$\rho: \mathcal{W}_{real} \rightarrow \mathcal{W}_{mental},$$

ist eine räumliche Repräsentation, falls die durch ρ respektierten Eigenschaften und Beziehungen in die Klasse *mathematischer Raumkonzepte* gehören. Aus der Charakterisierung der *Räumlichkeit* mentaler Repräsentationen auf der Basis *mathematischer Raumkonzepte* kann eine Systematisierung von „Aspekten der Räumlichkeit“ geleistet werden. Diese beruht auf einer Gliederung in vier Konzeptfamilien, die durch die Bezeichnungen *Topologie*, *Anordnung*, *Gestalt* und *Metrik* gekennzeichnet sind. Während durch *Topologie*, *Anordnung* und *Metrik* wohldefinierte Bereiche der Mathematik angesprochen sind – hierbei stehen *Anordnung* und *Metrik* für unterschiedliche Ausprägungen der Geometrie – ist der durch*Gestalt* angesprochene Bereich in der Mathematik nicht in der von Habel und Eschenbach vorgesehenen Weise systematisch erfasst. Wir fassen hierunter Eigenschaften von Objekten und Regionen auf, die unabhängig von ihrer Lokation im Raum aber wesentlich durch die *relative Position* ihrer Teile zueinander geprägt sind bzw. Eigenschaften und Relationen, die diese Eigenschaften mit der *relativen Position* der Teile zu anderen Objekte verknüpfen (z.B. Ausrichtung).

Die hier angesprochenen räumlichen Konzeptfamilien haben auch – über den physikalischen Raum hinaus – in anderen konkreten, aber auch abstrakte Domänen, wie etwa bei *Zeit*, *Farbe* aber auch *Ereignissen* bzw. *Situationen*. Bedeutung. Auf der Basis von Untersuchungen zur Korrespondenz derartiger Einteilungen argumentieren Habel und Eschenbach (1995a) dafür, die Räumlichkeit verschiedener kognitiver Domänen in gemeinsam zugrundeliegenden Strukturen zu suchen, die als Abstraktion aus dem konkreten, physikalischen Raum aufgefasst werden können. Unter dieser Perspektive abstrakter Raum-Strukturen kann die Beobachtung der kognitiven Grammatik, dass weite Bereiche der Semantik auf konkreten räumlichen Konzepten begründet sind, auch eine neue Erklärung erfahren:[10] Die gemeinsame abstrakte Struktur verschiedener Domänen ist für den gemeinsamen Gebrauch sprachlicher Ausdrücke verantwortlich. Der konkrete Raum ist eine besonders prominente kognitive Domäne, die die abstrakt erfassbaren Raum-Strukturen aufweist. Damit wird die Relevanz des konkreten 3-dimensionalen Raumes für kognitive Systeme nicht bestritten. Erfahrungen gerade in diesem Bereich können Ausgangspunkt der einschlägigen Abstraktionen sein.

Gemeinsame abstrakt räumliche Basisstrukturen sind die Basis der Intermodalität räumlicher Repräsentationen, die eine „Schnittstellenfunktion“ zwischen Sprache einerseits und visueller, auditiver und haptischer Perzeption, sowie Motorik und räumlichem Schliessen andererseits ausüben (siehe Abb. 3).

---

[10] Innerhalb der kognitiven Grammatik, in ihren Anfängen teilweise sogar als „space grammar“ bezeichnet, wird die Raumbasierung der Semantik zum Programm erhoben; vgl. u.a. Lakoff (1987), Langacker (1982, 1986), Talmy (1983).

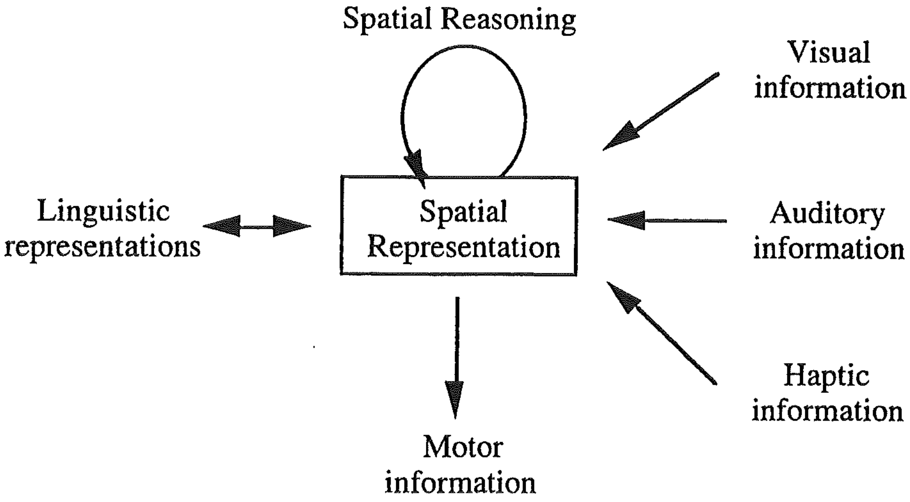


Abb. 3: Räumliche Repräsentationen als Schnittstelle zwischen Modalitäten der Kognition[11]

Wenn – wie oben erläutert – gewisse mentale Repräsentationen, etwa mentale Vorstellungen, einen räumlichen Charakter besitzen, ist zu klären, wie derartige Repräsentationen realisiert werden. Wichtige Hinweise hierzu geben die kognitiven Neurowissenschaften.

Mittlerweile liegt umfangreiche neurowissenschaftliche Evidenz für Kosslyns Forschungshypothese der letzten 20 Jahre „Mentale Bilder sind als Repräsentationen in einem räumlichen Format auf einem räumlichen Medium realisiert." vor: einerseits konnte nachgewiesen werden, dass z.T. dieselben Hirnareale, die bei der visuellen Wahrnehmung Aktivität aufweisen, auch in die Prozesse des Imaginierens und Inspizierens mentaler Bilder involviert sind, andererseits konnte gezeigt werden, dass dieser Areale retinotop organisiert sind (z.B. V17, vgl. hierzu: Kosslyn, 1994; pp.12-20, chap. 3, chap. 11).

Entsprechende Repräsentationsansätze sind auch für künstliche kognitive Systeme, d.h. für Systeme der Künstlichen Intelligenz, vorgeschlagen worden:[12] Glasgow und Papadias (1992) versuchen in einer als „Computational Imagery" bezeichneten Konzeption an die menschliche

---

[11] Diese Abbildung kombiniert die Sichtweise von Abb. 2 aus Habel (1987) und Abb. 1 aus Landau & Jackendoff (1993). Eine detaillierte Diskussion der hier zwischen den verschiedenen modalitätsspezifischen Repräsentationssystemen dargestellten Transformationen insbesondere im Hinblick auf die von Landau und Jackendoff erwähnte – aber in ihren Konsequenzen nicht explizierte – Unterscheidung in die Ausprägungen der „Heteromodalität" und der „Amodalität", kann im vorliegenden Aufsatz nicht geleistet werden.

[12] Die beiden im weiteren vorgestellten KI-Ansätze sind explizit dem Paradigma der mentalen Vorstellungen verpflichtet; dies zeigt sich u.a. in einer offensichtlichen und auch beabsichtigten Verwandtschaft zu Kosslyns eigenen Systemen der Computer-Simulation aus der ersten, „pre-neural-net"-Phase (vgl. Kosslyn, 1980).
Darüber hinaus gewinnt in den letzten Jahren der Bereich des „diagrammatischen Schliessens" zunehmend an Gewicht, in dem unterschiedliche Arten von räumlichen Repräsentationen und Repräsentationen des Raumes eine Rolle spielen (vgl. Glasgow, Narayanan & Chandrasekaran, 1995).

Imagination angelehnte Schlussmethoden als generelles Problemlösungsparadigma – in Ergänzung zu klassischen, symbolischen Methoden – zu etablieren. Hierbei werden räumliche und visuelle Repräsentationen im Arbeitsspeicher durch ineinander geschachtelte, symbolische Arrays bzw. dreidimensionale „occupancy arrays“ unterschiedlicher Detailliertheit realisiert. In Habel (1987) habe ich gezeigt, dass räumliche Repräsentationen der Art, die zu „mentalen Vorstellungen“ korrespondieren, eine zentrale Rolle in der Sprachverarbeitung einnehmen: Die Verbindung zwischen „Depiktionen“ (Repräsentationen in einem bildhaften Format von 2-dimensionalen Matrizen) und propositionalen Repräsentationen wird innerhalb dieses Ansatzes dadurch hergestellt, dass eine propositionale Annotation zu den Zellen der räumlichen Repräsentation existiert. Hierdurch erhält das Repräsentationssystem einen hybriden Charakter. Das am Beispiel von Annotationen zu Zellmatrizen skizzierte Verweissystem zwischen propositionalen und räumlichen Repräsentationen ist nicht an das Format der Zellmatrizen gebunden:[13] Habel, Pribbenow und Simmons (1995) schlagen die Berücksichtigung von speziellen Teil-Ganzes-Beziehungen konkreter Objekte im konzeptuellen Lexikon durch ein hybrides Repräsentationssystem vor, wobei das räumliche Repräsentationsformat an Biedermans Konzeption der „geometrical ions“ (geons) orientiert ist (vgl. Biederman, 1987).

---

[13] Gerade bei Matrizen bzw. „arrays“ scheint diese Konzeption besonders nahe zu liegen. Auch Glasgow und Papadias (1992) folgen dieser Idee, um eine hybride propositional–bildhaft-räumliche Aspekte integrierende Verarbeitung zu erreichen.

3 Räumliche Repräsentationen in der Sprachproduktion

3.1 Linearisierung – ein zentrales Problem der Sprachproduktion

Unter der in den vorangegangenen Abschnitten erläuterten Sichtweise, Sprachverarbeitungsprozesse als Prozesse anzusehen, die aus Äusserungen interne Repräsentationen erstellen bzw. ausgehend von internen Repräsentationen sprachliche Äusserungen erzeugen, ergibt sich für die Thematik „Sprache und Raumkognition" die folgende zentrale Fragestellung:

> Wie wird die Korrespondenz zwischen Sachverhaltsrepräsentationen mit räumlichem, d.h. insbesondere 2- oder 3-dimensionalem, Charakter und sprachlichen Äusserungen (Lautfolgen, Texten), denen insbesondere aufgrund der zeitlichen Bedingungen der Produktion bzw. Rezeption eine lineare (1-dimensionale) Struktur zugrundeliegt, hergestellt?

Das Problem, bei der Produktion eines Diskurses eine geeignete Linearisierung vorzunehmen,[14] wird von Denis (1991, p. 147) durch „How will people confronted with a two- or three-dimensional spatial entity 'linearize' information in order to produce descriptive discourse?" charakterisiert. Es ist ein spezieller Fall des generellen – nicht nur auf den Fall der Beschreibung räumlicher Gegebenheiten beschränkten – Linearisierungsproblems, das von Levelt (1989, p. 138) mit „... deciding what to say first, what to say next, and so on...." beschrieben wird.

Als grundlegendes Prinzip für die Linearisierung wird von Levelt das *Prinzip der natürlichen Anordnung* formuliert: „Arrange information for expression according to the natural ordering of its content" (Levelt, 1989, p. 380). Diesem Prinzip folgend wird, wenn der zu verbalisierende Sachverhalt eine geeignete interne Struktur besitzt, diese im Normalfall auch herangezogen. In experimentellen Untersuchungen zu Linearisierungsprinzipien werden – beginnend mit den von Levelt 1982 durchgeführten Arbeiten zur Beschreibung räumlicher Netzwerke – überwiegend solche Arrangements als Material verwendet, in denen – zumindestens partiell – lineare Strukturen vorliegen. Ziel dabei ist, Prozess-bezogene Einflussfaktoren („process-related determinants"; vgl. Levelt, 1989, pp. 139ff) zu ermitteln, also solche, die nach bzw. neben der Verwendung von im Material gegebenen linearen Teilstrukturen die Linearisierung bestimmen. Robin und Denis (1991) konnten zeigen, dass

---

14 Beim Textverstehen, insbesondere bei der Bildung von Sachverhaltsrepräsentationen, ist das inverse Problem des Aufbaus höherdimensionaler Strukturen aus linearem Diskurs zu lösen. Einige zentrale Arbeiten zum sprachlich induzierten Aufbau von Sachverhaltsrepräsentationen sind in Abschnitt 1 angesprochen; vgl. auch speziell im Hinblick auf die räumlichen Charakteristik derartiger Repräsentationen Taylor und Tversky (1992) und Tversky, Taylor und Mainwairing (1995).

die bei der Beschreibung perzipierter Konstellationen verwendeten Linearisierungsprinzipien auch bei der Beschreibung imaginierter Arrangements eingesetzt werden. Abb. 4 zeigt eine für derartige Experimente typische Netzstruktur, ähnlich zu solchen, die von Levelt (1982) bzw. Robin und Denis (1991) verwendet wurden.[15] Die typische Versuchsaufgabe besteht nun darin, das gesamte Layout zu beschreiben, also ein zwei-dimensionalen Arrangement (direkt wahrgenommen oder imaginiert) im Zuge der Äusserung in eine ein-dimensionale, lineare Abfolge zu überführen; als einzige Zusatzbedingung wird der Startknoten, hier die „Busstation", vorgegeben.

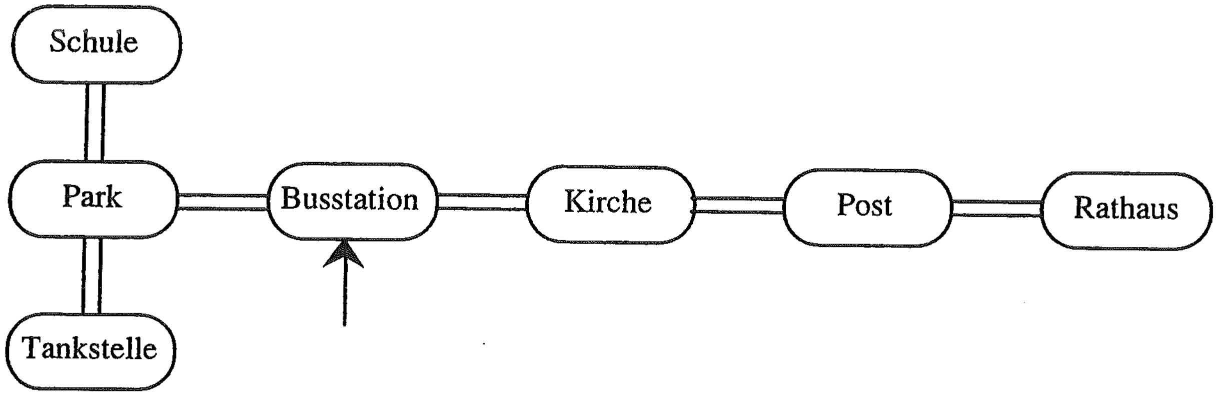


Abb. 4: Räumliches Netzwerk für Linearisierungsaufgabe

Für die Beschreibung von Netzwerken konnte Levelt (1982; 1989, p. 140-144) drei spezielle Prinzipien ableiten, denen die Versuchspersonen zu folgen scheinen, um die Verarbeitungs- insbesondere die Gedächtnisbelastung bei sich, aber auch bei den Adressaten der Beschreibung, möglichst gering zu halten:

Principle of connectivity „Wherever possible, choose as the next node to be described one that has a direct connection to the current node." (Levelt, 1989, p. 140)

Stack principle „Return to the last choice node in the waiting line." (Levelt, 1989, p. 143)

Minimal-load principle „Order alternative continuations in such a way that the resulting memory load for return addresses is minimal." (Levelt, 1989, p. 143)

Diese Prinzipien sind dem Anordnungsprinzip untergeordnet: das grundlegende Ziel, in der zu beschreibenden Struktur eine lineare Anordnung aufzudecken und somit dem Anordnungsprinzip zu folgen, kann dadurch erreicht werden, dass zumindestens eines der drei spezielleren Prinzipien angewendet wird.

Obwohl Levelt in der Formulierung des Konnektivitäts- und des Stapel-Prinzips explizit Bezug auf *Knoten*, d.h. auf Objekte in der konkret räumlichen Netzstruktur des Versuchs-

[15] Levelt (1982) verwendete Netze, deren Knoten farbige Kreise waren; die in Abb. 4 eingebrachte Motivation im Hinblick auf Landmarken an Strassen geht auf Fig. 1 von Robin und Denis (1991) zurück.

materials, nimmt, kann davon ausgegangen werden, dass entsprechende Prinzipien auch in anderen Domänen Gültigkeit besitzen, da sie allgemeine Eigenschaften der Wahrnehmung und des Gedächtnisses widerspiegeln (Levelt, 1989, p. 140). Nimmt man etwa – wie in Abschnitt 2 skizziert – an, dass räumliche Repräsentationen einen intermodularen, abstrakt räumlichen Charakter besitzen, so sind die von Levelt angesprochenen „general properties of perception and memory", die die Übertragbarkeit der genannten Prinzipien ermöglichen, gerade im Bereich der räumlichen Repräsentationen anzusiedeln. Unter dieser Sichtweise kann für die Leveltschen Versuche angenommen werden, dass durch die räumliche Konnektivität der netzartigen Reizmuster Konnektivitätsbeziehungen in den internen Repräsentationen induziert werden. Auch die von Levelt diskutierten Beispiele für eine Abschätzung der Gedächtnisbelastung (minimal-load principle) berücksichtigen räumliche Eigenschaften der Netze: Für die Entscheidung, bei der Beschreibung des Netzwerkes in Abb. 4 zuerst den rechten und erst anschliessend den linken Ast zu bearbeiten, ist nicht allein die Anzahl der Knoten der jeweiligen Äste ausschlaggebend, sondern insbesondere die Tatsache, dass der rechte Ast unverzweigt ist und der linke eine Verzweigung aufweist.[16] Die für das hier diskutierte Linearisierungsprinzip relevante Gedächtnisbelastung ist funktional von räumlichen Struktureigenschaften des Netzes abhängig.

Derartige indirekte Einflüsse räumlicher Konstellationen will ich im weiteren an einer anderen, meist als exemplarisch nichträumlich charakterisierten Domäne erläutern: der Domäne der Ereignisse. Ereignisse werden häufig als Standardfälle im Hinblick auf das Prinzip der natürlichen Ordnung angesehen: „What counts as natural ordering is different for different domains of discourse, and their is no general definition. Still, for certain important cases the notion is obvious. For event structures, the natural order is the chronological order of events." (Levelt, 1989, p. 138).

Probleme für die lineare Anordnung von Ereignisrepräsentationen entstehen insbesondere dann, wenn die Korrespondenz zwischen den für die sprachliche Beschreibung relevanten Ereignissen und der linearen Struktur der Zeit eine (n:1)-Beziehung ist; mit anderen Worten, wenn – partiell – gleichzeitig stattfindende Ereignissen beschrieben werden sollen. Beispielhaft wird dies im folgenden Text, der durch Glenberg und Langston (1992) angeregt ist, deutlich:

> „Zuerst legte Arthur die grundsätzliche Struktur des Aufsatzes fest und schrieb dann die erste Fassung. Danach überprüfte und verbesserte er die Argumentation und achtete besonders darauf, dass Terminologie und Darstellung für die

[16] Abb. 4 ist im wesentlichen Teil strukturgleich zum von Levelt (1989, p. 143f) diskutierten Netz 'e'.

Leserschaft angemessen sind. Abschliessend korrigierte er Rechtschreibung und Grammatik und glättete den Stil."

Die Notwendigkeit im Produktionsprozess zu linearisieren, zwingt den Produzenten, gleichzeitig stattfindende Ereignisse nacheinander zu beschreiben. Der in Abb. 5 dargestellte Graph kann während des Versprachlichungsprozesses nicht in der durch die zeitliche Ordnung gegebenen kanonischen Weise durchlaufen werden, vielmehr müssen Entscheidungen – ähnlich zu denen in der Beschreibung von räumlichen Netzwerken – getroffen werden, die der nicht linearen Ereignisstruktur eine zusätzliche lineare Anordnung aufprägen.

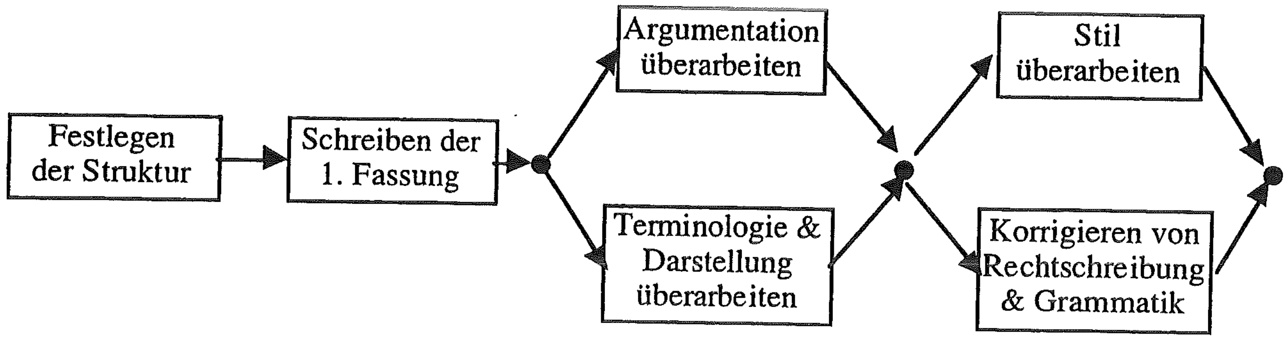


Abb. 5: Nichtlineare Ereignisstruktur: Nebenläufigkeit[17]

Nichtlineare Ereignisstrukturen sind nicht auf Fälle offensichtlicher Nebenläufigkeit – in Abb. 5 durch Verzweigungen dargestellt – beschränkt. In jedem Fall, in dem eine Ereignishierarchie vorliegt, treten damit auch nichtlineare Ereignisstrukturen auf.[18] Wenn etwa eine Person *p* mit eine Schreibmaschine „ein Wort" schreibt, so schreibt *p* nacheinander die Buchstaben dieser Phrase; Abb. 6 stellt die Struktur einer Ereignisrepräsentation, die die hierarchischen Beziehungen zwischen den beteiligten Ereignissen berücksichtigt, dar.[19] Die „natürliche Ordnung der Zeit" spiegelt sich hier jeweils nur in einer Ebene des Ereignisbaumes wider: die Zeit, zu der ein *Tochterereignis* stattfindet, ist enthalten in der

---

[17] Das hier vorgestellte Beispiel und die daraus abgeleitete nichtlineare Ereignisstruktur sind Varianten eines von Glenberg und Langston (1992) verwendeten Instruktionstextes (Table 1, p 132) und einer Ablaufstruktur (Fig. 1. p.133). Gegenstand dieser Untersuchung ist der Einfluss von Graphiken beim Textverstehen, insbesondere unter der Perspektive der Integration piktorieller und sprachlicher Information beim Aufbau mentaler Modelle.

[18] Dies trifft z.B. auf propositionale Strukturen und Situationsmodelle zu, wie sie von Van Dijk und Kintsch (1983) vorgeschlagen werden.

[19] Bei der Repräsentation von Satz-Schreib-Ereignissen wären ausserdem höhere Ebenen im Ereignisbaum zu berücksichtigen. Handelt es sich um mit-der-Hand-schreiben, so wäre es notwendig, die Knoten, die in Abb. 6 Blätter sind, noch weiter zu expandieren, etwa dadurch, dass „*p* produziert einen senkrechten Strich" und „*p* produziert einen kürzeren waagerechten Strich, der den senkrechten Strich im oberen Drittel kreuzt" als Feinanalyse für das Ereignis „*p* schreibt ein *t*" verwendet wird. Darüber hinaus besteht in diesem Fall ein erhebliche Schwierigkeit darin, die Grenzen zwischen aufeinanderfolgenden Ereignissen zu bestimmen.

Zeit, in der das *Muttereregnis* stattfindet. Betrachtet man im Gegensatz zum Schreibereignis ein „Stempel-Ereignis“, so kann – abhängig davon, wie der Stempel beschaffen ist – eine zu Abb. 6 isomorphe Ereignisstruktur vorliegen, in der jedoch alle Ereignisse, und zwar aller drei Ebenen, gleichzeitig stattfinden. Während die zeitliche Inklusion – im extremen Fall: die zeitliche Identität – zwischen den den Ereignissen kanonisch zugeordneten zeitlichen Entitäten zur „is_part_of“ Beziehung innerhalb der hier angesprochenen Ereignishierarchien korrespondiert, besteht für die zeitliche Präzedenz keine systematische Korrespondenz zu Beziehungen in der Ereignishierarchie.

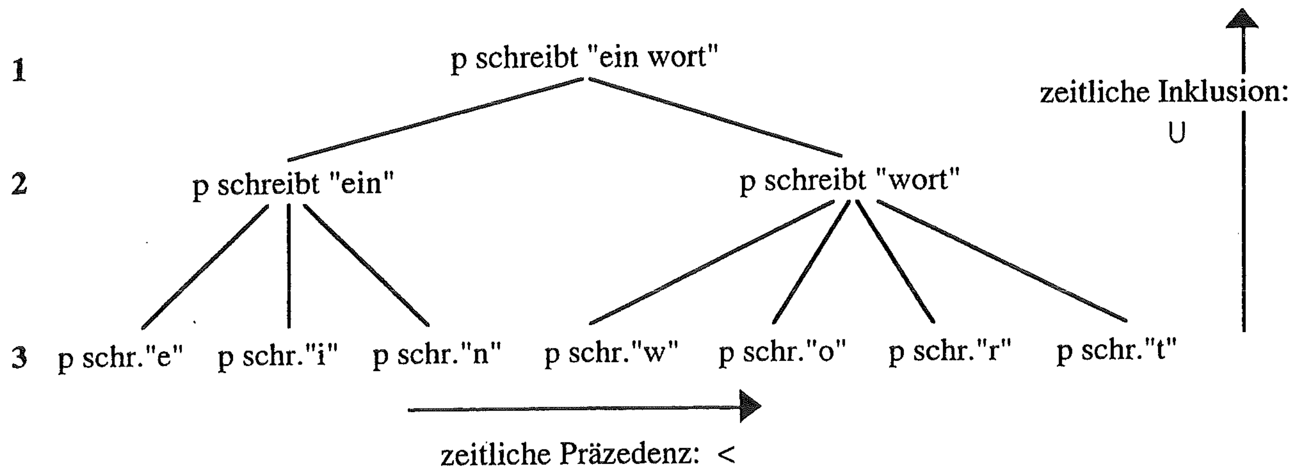


Abb. 6: Nichtlineare Ereignisstruktur: Ereignishierarchie und zeitliche Beziehungen

Soll ein wie in Abb. 6. veranschaulichtes Schreibereignis sprachlich beschrieben werden, so sollte – dem von Levelt formulierten Prinzip der natürlichen Anordnung folgend – der Ereignisbaum auf einer Ebene traversiert werden. Die Berücksichtigung eines Knotens und seiner Töchterknoten in einer Äusserung, wie etwa bei: „*p* schreibt ein ‘*e*‘, ein ‘*i* ‘und ein ‘*n*‘, also ‘*ein*‘“, ist durch das Prinzip der natürlichen Anordnung nicht abgedeckt. Ein Wechsel zwischen den Ebenen, etwa „*p* schreibt ein ‘*e*‘, ein ‘*i* ‘und ein ‘*n*‘, und dann ‘*wort*‘“, ist zwar verträglich zum Anordnungsprinzip würde jedoch sicherlich als ungewöhnliche oder sogar abweichende Äusserung angesehen werden.

Im folgenden Abschnitt werde ich exemplarisch an Hand einer Fallstudie erläutern, dass und wie in einer Domäne, in der Situationsmodelle einen explizit räumlichen Charakter aufweisen, räumliche Konzepte auch für die Ereigniskonstitution und die Linearisierung von Ereignisbeschreibungen herangezogen werden.[20]

---

20 Der Problemkreis der Ereigniskonzeptualisierung – und hierin eingeschlossen die Fragen danach, unter welchen Bedingungen aus Ereignissen komplexe Ereignisse gebildet werden, und wann Ereignisse als voneinander separiert angesehen werden – kann hier nicht im Detail behandelt werden; vgl. zur Konzeptualisierung von Ereignissen: Hanson und Hirst (1989) sowie Avrahami und Kareev (1994).

3.2 Skizzenerstellungsereignisse – Eine Fallstudie

Ereignisstrukturen, bei denen in der Linearisierung nicht direkt, d.h. ohne Modifikationen, auf die „chonologische Ordnung" zurückgegriffen werden kann, sind in komplexeren Domänen sicherlich eher die Regel als die Ausnahme. Hierbei sind insbesondere die beiden m vorangegangenen Abschnitt angesprochenen Typen von Nichtlinearität von besonderer Bedeutung: zum einen die „Echte Nebenläufigkeit" von Ereignissen, zum anderen die durch Teil-von-Beziehungen induzierte „partielle Gleichzeitigkeit" zwischen Ereignissen und ihren Teilereignissen. Für die Sprachproduktionsforschung gilt es also Faktoren zu bestimmen, die in diesen Fällen die Linearisierung leiten. Insbesondere ist zu untersuchen, inwieweit Prinzipien, die für die Linearisierung von räumlichen Strukturen angenommen werden können – also etwa die von Levelt vorgeschlagenen, im vorangegangenen Abschnitt vorgestellten Prinzipen der Konnektivität, des Stapelspeichers und der minimalen Gedächtnisbelastung –, auch in anderen Domänen einsetzbar sind. Aus der Perspektive der Forschungen zur Raumkognition stellt sich die Frage, welche Rolle räumliche Konzepte bei der Verwendung dieser Prinzipien in nichträumliche Domänen spielen.

Um eine systematische Untersuchung der Linearisierungsprozesse vornehmen und eine Isolierung der Phänomene erreichen zu können, erscheint es sinnvoll, eine Domäne auszuwählen, in der ausschliesslich einer der beiden oben erläuterten Typen von Nichtlinearität auftritt, und zwar durch Teil-von-Beziehungen strukturierte Ereignishierarchien ohne echte Nebenläufigkeit. Im weiteren werde ich eine Vorstudie[21] zu einer entsprechenden Domäne vorstellen. Als Untersuchungsgegenstand wurden Prozesse der Skizzenerstellung – vielleicht besser als „Skizzenentstehung" bezeichnet – gewählt. Den Versuchspersonen wurde die Entstehung einer Skizze visuell – auf einem Bildschirm – präsentiert; die Aufgabe der Vpn bestand darin, den Prozess der Skizzenentstehung – nicht die Skizze – zu beschreiben. Das Stimulusmaterial waren Zeichnungen, die auf einem Graphiktablett erstellt wurden,[22] so dass nicht nur das Resultat der Skizzenerstellung, d.h. die Skizze (siehe Abb. 7), sondern insbesondere der genaue Zeitverlauf der Skizzenentstehung in

---

[21] Die hier vorgestellten Überlegungen beruhen auf einem Voruntersuchungen, bei dem insbesondere Erfahrungen mit der verwendeten Methode gewonnen sowie Weiterentwicklungen des experimentellen Aufbaus erreicht werden sollten. Detaillierte, systematische Experimente sind in Vorbereitung.
Ich möchte an dieser Stelle den TeilnehmerInnen des Seminars „Repräsentation und Verarbeitung komplexer Ereignisse" und insbesondere Andrea Schopp, die das Seminar gemeinsam mit mir durchführte, sowie Barbara Kaup und Mark Erichsen , die die Experimentalumgebung – einschliesslich spezieller Eingabe-, Speicherungs- und Präsentationssoftware – realisiert haben, danken.

[22] Das Skizzenmaterial wurde durch Studierende im Rahmen eines Seminars erstellt. Vor Erstellung der Skizzen hatten die Teilnehmer ausreichende Gelegenheit, sich mit dem Zeicheninstrument vertraut zu machen.

einer Datei protokolliert und die Skizzenentstehung reproduzierbar war. Diese „Reproduktionen der Skizzenerstellung" werden als Stimulus für die Beschreibungsaufgabe verwendet: eine speziell hierzu entwickelte Wiedergabekomponente lässt – auf dem zu Beginn leeren – Bildschirm graphische Objekte auftauchen. Der Zeitlauf der Präsentation (Reproduktion) entspricht dem der Skizzenerstellung. Die Versuchspersonen sehen also die Entstehung einer Skizze, ohne den Produzenten der Skizze beobachten zu können. Da auf dem Bildschirm strikt sequentiell zusätzliche Pixel sichtbar werden, liegt eine der oben skizzierten Idealisierung entsprechende Ereignisstruktur vor, eine Struktur nämlich, in der die „am wenigsten komplexen Ereignisse" – das sind die Blätter der Ereignishierarchie – strikt linear der chronologischen Anordnung folgen.

Das Gegenstandsbereich der präsentierte Zeichnungen sind Wege- / Lageskizzen mit Beschriftungen, die als ergänzende Information zu einer Wegbeschreibung dienen können. Die Produzenten der Zeichnungen waren aufgefordert worden, eine Skizze zu erstellen, die später bei einer verbalen Wegbeschreibung für die Route vom Informatik-Gelände der Universität Hamburg (in Stellingen) zum Campus der Universität als illustrierendes Material verwendet werden könnte.[23]

Für die Sprachproduktionsaufgabe wurden die Vpn instruiert, den Prozess der Skizzenentstehung – nicht die Skizze – zu beschreiben; die Aufgabenstellung wurde in zweierlei Hinsicht variiert: (i) „on-line"-Beschreibungen vs. nachträgliche Beschreibungen, (ii) Beschreibungen mit vorheriger Information darüber, dass eine Wegeskizze präsentiert wird vs. Beschreibungen ohne Information darüber, welche Art von Objekten auf dem Bildschirm erscheinen wird. Die Präsentationsgeschwindigkeit, die in allen 2 x 2 Versuchsbedingungen gleich war, wurde so gewählt, dass eine on-line-Beschreibung möglich und sinnvoll ist.[24]

Die Phänomene der Skizzenerstellungsbeschreibung werden im folgenden am Beispiel einer Zeichnung (Abb. 7; im weiteren häufig mit dem Namen ihres Produzenten „Soenke" benannt) bzw. deren Entstehung erläutert.

---

[23] Start und Ziel sowie der Verlauf geeigneter Routen, die eine Länge von etwa 6 km besitzen, waren allen Skizzenerstellern aufgrund eigener Erfahrung gut bekannt; die Art des in der Routenbeschreibung zu verwendenden Verkehrsmittels wurde nicht vorgegeben.

[24] Die Präsentationsgeschwindigkeit wurde gegenüber dem Skizzenerstellungsprozess erhöht. Bei Verwendung des ursprünglichen Zeitverlaufs wäre u.E. keine spontanen Sprachproduktion erfolgt.

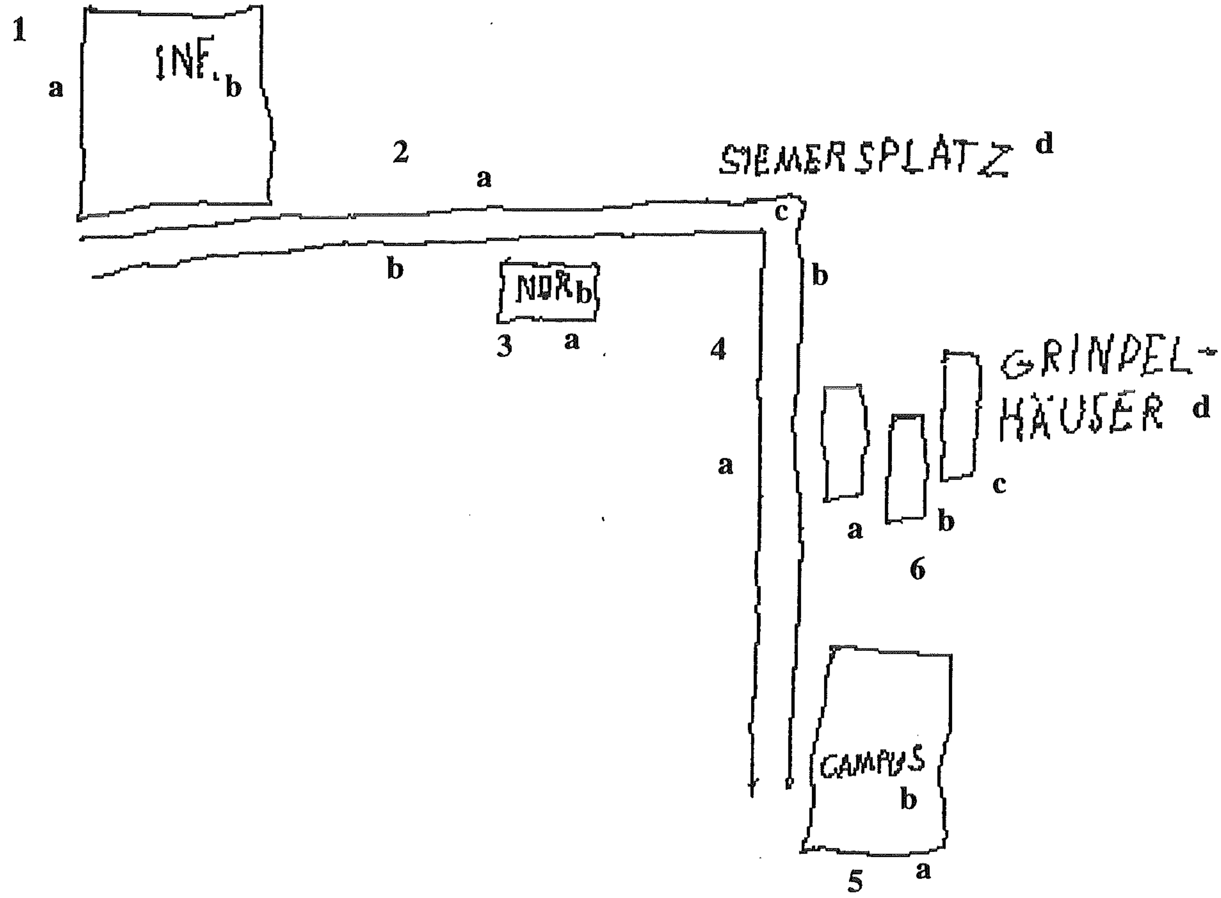


Abb. 7: Skizze „Soenke“: Vom Informatikum zum Campus[25]

Auch wenn die Präsentation der Skizzenentstehung – aus einer objektiven Perspektive – als eine monotone, im wesentlichen gleichförmig andauernde „Vermehrung von Bildpunkten“ angesehen werden kann, stellt sie sich für den Betrachter und Sprachproduzenten – subjektiv – als eine gegliederte Abfolge von Ereignissen dar. Aufgrund anderer Untersuchungen kann davon ausgegangen werden, dass die Fähigkeit den „Strom der wahrgenommenen Eindrücke“ zu segmentieren, d.h. Segmentgrenzen zu postulieren, zentral dafür ist, gewisse Segmente als Ereignisse zu individuieren (vgl. „cut hypothesis“ von Avrahami und Kareev, 1994).

Die Segmentierung ergibt sich aus der Anwendung von Segmentierungsprinzipien und -regeln; unter diesen sind einerseits solche, die den generellen Prinzipien der visuellen Wahrnehmung zuzurechnen sind, und andererseits können einige als Reflex gewisser Regularitäten der Skizzenerstellung und Skizzeninterpretation angesehen werden. Neben

[25] Annotationen (gedruckte Ziffern und Buchstaben) sind nicht Bestandteil der Skizze/Präsentation sondern betreffen die Segmentierung; vgl. die folgenden Erläuterungen und Tabelle 2, insbesondere die beiden rechten Spalten.

dieser Differenzierung in generelle, domänenunabhängige Prinzipien und domänenspezifische Regeln ist eine weitere Unterscheidung relevant, die in „gruppierende" und in „abgrenzende" Segmentierungsvorgänge, oder anders charakterisiert: in solche, die Segmentgrenzen postulieren, und solche, die für einen Abschnitt des Eingabestroms signalisieren, dass keine Segmentgrenze vorliegt.

Für die Formulierung der Segmentierungsprozesse wird davon ausgegangen, dass in der *Schicht* der Skizzenobjekte[26] eine Sortierung in *graphischen Entitäten*, wie *Linien* und *Flächen*, und *Schriftobjekte* vorliegt. Diese Sortierung der Skizzenobjekte korrespondiert zu einer Sortierung der Ereignisse in die Modalitäten *zeichnen* und *malen* einerseits,[27] und *schreiben* andererseits.

Eine wichtige, für die Formulierung von Segmentierungsregeln hochwirksame Regularität in der Skizzenerstellung ist, dass häufig Schreibereignisse als Kommentierung von Zeichenereignissen die skizzenhafte Darstellung eines Objektes der realen Welt abschliessen und insofern das Ende eines komplexen Skizzenerstellungsereignisses signalisieren. Diese Analyseregel korrespondiert zu einer Skizzenproduktionsregel

$$\text{ZSK} \;\;\text{->}\;\; \text{ZAGO}^{+} \;\; \text{KOM}^{*},$$

die hier in einer an die Standardnotation formaler Grammatiken angelehnten Weise formuliert ist.[28] So wird etwa durch die Beschriftung „INF" (1.b) die Erstellung des Skizzenobjekts, das das Informatikgelände der Universität Hamburg darstellen soll, abgeschlossen.

---

26 Neben Skizzenobjekten spielen in den Beschreibungen auch die „durch Skizzenobjekte repräsentierten Objekte der Realität" – auch als „intendierte Objekte" bezeichnet – eine Rolle. Ich verwende im weiteren die Bezeichnung *„Schicht der ...“*, um diesen beiden Typen von Entitäten zu unterscheiden, da andere – ebenfalls naheliegende Bezeichnungen – wie *Ebene* oder *Stufe* eine hierarchische Konnotation aufweisen, die insbesondere im Zusammenhang mit den hier auftretenden Ereignishierarchien zu Missverständnissen anlass geben könnten.

27 Während einige Vpn *zeichnen* und *malen* als Synonyme zu verwenden scheinen, bzw. sogar eine Präferenz für *malen* aufweisen, wird von anderen ein systematischer Unterschied zwischen den beiden Konzepten gesehen, der – informell charakterisiert – die Beziehung zwischen Linienhaftigkeit und Flächigkeit betrifft. Die eigenständige Ausarbeitung von Flächenelementen (z.B. durch Schraffur) bzw. die Verwendung von Linien unterschiedlicher Breite wird häufig als charakteristisch für den Übergang vom Zeichnen zum Malen angesehen. Für das vorliegende Material sind derartige Unterscheidungen m.E. jedoch kaum relevant, so dass im weiteren eine Trennung der beiden Konzepte nicht vorgenommen wird.

28 Hierbei steht ZSK für „Zeichnen eines Skizzenobjekts", ZAGO für „Zeichnen eines atomaren graphischen Objekts" und KOM für „Schreiben eines Kommentars". Somit ist die Regel in der folgenden Weise zu lesen: „Das Zeichnen eines Skizzenobjekts erfolgt dadurch, dass atomares graphisches Objekt oder eine Folge atomarer graphischer Objekte gezeichnet wird und anschliessend – nicht notwendigerweise – kommentierende Beschriftungen angebracht werden."
Die hier aus Platzgründen nur angedeuteten Skizzenproduktionsregeln sind Teil eines zu Rumelharts (1975) *story grammars* verwandten Konventionensystems, das ich mit *sketch grammar* bezeichne.

Eine andere wichtige Klasse von Segmentierungsregeln macht in wesentlicher Art und Weise von räumlichen Konzepten, wie *Nähe*, Gleichheit/Ähnlichkeit von *Form* oder *Ausrichtung*, Konstanz und Änderung der *Richtung, Geschlossenheit, Zusammenhang*, etc., Gebrauch. Die räumliche Fundierung der hier skizzierten Prinzipien scheint nicht nur auf den räumlichen Eigenschaften der beteiligten bzw. resultierenden Objekte zu beruhen, sondern durchaus auf den räumlichen Eigenschaften der Ereignisse. (Ähnlich hierzu ist, dass für die Charakterisierung gewisser Ereigniskonzepte wie *folgen* oder *überholen* die „Gestalt der Trajektorien" (in einem raum-zeitlichen Sinn) bestimmend ist; vgl. Habel, 1990).

Beginn (Ansetzen) und Ende (Absetzen) eines Linienerstellungsvorgangs, also Ereignisse, die zu Lücken im Resultat des Zeichnungsprozesses (der Zeichnung) führen, signalisieren das Ende eines elementaren Zeichnungsereignisses. So werden zwei individuelle Ereignisse des Zeichnens horizontaler Linien, (2.a) und (2.b), festgestellt. Andererseits ist das Ereignis (1), *Zeichnen-eines-Rechtecks,* dadurch gegenüber vier aufeinanderfolgenden *Zeichnen-einer-Strecke-* Ereignissen ausgezeichnet, dass ein stetiger Prozess des Zeichnens stattfindet, d.h. in der Schicht der Skizzenobjekte, dass jeweils das Ende einer der gezeichneten Strecken Anfang der nächsten ist. Signifikante Richtungsänderungen bei der Linienerstellung können jedoch auch als Signal für das Ende eines elementaren Zeichnungsereignisses angesehen werden; diese Regel findet in der Beschreibung anderer in der Studie untersuchter Skizzen systematische Anwendung. Neben der Stetigkeit des Zeichenprozesse – korrespondierend zum Zusammenhang der erstellten Skizzenobjekte – spielt auch die Nähe von Objekten bzw. Anfangs- und Endpunkten von Ereignissen eine wesentliche Rolle als Segmentierungshinweis: Beim Übergang von den Ereignissen (2a) und (2b) zu(3a), *zeichnen eines Rechtecks,* liegt sowohl Unähnlichkeit der Trajektorienform als auch fehlende räumliche Nähe zwischen den Endpunkten der Ereignisse aus (2) und dem Anfangspunkt des Ereignisses (3) vor.

Aufeinanderfolgende elementare Skizzenerstellungsereignisse, bei denen die Trajektorien bzw. die Resultate des Zeichnens (Linien etc.) eine „gute Gestalt" haben, werden üblicherweise einem Segment der nächst höheren Ebene der Ereignisse zugerechnet. Für die Charakterisierung von (ii) werden einerseits *Zusammenhang* und *Nähe* und andererseits verwendet. Als Kriterien für „gute Gestalt" (im Sinne der Gestaltpsychologie) sind insbesondere *Nähe, gute Fortsetzung, Geschlossenheit* und andere *nichtakzidentelle Eigenschaften* der Segmente zu nennen (vgl. Biederman, 1987, pp. 199f; Lowe 1985). So können etwa die individuellen Linienzeichnungsereignisse (2.a) und (2.b), das nacheinander-Zeichnen-von-zwei-parallelen-Linien, zu (2), Zeichnen–einer-Doppellinie, aufgrund der Parallelität der beiden Linien, die als nichtakzidentelle Beziehung angenommen werden kann, zu einem komplexen Ereignis zusammengefasst werden; zur Formähnlichkeit bei den Skizzenobjekten korrespondiert die Bewegungsähnlichkeit bei den Zeichenereignissen.[29]

Die hier skizzierten Segmentierungsprinzipien und -regeln sind die Basis dafür, dass Segmentierungsprozesse den „Strom der wahrgenommenen Eindrücke“ zerlegen, hierbei gewisse Segmente als Ereignisse zu individuieren, und somit die Grundlage für Ereignisstrukturen, gesehen als Entitäten der konzeptuellen Ebene eines kognitiven Systems, legen.

Ob das System der Prinzipien und Regeln so spezifiziert werden kann, dass es deterministisch zu genau einer Segmentierung bzw. Ereignisstruktur führt, ist zum gegenwärtigen Zeitpunkt nicht geklärt. Es ist zu erwarten, dass Parameter, die dem situativen Kontext bzw. dem Vorwissen zuzurechnen sind, die Prozesse der Regel- uns Prinzipienverwendung steuern. In diesem Sinne sind die in den Tabellen 1 und 2 dargestellten Ereignisstrukturen zur Skizze „Soenke“ Beispiele für erfolgreiche, besonders plausible Segmentierungen.

Zeit (in sek)	Ereignisse der Skizzen-erstellung	Segmentierungshinweis	Seg-ment		intendiertes Objekt
	Zeichnen einer senkrechten Linie	Richtungsänderung	1	a1	
	Zeichnen einer waagerechten Linie	Richtungsänderung		a2	
	Zeichnen einer senkrechten Linie	Richtungsänderung		a3	
	Zeichnen einer waagerechten Linie	Richtungsänderung		a4	
zusammen:					
3	Zeichnen eines Rechtecks	Gruppierung wegen Geschlossenheit	1	a	Informatik-Gelände
5	Beschriftung „INF“	Beschriftungskonvention		b	

Tabelle 1: Skizzenerstellung „Soenke“ – Detailanalyse der ersten Segmentierungsschritte

[29] Das Konzept der Bewegungsähnlichkeit enthält offensichtlich räumliche Aspekte. Dies ist jedoch nicht verwunderlich, da die Trajektorie einer Bewegung ein primär räumliches Konzept ist.

Zeit (in sek)		Ereignisse der Skizzenerstellung	Segmentierungshinweis (Hinweis auf das Ende des Ereignisses)	Segment		intendiertes Objekt
	3	Zeichnen eines Rechtecks	Abschluss durch Gruppierung (Gestalt)		1a	Informatik-Gelände
	5	Beschriftung „INF“			1b	
0–5			Beschriftungskonvention	1		Gebäudekomplex des FB Informatik
	9	Zeichnen einer waagerechten Linie	Unstetigkeit: Neuansetzen an einem „fernen Punkt“		2a	
	13	Zeichnen einer waagerechten Linie			2b	
5–13		Zeichnen einer waagerechten Doppellinie	Gruppierung wegen Parallelität;	2		Strassenzug
	14	Zeichnen eines Rechtecks	Abschluss durch Gruppierung (Gestalt)		3a	NDR (Fernseh-Studios)
	16	Beschriftung „NDR“			3b	
13–16			Beschriftungskonvention	3		Gebäudekomplex des NDR
	20	Zeichnen einer senkrechten Linie	Unstetigkeit: Neuansetzen an einem „fernen Punkt“		4a	
	22	Zeichnen einer senkrechten Linie	Unstetigkeit: Neuansetzen an einem „fernen Punkt“		4b	
		Zeichnen einer waagerechten Doppellinie	Gruppierung wegen Parallelität;			Strassenzug
	24	Zeichnen einer kurzen waagerechten Linie			4c	Kreuzung von zwei Strassenzügen
	29	Beschriftung „Siemersplatz“			4d	
16–29			Beschriftungskonvention	4		Strassenzug + Kreuzung
	33	Zeichnen eines Rechtecks	Abschluss durch Gruppierung (Gestalt)		5a	Campus der Univ. Hamburg
	36	Beschriftung „Campus“			5b	
29–36			Beschriftungskonvention	5		Campus der Univ. Hamburg
	38	Zeichnen eines Rechtecks	Abschluss durch Gruppierung (Gestalt)	6	6a	Häuserreihe
	40	Zeichnen eines Rechtecks	Abschluss durch Gruppierung (Gestalt)		6b	Häuserreihe
	41	Zeichnen eines Rechtecks	Abschluss durch Gruppierung (Gestalt)		6c	Häuserreihe
	49	Beschriftung „Grindelhäuser“			6d	
36–49			Beschriftungskonvention			Häuserkomplex

Tabelle 2: Zeitverlauf und Segmentierung der Skizzenerstellung „Soenke“

Bevor ich detaillierter auf die aus den Segmentierungen abgeleiteten Ereignisstrukturen eingehe, möchte ich darauf hinweisen, dass in der vorliegenden Untersuchung Segmentierungsprinzipien, -regeln und -prozesse unter verschiedenen Perspektiven betrachtet und verwendet werden:

- sie dienen dazu, eine „objektive Referenzsegmentierung" der zugrundeliegenden Ereignisabfolge zu produzieren, auf die dann bei der Analyse der Beschreibungsprotokolle Bezug genommen werden kann,
- sie werden als kognitive Entitäten postuliert; als Evidenz hierfür werden im weiteren Beschreibungsprotokolle daraufhin untersucht, ob und inwiefern Struktur und Inhalt der Produktion natürlichsprachlicher Schilderungen von Skizzenerstellungen durch die postulierten kognitiven Entitäten beschrieben und erklärt werden können.
- und schliesslich bilden sie den Ausgangspunkt für die maschinelle Produktion von natürlichsprachlichen Beschreibungen von Skizzenerstellungsereignissen.[30]

Durch den Segmentierungsprozess wird der Eingabestrom zu einer Ereignisfolge strukturiert, die die beiden untersten Ebenen einer Ereignishierarchie spezifizieren: einerseits werden durch separierende Segmentierungen elementare Skizzenerstellungsereignisse (terminale Ereignisse Knoten des Ereignisbaumes) individuiert, andererseits werden durch gruppierende Segmentierungen die Positionen in der Ereignisfolge gekennzeichnet, die den Übergang zwischen Knoten der darüber liegenden Ebene der Ereignishierarchie ausmachen.

Aus der in Tabelle 2.b dargestellten Segmentierung
1a, 1b | 2a, 2b | 3a, 3b | 4a, 4b, 4c, 4d | 5a, 5b | 6a, 6b, 6c, 6d
kann direkt, d.h. ohne weitere Schritte, eine flache Ereignisstruktur aufgebaut werden; siehe Abb. 8.

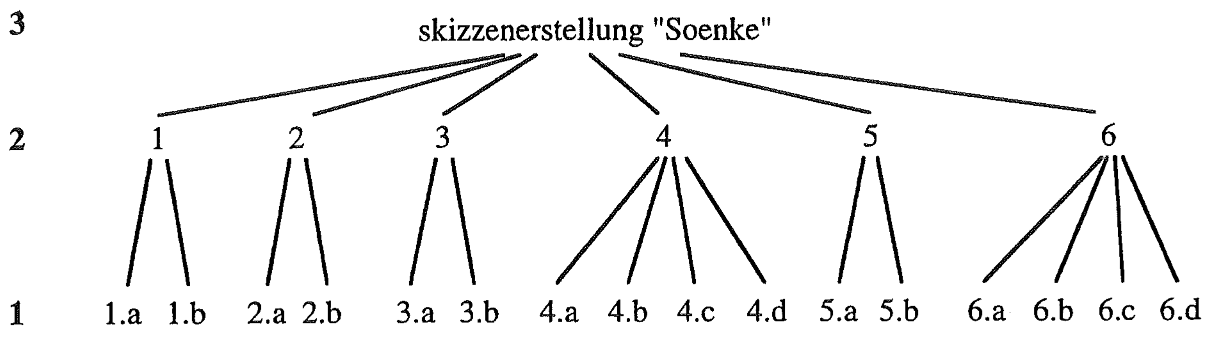


Abb. 8: Flache Ereignishierarchie zur Skizzenerstellung „Soenke"

30 Diese gegenwärtig beginnenden Arbeiten werden im Rahmen der SYNPHONICS-Konzeption eines maschinellen Sprachproduktionssystems, das an der psycholinguistischen Architektur Levelts (1989) orientiert ist, durchgeführt; vgl. hierzu Günther, Habel, Schopp und Ziesche (im Ersch.).

Ähnlich wie bei der syntaktisch-semantischen Analyse von Sätzen bzw. Texten werden im Normalfall sicherlich bei der Wahrnehmung und Konzeptualisierung einer Ereignisfolge nicht ausschliesslich flache Strukturen aufgebaut werden. Vielmehr ist davon auszugehen, dass gerade die Konstruktion von tieferen, und insofern strukturreicheren Ereignishierarchien wichtiger Teil der individuellen Konzeptualisierung, d.h des Bildung eines spezifischen Situationsmodells, ist. In Abb. 9 ist eine der möglichen Ausstrukturierungen der flachen Ereignishierarchie (aus Abb. 8) dargestellt.[31]

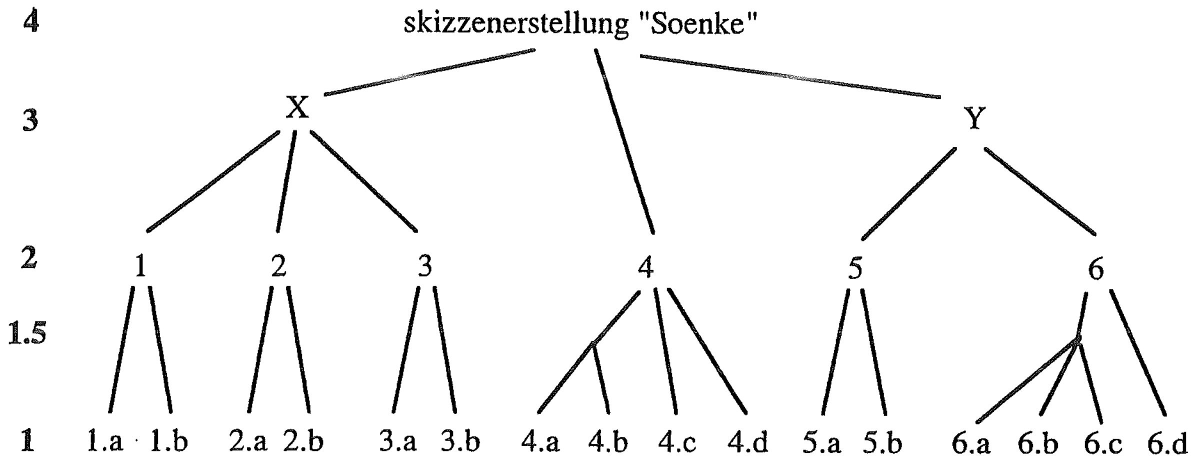


Abb. 9: Eine tiefe Ereignishierarchie zur Skizzenerstellung „Soenke"

Im folgenden möchte ich einige Befunde aus der Untersuchung exemplarisch an Hand zweier Protokolle natürlichsprachlicher Beschreibungen von Skizzenerstellungen erläutern (siehe Tabelle 3); in beiden Fällen handelt es sich um on-line-Beschreibungen, variiert ist der Parameter „vorherige Information über Skizzentyp".

Die Vorabinformation über den Skizzentyp führt dazu, dass die Vpn die für den Skizzentyp „Wegeskizze" spezifische Konventionen des Zeichnens und insofern auch des Interpretierens berücksichtigen kann. Hieraus ergibt sich, dass über die Schicht *Skizzenobjekte* hinaus auch die Schicht der *intendierten Objekte* in der Beschreibung verwendet werden kann, d.h. dass ein Übergang von graphischen Repräsentationen von Objekten der Realität zu diesen Objekten stattfinden kann. So werden etwa die Ereignisse (2.a, 2.b) im Transskript Soenke-1 (ohne Vorabinformation; Tab. 3.a) durch „...gemalt,...und da drunter eine Linie, noch eine,..." beschrieben, in Soenke-2 (mit Vorabinformation; Tab. 3.b) durch „da drunter wird 'ne Strasse gemalt,..."; d.h. die Objektpositionen von *malen* werden im ersten Fall durch Skizzenobjekte im zweiten durch Realwelt-Objekte besetzt. In beiden Fällen werden die

[31] Zu beachten ist, dass meist verschiedene Strukturanreicherungen existieren; diese Alternativen entsprechen unterschiedlichen Sichtweisen auf eine Ereignisfolge. Eine detaillierte Interpretation der in Abb. 9 dargestellten Struktur möchte ich den Lesern überlassen.

räumlichen Beziehungen zwischen den Objekten durch das in den Zeichnungen vorliegende Bezugssystem beschrieben, hier durch *drunter*.

natürlichsprachliche Schilderung	betrifft Segment
Da wird jetzt 'n Quadrat gemalt, und da steht INF drinne,	1.a, 1.b
und da drunter eine Linie, noch eine,	2.a, 2.b
---[32]	
und dann noch so 'n Dödel mit NDR drinne gemalt,	3.a, 3.b
dann geht die Linie nach unten, 'ne zweite da ran, das scheint also irgendwie 'n Weg oder sowas zu sein,	4 a,4.b
dann wird Siemersplatz geschrieben, oben an der Ecke,	4.d, 4.c
---	
dann kommt noch so 'n Rechteck, mit Campus,	5.a, 5.b
und da drüber kommen drei Rechtecke	6.a, 6.b, 6.c
---	
und da steht Grindel, Grindel Häuser.	6 d
Tabelle 3.a: Soenke-1 (on-line / ohne Information über Skizzentyp)	

Da ist 'n Gebäude und da drunter wird 'ne Strasse gemalt,	1.a, 2
in dem Gebäude steht INF,	1.b
dann kommt'n zweites Gebäude, da steht NDR drin.	3.a, 3.b

Die Strasse geht nach unten.	4 (a-b)

Über dem Knick steht Siemersplatz.	4 (c-d)

Bei der Strasse, die nach unten geht, kommt unten jetzt noch ein Gebäude	5 a
oder halt ein Platz, da steht Campus.	5.a, b

Und da drüber entstehen drei Gebäude,	6.a, 6.b, 6.c
und da steht Grindelhäuser dran.	6.d
Tabelle 3.b: Soenke-2 (on-line / mit Vorinformation über Skizzentyp)	

Wie oben erläutert, besteht die zentrale Aufgabe der Sprachproduzenten bei einer on-line-Beschreibung darin, eine hinreichend strukturreiche Ereignishierarchie aufzubauen, und diese gleichzeitig hierzu, geeignet zu verbalisieren; es handelt sich hierbei wieder um die oben erläuterte Linearisierungsaufgabe. Als Verbalisierungsebene die Traversierung des Ereignisbaums auf der Ebene der Blätter vorzunehmen würde in verschiedener Hinsicht problematisch bzw. inadäquat sein, zum einen weil der Zeitdruck der on-line-Beschreibung

32 „---" wird verwendet, um längere Pausen zu markieren.

auf dieser Ebene am grössten ist, aber auch deswegen, weil die Ereignisse der Basisebene (Ebene 1) unter Umständen keine Entsprechung in der für die Verbalisierung gewählten Schicht finden. So sind die durch 2.a und 2.b produzierten Linien in der Schicht der Skizzenobjekte wohldefiniert und auch sinnvolle eigenständige Diskursobjekte, aber erst die durch Gestaltprinzipien induzierte Doppellinie (als Resultat des komplexen Skizzenerstellungsereignisses 2) besitzt ein korrespondierendes intendiertes Objekt in der Schicht *Objekte der Realität.* Das Protokoll Soenke-2 gibt Anlass zu der Annahme, dass der Sprachproduzent eine Struktur aufgebaut hat, die in etwa der in Abb. 9 dargestellten entspricht. Die Beschreibung von (1b) erfolgt nach der Verbalisierung von (2); die Ereignisse (1) – (3), die durch räumliche Kontiguität verbunden sind, können als ein komplexes Ereignis (der Ebene 3), „zeichnen des Weges bis zur ersten Richtungsänderung" aufgefasst werden. Nach Abschluss der Beschreibung diese Ereignisses folgt eine merkliche Pause, die als Indiz für einen Ereignisübergang auf hoher Ebene gedeutet werden kann.

Eine in der Sprachproduktion besonders wichtige Entscheidung ist, ob ein Objekt durch eine definite oder eine indefinite Nominalphrase beschrieben werden soll. Die weitverbreitete Annahme, dass definite NPs überwiegend (fast ausschliesslich) dann auftreten, wenn durch sie auf ein vorhandenes Diskursobjekt Bezug genommen wird, ist, wie von Fraurud (1990) überzeugend belegt wurde, nicht gerechtfertigt; vielmehr kann davon ausgegangen werden, dass „first-mention definite NPs" (f-mDefNP) einen grossen Anteil der definiten Nominalphrasen ausmachen. Auch wenn durch eine f-mDefNP nicht auf einen vorhandenen Diskursreferenten zugegriffen wird, existiert in vielen Fällen – wie Fraurud zeigt – eine *Verankerung* an gegebenen Diskursreferenten; oder in anderen Worten: die durch f-mDefNP ausgelöste Einführung eines neuen Diskursreferenten erfolgt in Relation zu bestehenden Diskursreferenten oder zu markanten Objekten des Hintergrundwissens. Die vorliegenden Skizzenerstellungsbeschreibungen sprechen dafür, dass *räumliche Nähe* bzw. *Zusammenhang* im Situationsmodell derartige Verankerungen auslösen bzw. unterstützen. In den Verbalisierungen des Ereignisses (4) bzw. (4.a) werden f-mDefNPen verwendet, „dann geht *die* Linie nach unten" (Soenke-1), „*die* Strasse geht nach unten" (Soenke-2), bzw. „über dem Knick" (Soenke-2).

Wie die beiden hier exemplarisch vorgestellten Phänomene gezeigt haben, sind Einflüsse des Situationsmodells, bzw. der räumlichen Aspekte der internen Repräsentation, auf der Ebene der Satz- und auf der der Diskursplanung wirksam.

4 Zusammenfassung

In verschiedenen kognitiven Modellen des Sprachverstehens wird seit längerem davon ausgegangen, dass neben propositionalen Repräsentationen zur Darstellung des Textinhalts auch Sachverhaltsrepräsentationen, die einen räumlichen Charakter aufweisen können, eine zentrale Funktion im Prozess des Sprachverstehens haben. Nach einem ersten Schritt zur Präzisierung des Begriffs der „Räumlichen Repräsentation" habe ich an Hand einer Fallstudie zu Skizzenerstellungsprozessen gezeigt, dass derartige räumliche Repräsentationen auch für die Sprachproduktion von Bedeutung sind: über die Beschreibung räumlicher Eigenschaften konkreter Objekte hinaus werden räumliche Konzepte und räumliche Repräsentationen insbesondere für die Konstituierung von Ereignissen, den Aufbau von Ereignisstrukturen und die anschliessende Linearisierung dieser Strukturen bei der Sprachproduktion eingesetzt.

5 Literatur

Avrahami, J. & Kareev, Y. (1994). The emergence of events. *Cognition, 53*. 239-261.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding.*Psychological Review, 94*. 115-147.

Block, N. (ed.) (1981). *Imagery*. Cambridge, MA: MIT-Press.

Denis. M. (1991). *Image and cognition*. New York: Harvester.

Farah, M.J (1988). Is visual imagery really visual? Overlooked evidence from neuropsychology. *Psychological Review, 95*, 307-317.

Finke, R. A. (1989). *Principles in mental imagery*. Cambridge, MA.: MIT-Press.

Fraurud, K. (1990). Definiteness and the processing of NPs in natural discourse. *Journal of Semantics, 7*. 395-433.

Gärdenfors, Peter (1995). Konzeptuelle Räume. *Kognitionswissenschaft, 4*. 185–189.

Glasgow, J., Narayanan, H., & Chandrasekaran, B. (Eds). (1995). *Diagrammatic Reasoning: Cognitive and Computational Perspectives*. Cambridge, MA: MIT-Press.

Glasgow, J. & Papadias, D. (1992). Computational imagery. *Cognitive Science, 16*, 355-394.

Glenberg, A. M. & Langston, W. E. (1992). Comprehension of illustrated text: Pictures help to build mental models. *Journal of Memory and Language, 31*. 129-151.

Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language, 26,* 69-83.

Günther, C., Habel, Ch., Schopp, A., & Ziesche, S. (im Ersch.). Perspektivierungsprozesse in SYNPHONICS. In Ch. Habel, S. Kanngießer & G. Rickheit (Hrsg.) *Perspektiven der kognitiven Linguistik – Modelle und Methoden.* Wiesbaden: Westdeutscher Verlag.

Habel, Ch. (1987). Cognitive linguistics: The processing of spatial concepts. In: *T.A. Informations -ATALA (Association pur le Traitement Automatique des Langues),* 28. 21–56.

Habel, Ch. (1990). Propositional and depictorial representations of spatial knowledge: The case of *path* concepts. In: Rudi Studer (ed.): *Natural Language and Logic.* Lecture Notes in Artificial Intelligence. Berlin: Springer-Verlag. 94-117.

Habel, Ch. (1994). Discreteness, finiteness, and the structure of topological spaces. In: C. Eschenbach, Ch. Habel & B. Smith (Hrsg.): *Topological Foundations of Cognitive Science. Papers from the Workshop at the FISI-CS, Buffalo, NY. July 9–10, 1994.* (81–90). Hamburg: Graduiertenkolleg Kognitionswissenschaft Hamburg. Report 37.

Habel, Ch. & Eschenbach, C. (1995a). Abstrakte Räumlichkeit in der Kognition. *Kognitionswissenschaft, 4.* 171–176.

Habel, Ch. & Eschenbach, C. (1995b). Formale Spezifikation von räumlichen Konzepten und Strukturen mit Definitionen und axiomatischen Charakterisierungen. Projektantrag. Deutsche Forschungsgemeinschaft.

Habel, Ch., Pribbenow, S. & Simmons, G. (1995). .Partonomies and depictions: A hybrid approach. In J. Glasgow, H. Narayanan & B. Chandrasekaran (Eds.), *Diagrammatic Reasoning: Cognitive and Computational Perspectives.* (pp. 627–653). Cambridge, MA: MIT-Press.

Hanson, C. & Hirst, W. (1989). On the representation of events: A study of orientation, recall, and recognition. *Journal of Experimental Psychology: General, 118*, 136-147.

Hovy, E. H. (1993). Automated discourse generation using discourse structure relations. *Artificial Intelligence, 63*, 341–385.

Johnson-Laird, P. N. (1983). *Mental models.* Cambridge: Cambridge University Press.

Johnson-Laird, P.N. (1995). Mental models, deductive reasoning, and the brain. In Michael S. Gazzaniga (Ed.). *The Cognitive Neurosciences.* (pp. 999–1008). Cambridge, MA: MIT Press.

Kelter, S.; Kaup, B.; & Habel, Ch. (i. Vorb.). Taking the functional aspect of mental models as a starting point for studying discourse comprehension. Erscheint 1996/97 in: Ch. Habel & G. Rickheit (eds.): *Mental models in discourse processing and problem solving.*

Kosslyn, S.M. (1980). *Image and mind.* Cambridge, MA: Harvard UP.

Kosslyn, S.M. (1994). *Image and brain.* Cambridge, MA: MIT-Press.

Kosslyn, S. M., Ball, T., & Reiser, B. J. (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance, 4.* 47-60.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things.* Chicago: University of Chicago Press.

Landau, B. & Jackendoff, R. (1993). „What“ and „where“ in spatial language and spatial cognition. *Behavioral and Brain Sciences, 16,* 217–238, 255–266.

Langacker, R.W. (1982). Space grammar, analysability, and the English passive. *Language, 58,* 22–80.

Langacker, R.W. (1986). An introduction to cognitive grammar. *Cognitive Science, 10,* 1–40.

Levelt, W.J. (1982). Linearization in describing spatial networks. In S. Peters & E. Saarinen (Eds.). *Processes, beliefs, and questions.* (pp. 199-220) Dordrecht: Reidel.

Levelt, W J. (1989). *Speaking.* Cambridge, MA: MIT Press.

Lowe, David G. (1985). *Perceptual organization and visual recognition.* Boston, MA:: Kluwer.

McKoon, G. & Ratcliff, R. (1992). Inference during reading. *Psychological Review, 99,* 440-466.

Metzler, J., & Shepard, R. N. (1974). Transformational studies of the internal representation of three-dimensional objects. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium.* (pp. 147-201). Potomac, MD: Erlbaum. [shortened version as chap. 3 of Shepard, R.N. & Cooper, L.A. (1982). *Mental Images and Their Transformations.* Cambridge, MA.: MIT-Press. (pp.25-71)]

Morrow, D. G., Greenspan, S. L. & Bower, G. H. (1987). Accessibility and situation models in narrative comprehension. *Journal of Memory and Language, 26,* 165-187.

Palmer, M. S., Passonneau, R. J., Weir, C.,& Finin, T. (1993). The KERNEL text understanding system. *Artificial Intelligence, 63,* 17–68.

Palmer, S.E. (1978). Fundamental aspects of cognitive representations. In: E. Rosch & B. Lloyd (Eds.). *Cognition and Categorization* (pp. 259–303). Hillsdale, NJ.: Erlbaum.

Perrig, W. & Kintsch, W. (1985). Propositional and situational representations of text. *Journal of Memory and Language, 24,* 503-518.

Pylyshyn, Z. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review 88.* 16-45.

Rehkämper, K. (1991). *Sind mentale „Bilder" bildhaft? Eine Frage zwischen Philosophie und Wissenschaft.* Dissertation, Univ. Hamburg, FB Philosophie und Sozialwissenschaften.

Robin, F. & Denis, M. (1991). Description of perceived or imagined spatial networks. In R.H. Logie & M. Denis (Eds.). *Mental images in human cognition.* (pp. 141-152) Amsterdam: North Holland.

Rumelhart, D. (1975). Notes on a schema for stories. In D. Bobrow & A. Collins (Eds.) *Representation and understanding.* (pp. 211-236) New York: Academic Press.

Shepard, R. & Cooper, J. (1982). *Mental images and their transformations.* Cambridge, MA.: MIT-Press.

Talmy, L. (1983). How language structures space. In: H. Pick & L. Acredolo (eds.), *Spatial Orientation* (pp. 225–282). New York: Plenum.

Taylor, H. A. & Tversky, B. (1992). Descriptions and depictions of enviroments. *Memory & Cognition, 20.* 483-496.

Tversky, B., Taylor, H. A., & Mainwairing, S. (1995). Spatial perspective in language. to appear in M. Denis (Ed.). *Langage et cognition spatiale.* Paris. Masson.

Van Dijk, T. A. & Kintsch, W. (1983). *Strategies of discourse comprehension.* New York: Academic Press.

Conceptual and semantic knowledge: a psychological perspective

*Theo Herrmann & Ralf Graf**

Abstract

The terminological comparison of "conceptual" and "semantic", developed within Cognitive Linguistics has no adequate equivalent in Cognitive Psychology. Using the theoretical relationship between concepts and figures, we will discuss ways to find psychological equivalents of the dichotomy of "conceptual" and "semantic".

We refer to the meaning of figures as an association of figures with concepts. Figures may be described by underlying structural principles. However, the perception as well as the production of figures should not be understood as step-by-step cognitive testing of such structural principles. Even when figures and concepts are closely associated with one another, the context can lead to selectional restrictions in deriving a figure from a concept. This can be described and explained within a framework of common cognitive psychological theory. The assumption of a specific "semantic" type of knowledge in addition to the regular cognitive dynamics concerning the associations between concepts and figures appears superfluous.

We will discuss to what extent the views developed in this paper about the theoretical relationship between concepts and figures can be transferred to the theoretical relationship between concepts and words. In this context we will discuss the explanatory value of linguistic rules for psychological explanation. Regarding the role of linguistic rules in this matter, which would not appear to be of particular importance, some distinctions must be made. From a psychological point of view we must warn against any form of ontologisation of linguistic rules or grammar. Theoretical differentiation and construction must serve scientific problem solving processes. Both linguistics and psychology have widely divergent traditions of problem definition.

*We are grateful to Gisela Harras and Joachim Grabowski for the critical and constructive review of an earlier version of this paper. We are also grateful to Jane Gordon-Pietsch for helping to translate the original paper into this english version.

1 "Conceptual vs. semantic" – a non-psychological comparison

The term "semantic" appears in both linguistic and psychological texts, its usage in linguistics is particularly inconsistent. Besides the traditional semiotic usage (Morris, 1983), following usages dominate: (1) In the tradition of the Generative Transformational Grammar, "semantic" is confronted by term "syntactic" (for example, Lewandowski, 1990, pp. 948ff). Sentences have a syntactic and a semantic structure. Syntactic representations can be given a semantic interpretation, etc. In Schwarz, (1992, pp. 114ff) the terms "semantic-conceptual" and "formal-syntactic" are juxtaposed. (2) In the linguistic context, and this is the version we are interested in here, the term "conceptual" (or "encyclopedic") also serves as the opposite of the term "semantic" (Lewandowski, 1990, p. 951). This is how it is used in cognitive linguistic terminology. The tandem concept of "semantic vs. conceptual" was introduced by Bierwisch & Lang (1987) as part of a Two-Level-Semantic-Theory. They distinguish a semantic and a conceptual level of representation as two different modules of cognition, both distinct from the grammatical level of representation. Semantic units are tied to lexical items, and governed by the principles of the language system. The conceptual system, on the other hand, is independent of language, and serves as a frame of reference for personal experience, Schwarz, (1992, p. 49). The contradiction "semantic knowledge vs. conceptual knowledge" should also be mentioned in this context. It should be pointed out, that Jackendoff (1983), for example, rejects the comparison of semantic and conceptual structural levels in his semantic theory. Kelter has also spoken out against the seperation of a semantic and a conceptual module (Kelter, 1994).

Looking at current psychological terminology, we have to diagnose a veritable chaos, not only in the usage of the term "semantic", but also of the terms "meaning", "symbol" or "semantic knowledge". The term "semantic" is used in a completely heterogeneous, and generally non-explicit manner, when talking about the "contents" of some matter, or "something intellectual", "something meaningful", and often in contrast with "material", "perceptive" or "formal". The term "meaning" is used, as in everyday speech, very often, however its usage is even vaguer and more heterogeneous than that of the term "semantic". For example, in the representative psychology text book by Zimbardo (1995, p. 327), both words and concepts are said to have meaning. Friedhard Klix (1992) uses the term "semantic" often, however predominantly with reference to conceptual distinctions and interrelations (see. e.g. Klix, 1992, p. 268; 303ff).

A certain conceptional standardisation has been reached in the next two theoretical contexts: (1) a distinction is made between semantic networks and connectionist networks (see amongst others Helm, 1991; Rojas, 1993), which is, however, of no further interest here. (2) Semantic knowledge is compared to episodic knowledge (Tulving, 1972). Barsalou (1992), for example does not accept Tulving's use of "semantic", he would rather have the term "semantic" reserved for the meaning of language forms. He suggests that "conceptual knowledge" (!) be used in place of "semantic knowledge" (1992, p. 128). Despite these attempts at standardization, psychologists are far removed from using "semantic" and related combinations with anything like the clarity and consensus required.

Given the chaotic variety in the usage of "semantic" and related terms, the only thing which is clear is that psychologists make neither a strictly terminological nor theoretical differentiation between "conceptual knowledge" and "semantic knowledge". When such comparisons by psychologists can be found, they turn out to have been imported from cognitive linguistics, and are lacking theoretical incorporation into the inventory of general psychological theory. More frequently, a commonplace distinction can be observed, expressed in a very lenient fashion by the term "semantic vs. conceptual": People have a knowledge of words and other language entities, and they have a knowledge of the "world", insofar as this has nothing to do with language knowledge. You can know a word, and you can know objects, events or facts. It is immediately clear that this rather everyday distinction is not equivalent to the modular approach of Bierwisch & Lang (1987), or other similar linguistic distinctions, simply because it is extremely underdeveloped.

Conclusion: When dealing with theoretical comparison made by linguists of "semantic knowledge" and "conceptual knowledge" from a psychological perspective, we have no established psychological theory which could reflect the linguistic theory this comparison is based upon. In psychology neither a comparable terminological distinction, nor a corresponding theory of the standard of the linguistic "Two-Level-Semantic-Theory" is available. In our opinion, interdisciplinary incompatabilities of this kind tend not to arise by chance. The construction of theories with corresponding terminology reflects the scientific approach to problems in disciplinary research programs, theories are problem-solving tools (Herrmann, 1979). The lack of theoretical compatability, or even correspondence, between scientific disciplines is an indication that they each define problems differently (see also Herrmann, 1979; 1994a). As a consequence of this situation, we will examine, from a genuinely psychological perspective, what form a transfer of the distinction between "semantic knowledge" and "conceptual knowledge" could take for psychologists, and to what extent it could prove theoretically meaningful and useful.

2 Figures and concepts

Some linguists distinguish, as discussed, a semantic and a conceptual level from the grammatical level (cf. esp. Bierwisch & Lang, 1987). Or they confront the grammatical level with a combined conceptual-semantic system, Jackendoff (1983). Regarding the number of differentiable levels (or "modules"), we can tell a linguistic three-level-model from a linguistic two-level-model (cf. Kelter, 1994). As already stressed, these differences cannot be reflected by established theory within general psychology. In this paper we will discuss the possibility of bridging the gap between psychology and linguistics, based rather on the theoretical relationship of figures and concepts, than on the relationship of words and concepts (see Herrmann, 1994b). We have made this decision in order to provide better theoretical contour. If we had chosen, for example, the relationship between words and concepts, it would have been difficult to avoid linguistic facts infiltrating the description of psychological subject matter.

According to general psychological opinion, concepts, which combine to form various conceptual structures (propositions, semantic networks, cognitive schemes, scripts, mental models etc.) are the building-blocks of human knowledge. Concepts are a kind of result of abstraction or schematizing, which people learn to develop resp. to generate via an intensive learning process. Concepts arise out of the necessity of regulating behavior. The common conceptual factor of a group of objects is often their similar application or their similar naming. Investigation into the conditional aspects and aspects of processing of concepts is well advanced within psychology (cf. among others Kintsch, 1974; Hoffmann, 1986, 1988; Klix, 1992). – As figures we understand what used to be referred to in the period of classical psychology, as "gestalts". The result of perception, perceptional information, is represented internally, in a manner which is distinguishable in many ways from the representation of concepts. We have discussed the theoretical differentiation of figures and concepts in detail elsewhere (Graf, Herrmann, Grabowski & Schweizer, in print; Herrmann, Grabowski, Schweizer & Graf, in print).

Figures do not have to be merely of visual character, they can be auditory, tactile etc. Melodies, harmonies, beats etc. can also be included amongst the figures. The distinction we are interested in here can be easily explained with the following example: A swiss cheese and a pile of whipped cream are quite different in terms of their figure/shape, they are, however, conceptually (as dairy products) very close together. The swiss cheese has conceptually hardly anything to do with a wheel, although they have a considerable figural similarity. Flores d'Arcais and colleagues have studied the influence of figural vs. conceptual similarity in depth (with other theoretical requirements), in a series of trials (Flores d'Arcais, Schreuder & Glazenborg, 1985; Flores d'Arcais & Schreuder, 1987). For example, while a banana and a cherry are figurally (perceptually) very different, they are conceptually very closely connected to each other. A ball and a cherry, on the other hand, are figurally very similar but conceptually different. The same is true for the comparison of tennis rackets, violins and accordions; figurally fiddles and tennis rackets are more similar than fiddles and accordions, while conceptually it is the other way around.

We have developed a theory which explains the relationship between the cognitive processing of information about concepts, figures and also words with the help of an spreading activation theory after the connectionist paradigm. The implication is, seperate figure, concept and word networks are connected to one another in a specific manner. The theory defines concepts, words and figures as patterns of activation of the output layer of the network. The procedural interaction of concepts, words and figures is achieved via activity spreading from one network to another. For example, it is assumed that spontaneous deactivation is more intensive in the word and figure networks than in the concept network. (See Graf et al., in print)

Put plainly, not all figures have "meaning". Not every figure (resp. class of figures) has such a strong association for an individual (or collective group) with certain concepts, that these concepts could be co-activated via perception of the figure. Other figures are more strongly associated with concepts. For these the perception of a figure, or even a complex perceptual informational structure, "triggers" a specific concept. In this way a sketch can be identified as Cologne Cathedral, we cognitize a curved yellow form as a banana, recognize Mr. Smith at dusk, etc. In this context we can explain "meaning" as a figure-concept association – "starting at the figure".

The terms "meaning" and "designation" do not correspond to one another when indicating the relationship of concept to figure ("starting at the concept"). Words do have meaning when associated with concepts; and concepts can be designated by being associated, amongst other things, with words. Figures also have meaning, when associated with concepts. Concepts, however, are not designated through being associated with figures. Concepts "manifest themselves", "realize themselves", "portray themselves" in figures. In this respect, there is a semiotic difference between figures and words. However, there is no need to change the theoretical treatment of the association between associates purely because various semiotic relationships between these associates can be made. It is generally applicable, that it is possible, but not necessary to draw a theoretical parallel between words and figures. The decision on whether to emphasize the mutuality or the discrepancy between figures and words depends on the theoretical context. Within our psychological network model, we emphasize the mutuality between figures and words, although we also assume some (subordinate) differences in this context, which will not be further discussed here (Graf et al., in print).

At this point it should be noted, that we regard the ontogenetic development of concepts and concept structures as strongly influenced by a person's physical environment and, as such as being influenced by figures which are generated over their lifetime. Concepts fulfill, to a large extent, the purpose of cognitive organisation in the physical environment. Our mentality is not only formed by language, in the sense of a kind of watered-down linguistic realism (Schlesinger, 1986), but also by representations of our physical environment in the form of figures. To a certain extent these also play a part in forming our conceptual system. For the present we will concentrate on figures and concepts as they exist after ontogentic development has been completed.

Figures and concepts (like words and concepts) are hardly even related to one another. A figure, resp. a class of figures is predominantly associated with several concepts; accordingly a concept is associated with several figures. Apparently these associations are not all of the same strength. Principally it seems that we can create a descending hierarchy of associations with various figures for a given concept, and a descending hierarchy of associations with concepts for a given figure (class of figures). Hierarchies of this kind have been known within psychology for a long time now (cf. Deese, 1962). If we examine the associations hierarchy which combines a figure with a set of concepts for a given individual or collective group, we can define this hierarchy as the "meaning" that this figure can take for that individual or collective group. If there were encyclopedia for figures as well as for words, then the associations hierarchy established for a collective group (over a period of time) would correspond to the entry for the figure (or an important part of it). As the strength of the associations in the hierarchy continually decreases, the "Lexicographical Encyclopaedia of Figures" would have to have some criteria for making a break at a certain point of the hierarchy, so that only the relatively strong associations connecting the figure with the concept could be included. The result would be something close to a lexical meaning of the figure in question. It would inform the reader which concepts were associated (by the collective group) with a given figure. A small native tribe in the Amazon, which had come into contact with white people, had acquired a skimmer (a slotted metal cooking spoon). They didn't use it for preparing food (i.e. as a skimmer), but rather as a kind of net for catching small fish. The (fictitious) Encyclopeadia of Figures for this tribe would therefore contain a different entry from the one from this part of the world. For the native tribe there is a relatively strong association between the figure we call a skimmer and the concepts FISHING NET, THINGS-FOR-CATCHING-FISH-WITH.

The association between figures and concepts is, (as is the association between words and concepts) open to the influence of historical change. Illustration 1 shows a figure (a) which was very strongly associated with the concept SHIP 400 years ago. Today it would tend to have a much stronger association with the (subordinate) concept SAILING SHIP. Today figure (b) is more closely associated with SHIP. (We will let the reader decide which figure is most likely to be associated with the concept BOAT.)

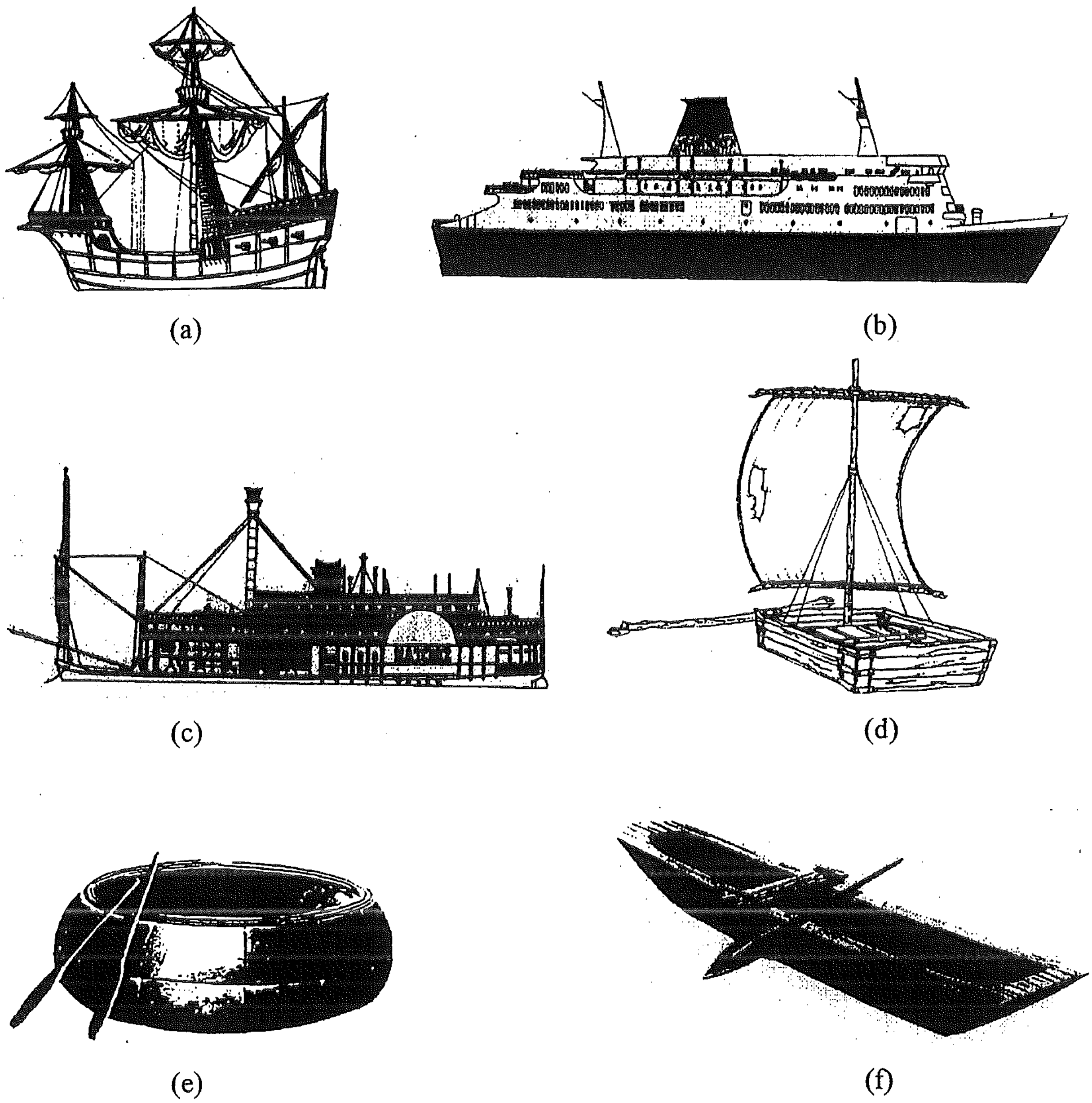


Fig.1: ships and boats (see text).

Now we can consider whether this gives us a psychological equivalent for what linguists call "semantic knowledge". Can semantic knowledge be defined as the structures of association that exist for individuals or populations between figures and concepts? In that case semantic knowledge would be the same thing as conceptual knowledge. Or is there any surplus? Before we deal with this question, we will have a look at figures, and certain restrictions governing their informational construction.

Regardless whether figures mean anything or what they mean (see above), we can say that their perceptual construction and structure are governed by certain rules, or restrictions. We can (re)construct such restrictive rules in a variety of ways and, with their help, tell well-formed figures and ill-formed figures apart.

We can regard a hexahedron, for example, as being made up of specific basic elements. These basic elements can be varied in a number of ways and these variation options follow restrictive rules. The basic elements can also be combined to form a variety of shapes (hexahedra) and for these combinations restrictions can also be identified. If the rules are "violated" then the resulting shapes are "ill-formed". (Linguists might like to consider what a set of rules of this kind has to do with grammar.)

As an illustration we will take a quick look at the structural form of hexahedra, which can be reconstructed in various ways, just as the the form of language can be reconstructed with the help of various grammars. Basic elements of hexahedra are, for example, "forks" and "arrows", (see Figure 2). This is how the Necker-Cube, for example, was reconstructed (cf. Zimmer, 1989).

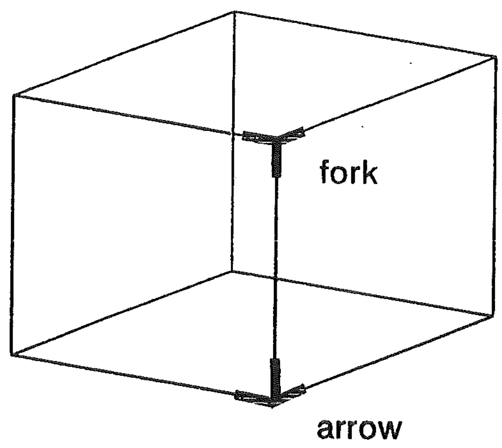


Fig. 2: Necker-Cube (forks and arrows).

Another option we used here was to define the construction of hexahedra using the side panels and combinations of side-panels as the basic elements (for a taxonomy of twodimensional figures see also Opwis & Luer, 1996). Rules or principles for the construction of figures are hypothetical constructs, which should definately not be ontologized, and which in any case, do not come into consideration as potential "real causes" for a psychological explanation of behavior. (From our psychological perspective no language grammars should be allowed to become essential; see below.)

According to the idea presented here, "normal" resp. "well-formed" hexahedra can be distinguished from "ill-formed" ones: (1) we can tell the set of acceptable basic elements for building hexahedra from the unacceptable set. Therefore, while the geometric form of the basis elements can vary, this variation has defined limits. (2) Hexahedra are built out of (a combination of) six basic elements, where subsets of the basic elements and their combinations constitute subsets of hexahedra: It follows that cubes are made up of a different subset of basic elements than cuboids, skewed cuboids or pyramid bases. The combination options of the basic elements of hexahedra are limited; there are possible combinations of (acceptable) basic elements which do not result in "well-formed" hexahedra.

We will discuss the variation aspect of the basic elements (BE) first. As basic elements for building a hexahedron we will use six convex and even squares (quadrangles) (squares which can be drawn such that all corners of each square are captured by one circle and which create areas of first order). No other geometric forms can act as basic elements. For our current purpose it is sufficient to divide the set of squares – in line with school mathematics – into the subsets quadrangles (quadr), rectangles (rect), parallelograms (paral) (rhombi are a special case in this subset), trapeziums (trape) and uneven squares (unsquar). In all cases we will refrain from defining specific lengths for the sides of the members of a sub-set – we will not categorize further by size. Furthermore, we ignore the proportions of the lines and angles within a sub-set, as long as these are not part of the mathematical definition itself (e.g. quadrangles). Therefore while the acceptable basic elements can vary, they are limited to the following set of geometric forms:

$BE \in \{quadr, rect, paral, trape, unsquar\}$.

Figure 3. shows geometric forms which are acceptable as basic elements for hexahedra ((a) – (e)) and geometric forms which are unacceptable as basic elements ((f) – (h)). The forms (f) – (h) are either not squares, not convex squares, or have not (only) straight sides.

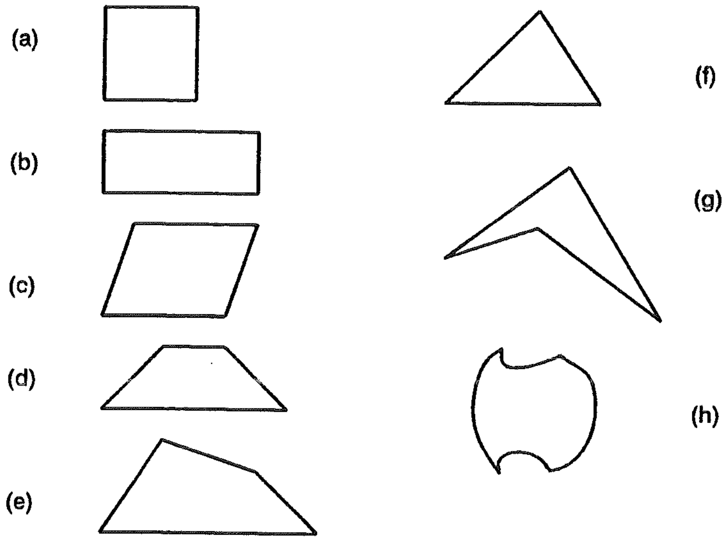


Fig. 3: Geometric forms acceptable and unacceptable as basic elements (see text).

As a first approximation, the suitable <u>combinations</u> of the basic elements for forming hexahedra can be described as the joint appearance of those basic elements. If six of the basic element *quadr* appear, we have a cube; six of the basic element *rect* produces a cuboid; if we combine four of the basic element *rect* with 2 of the basic element *quadr*, we have again a cuboid. It is impossible to construct a hexahedron by combining five of the basic element *quadr* with one basic element *unsqua*. The production of hexahedra therefore limits the free combination of basic elements. In a different theoretical context, such limitations are termed "constraints". For our purpose it is sufficient to further divide possible hexahedra (HEX) into sets of forms familiar from school mathematics – the set of cubes (cub), cuboids (cubo), prisms (pri), pyramid bases (pyb) and obelisks (obe). Obviously divergent classifications and/or additions are conceivable, however we will limit ourselves to the following:

$HEX \in \{cub, cubo, pri, pyb, obe\}$.

Figure 4 shows hexahedra correctly combined from the basic elements ((a) – (d)), an attempt to create a hexahedron with an unacceptable combination of acceptable basic elements (e), and an attempt to arrive at a hexahedron using a combination of unacceptable basic elements (f).

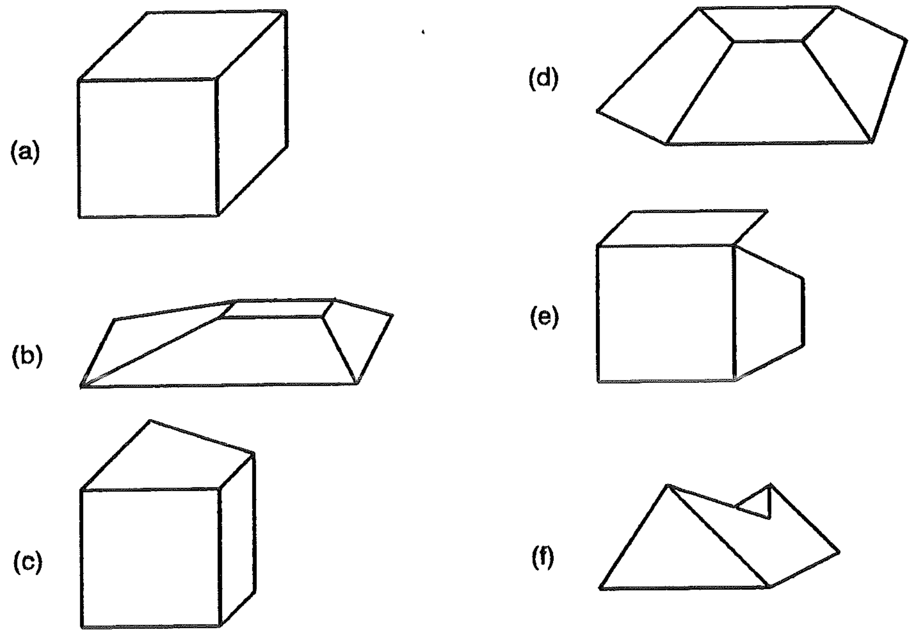


Fig. 4: Hexahedra (see text).

The acceptable variations of the basic elements, and acceptable combinations of the basic elements, can be shown formally in various ways. We will dispense with a presentation of these options to save space and limit ourselves to the formulation of a set of rules which, when followed sequentially, allows us to define a set of hexahedra.

R1: IF there is an object,

THEN check whether the object has edges.

R2: IF the object has edges,

THEN check whether the object consists exclusively of flat sides.

R3: IF the object consists exclusively of flat sides,

THEN check whether the number of sides is six.

R4: IF the number of sides is six,

THEN check whether all side panels are quadrilateral

R5: IF all side panels are quadrilateral,

THEN check whether all squares are convex.

R6: IF all squares are convex,

THEN check whether pairs of opposite sides are parallel.

R7: IF one pair of opposite sides is parallel, and two pairs of opposite sides are not parallel,

THEN check whether all non-parallel sides are rhombi.

R8: IF all non-parallel sides are rhombi,

THEN check whether lines drawn as an extension of the non-parallel sides cross at one point.

R9: IF straight lines drawn as an extension of the non-parallel sides of the rhombus cross at one point,

THEN check whether the parallel sides are rectangles.

If the IF-parts of the nine rules are given, then the resulting object is a pyramid base (pyb). If the condition in the THEN-part of rule nine can be answered affirmatively, then the object is a pyramid base with a rectangular top and bottom. To arrive at a complete determination of further sets of hexahedra (cub, cubo, pri, obe) – or a further differentiation of this set, other rules in the regulatory body come to apply.

Figure 5 shows a token of the pyramid base described above. Regarding the association of this pyramid base with a concept, it is easy to recognize the stump, after a 180° rotation within its "canonical" position in space, as the hull of one of the ships from Figure 1.

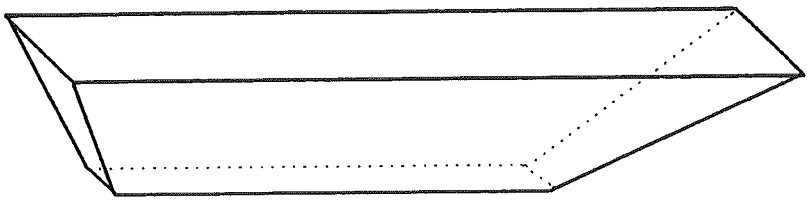


Fig. 5: Ship's (or boat's) hull.

It is of especial importance for our whole argumentation, that rules for the construction of figures say absolutely nothing about how people perceptively generate figures, i.e. how they perceive them. A glance at the psychology of perception (cf. for example, Bruce & Green, 1990; Goldstein, 1989; Marr, 1982; Marr & Hildreth, 1980; Prinz, 1983) shows, that the perception of figures, and therefore also the perception of an irregular pyramid base, is physiologically, neurologically and psychologically completely different from a step-by-step following of construction rules.

Let us suppose, that we can differentiate two psychological levels of examination: a level of available restrictions for the construction of figures, and a level of figural meaning resp. variable associations between figures and concepts. We can also conceive a comparison of this dichotomy with the linguistic dichotomy of grammatical and conceptual levels. If we disregard the profound theoretical implications (see below), it should be relatively unproblematical to set the two side-by-side. The restrictions which can be (re)constructed for the construction of figures, are something similar to the grammar of the figures. This grammar allows for a differentiation of restrictions governing variation, i.e. quasiflexion of figural elements from restrictions governing the combination of (varied) elements. This could be recognized as the differentiation between morphology and syntax. This side-by-side setting is quite superficial, and does not lead us to any substantial theoretical implications. Regarding associative networking of figures and concepts, this can easily be related to the conceptual level of linguistics. We "interpret" or "identify" perceived (or imagined) figures with the help of our conceptual knowledge.

Psychologists are unlikely to see it as an urgent task of theirs to equate the theoretical discriminability of restrictions governing the construction of figures from their associative connections with the concept level of a linguisitic two-level-approach (Kelter, 1994.) Such a comparison is hardly likely to have any important heuristic consequences. Nevertheless, the offer of a comparison of this kind could be of interest to the interdisciplinary debate, especially as our subject was, for once, not the relationship between words and concepts, but rather the relationship of a different mental type of representation (Herrmann et al., in print) with the conceptual level. Thus, we can make a theoretical attempt at generalization. As long as we don't construe "knowledge" as completely decompositional, resp. cognitively penetrable, which to our knowledge hardly any psychologist or linguist does nowadays, we can make an attempt, by regarding the figures we examined as a special mental type of representation, at distinguishing between two kinds of knowledge, and we can consider whether this distinction can be paralleled by a pertinant linguistic distinction.

On the one hand there is some knowledge, or better, a skill, which enables people to organize perceived figures into their knowledge of the world, i.e. to recognize their "meaning". This knowledge could be called "conceptual knowledge". To this knowledge we can perhaps also include the fact, that a few people (namely appropriately specialised psychologists) could also know, explicitly, which figure-concept-structures resp. associative networks or "meaning structures" exist for certain people. If there were in addition, "figurists" and figure encyclopedia, then the knowledge of collective groups accumulated by them would be something different again from the procedural "conceptual knowledge" of the man in the street, and the explicit knowledge of the psychologist. We would now be dealing with at least three varieties of "conceptual knowledge": the procedural knowledge of the man in the street, the declarative-reflexive knowledge of the psychological specialist and the quasi-lexical knowledge about collective groups.

On the other hand we can quasi-grammatically reconstruct the construction principles of figures and attempt to label them as a kind of knowledge. A question, which has to be negated, arises as to whether these (as demonstrated in our example) options give us the right to assume that the man in the street and everyday situations have something like a "quasi-grammatical knowledge of figures" of a <u>procedural kind.</u> In any case it is certain, that the actual perception of figures has nothing to do with a "quasi-grammatical knowledge of figures". It would be absurd to want to interpret the perception of a hexahedron as the application of the rules for its construction (see above).

As soon as certain parallels with the two-layer differentiation of grammatical and conceptual knowledge are presumed in the rather precarious way indicated above, the question immediately arises regarding "<u>semantic knowledge</u>". In linguistics, (Bierwisch & Lang, 1987; Bierwisch & Schreuder, 1992) "semantic knowledge" is tied to lexical items. "Semantic knowledge" is determined by the principles of the language system concerned. A semantic restriction which cannot be reduced to either a grammatical or general conceptual restriction, is illustrated by linguists with the following example (cf. Fellbaum, in print): The English verbs "to go up", "to rise" and "to ascend", all describe the same concept. Despite this they cannot be used interchangeably in every context. For example, "to ascend" is not permitted in the context of temperature. Implementing these "selective restrictions" in speech (and even explicitly knowing about these nuances of the English language) belongs to the wide area of "semantic knowledge", as far as certain areas of linguistics construe this "knowledge". Naturally, it cannot be our task to take a stand on this example, or in general on the linguistic seperation of "semantic" and "conceptual" knowledge. Still, we would like to pick up on the linguistic example given above and consider whether any parallels with the psychology of concepts and figures can be found here. (For reasons of space, we cannot deal with other kinds of paradigmatic examples of "semantic knowledge" (cf. for example, Bierwisch & Schreuder, 1992).)

The example given above concerns the choice of words when speaking, where speech is based on concepts. When a certain concept is to be verbalized, the choice of one of a selection of verbs is prohibited by "semantic knowledge". Concepts can also be represented (and communicated) as figures, for example by making a sketch. In this way someone might want to illustrate the concept of a MILLING CUTTER by a sketch. The production of figures from concepts can also take place when someone imagines a figure, or actively looks for it in the environment. When we think of the tool concept MILLING CUTTER, this might lead us to remember a singular "prototypical" milling cutter which we could imagine relatively clearly. Or we could open an appropriate reference book and look at an illustration of a milling cutter. In each of these cases previous activity on the conceptual level was the impulse for the generation of a figure. (For a network-theoretical explanation on this subject see Graf et al. in print)

The following types of (selective) restrictions for figures exist: In order to illustrate a concept with a figure, several different sketches can be made; one of them can be in some way contextually forbidden, even when the sketched figure is strongly associated with the concept. Figure 6 shows three figures (a), (b) and (c) which illustrate sketched manifestations of the concept TELEPHONE.

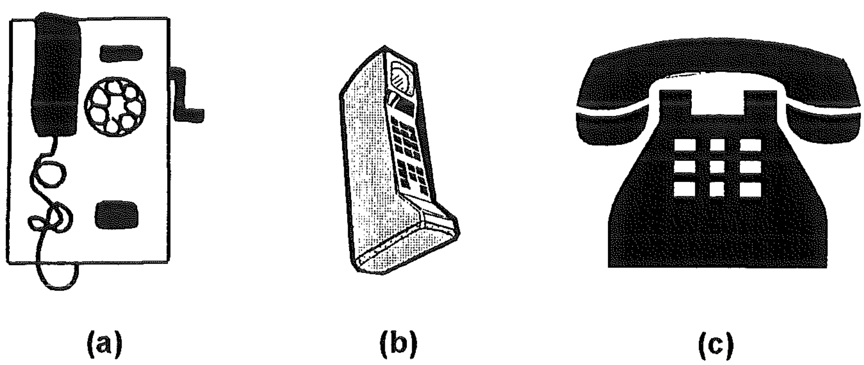


Fig. 6: TELEPHONE-figures.

All three figures follow the common construction rules for visual figures. (As an analogy, they would be grammatically well-formed.) Nevertheless, figures (a) and (b) are not valid illustrations of TELEPHONE when we have the following context: One of the figures is to be used as a pictogram in letterheads, on posters etc. to indicate telephone numbers. Figure (a) is unsuitable because its contours are much too complex for an iconic symbol and it cannot be easily identified. Figure (b) would be distinct enough regarding ist contours, however the portrayal of the telephone is not conventional or prototypical enough. Figure (c), on the other hand, is well-suited.

Generally it can be said with a degree of plausibility, that, the greater the semiotic-iconic content of figures, the more the sketches of them are governed by selective restrictions (which have nothing to do with a breach of construction rules). The concept TELEPHONE can be figurally depicted in a large variety of ways, as illustrated in figure 6. Regarding the indication of a telephone number in a letterhead, however, only figure (c) comes into consideration. Here the figural depiction of concepts is constrained by convention.

Figural representations with strong symbolic and conventional character offer closer parallels to the linguistic concept of "semantic knowledge" than everyday figures as we perceive, imagine or actively search for them. The everyday internal representation of acts of perception is only very loosely dictated by convention; and certainly does not require a three-layer-approach, rather, if at all, a two-layer one. Therefore special "semantic knowledge" hardly plays any role in the prototypical context of the psychology of perception and cognition. There is no reason for psychologists to seperate "conceptual" and "semantic" knowledge unless they are particulary interested in a psycholinguistic exchange. Contextual restrictions can be theoretically dealt with under the aspect of internally represented figures with both excitatory and inhibitatory connections with a majority of concepts and concept structures, (Graf et al., in print). It is possible that a figure cannot be generated in the imagination, although strongly associated with a specific concept, due to multifarious excitatory and inhibitatory connections in the resulting activation pattern in the concept network. For example, the generation of one of the figures in figure 6 does not come into consideration if the context requires a letterhead pictogramm. Nowadays, the figure of a sailing ship is unlikely to be generated when visualizing SHIP. In car commercials the advertised make is not illustrated by a dirty, dented car figure with kitschy paintwork, even if this is the figure most closely associated with that make. Psychology does not require a special type of knowledge for the theoretical treatment of these cases.

For now we can draw the following conclusions:

- The way figures are represented can be regarded, either under the aspect of a structural (quasi-grammatical) construction, or under the aspect of "meaning", i.e. the connection of figures with a concept level. In this context we have to ensure that the perception of figures is not interpreted as the cognitive step-by-step following of quasi-grammatical construction regulations. The theoretical conceptualization of language perception as the cognitive processing of "psychologically existent" grammatical regulations, appears to us to be extremely reduced, simplified and altogether false (Herrmann, 1995, pp.125ff., 160ff.).

- The distinction between a general "conceptual knowledge" and a specific "semantic knowledge" should be regarded as psychologically irrelevant when referring to the relationship of figures to concepts. Also regarding the theoretical assignment/classification of words and concepts, psychologists should generally be only mildly interested in the distinction between "semantic" and "conceptual" knowledge. Whenever the distinction between "semantic" and "conceptual" knowledge was examined from the inquiry perspective of Language Psychology, the justification for this distinction has proven doubtful (Kelter, 1994). For as far as psychologists are interested in the semantic selection of "to ascend" in the English language, the theoretical idea, that the selective restriction on the combination of "to ascend" with temperature, is the result of an individual learning process, is sufficient (see among others Moerk, 1985). Learning processes lead to appropriate patterns of connective weights within and between connectionist networks: Between the concept TEMPERATURE and the words "to ascend" there are – put simply – only weak excitatory, but strong inhibitatory connections. (See also the concept of "planning"-units based on Jordan (1986) in Graf et al., in print). – Assumptions of this kind appear to us to be richer in psychological content, and more easily incorporated into the body of psychological learning than the declaration of an especial "semantic" kind of knowledge. The question whether, or to what extent, "to rise" and "to ascend" are synonymous is not even a psychological one.

3 Some comments on the explanatory value of linguistic rules to psychology

The problem of the distinction between "conceptual knowledge" and "semantic knowledge" does not pose itself to psychologists as the (at least since Popper (1953) repudiated) question, whether "semantic knowledge" actually "exists" along with "conceptual knowledge". In order to achieve better theoretical contour, we have discussed here the relationship of figures to concepts rather than the relationship of words to concepts. We have come to the conclusion, that with the exception of psycholinguistic exchange projects, the hypothetical assumption of a construct called "semantic knowledge" next to a construct called "conceptual knowledge" is of little interest to psychologists due to its theoretical parsimony and lack of heuristic content. Empirical and theoretical research into the discrimination of figures and concepts and their, often complex – and inhibitatory – relationship is adequate for psychological inquiry. The other question which arose regarding the rules for figure construction, should not be confused with the question of figure perception. Reconstructionable regulations for construction cannot explain the genesis of individual figures, and it is doubtful that they comprise a "type of knowledge". – Based on an analysis of figures we have already attempted some generalisation which, in our view, should have consequences for the psychological treatment of words. We will now discuss the overall explanatory value of linguistic rules with regard to psyschological theory.

First, we must briefly make the following (highly simplified) distinction: we distinguish nomological hypotheses, algorithmic rules and deontic rules. (We refrain from a discussion of constitutive rules here (see Searle, 1969).)

General psychology has always set up nomological hypotheses ("laws"), integrated them into the structure of theory and proven the validity of these structures experimentally, or by other empirical means (Gadenne, 1984; see also Stegmueller, 1969). Nomological statements have the status of truth values; they can be empirically right or wrong. Hypothetical (law-like) 'if-then'-assumptions can be formulated and proven, for example in the research field of object perception, or the perceptual generation of figures and their association with concepts. The theories of the psychology of perception contain empirically testable hypotheses (laws). They can be applied (once sufficiently empirically proven) to explain systematically gathered (observed or measured) results. Two arbitrary examples are Emmert's Law (Emmert, 1881) and Hick's Law (Hick, 1952). Production and reception of spoken statements and their elements are also being intensively researched using nomological methodology (cf. for example, Levelt, 1989; Herrmann & Grabowski, 1994). – As will shortly become apparent, nomological laws are not in any sense the same as rules.

Besides the classical nomological methodology of general psychology, there are further means of scientific systematization. Theories and models of the human mind resp. knowledge with a cognitive-scientific orientation have also enjoyed a boom in the field of psychology. A description of human characteristics is attempted via reconstruction of human goals, convictions, and action alternatives – in a wide sense: knowledge. (cf. in.al. Newell, 1981; Tack, 1995.) When describing knowledge here, no difference is made between (objectively) right and wrong knowledge; knowledge here means that which someone believes himself to know (Tack, 1995; p.54). Application of knowledge is construed as rule-based symbol manipulation. Knowledge consists of data and rules. Data are mainly formulated in terms of propositions . A database consists of a number of such elements of knowledge. A rule is a pair consisting of a condition, represented by a model language, plus an action resp. an operation depicted in the same model language. Conditions are fulfilled by the presence of specific data. In this sense we will now discuss algorithmic rules. Algorithmic rules do not have the status of truth values and per se cannot be empirically tested, they cannot be empirically right or wrong. A rule like "If the 'number is even', add 2!" has no truth value, and as such cannot be tested.

However, whole systems of implemented data and rules can be "holistically" examined to see if, when implemented on a computer, they lead to output patterns which reflect observed human behavior (= Turing-Criterion). (Turing, 1950; cf. also Massaro & Cowan, 1993.) What human behavior actually consists of in comparison with computer output, has to be empirically established, unless it is already undeniably well-known (cf. McKay, 1988). condition-operation pairs as rules, which can be construed as elements of rule-based symbol manipulation systems, have a different theoretical and methodological function from nomological statements (see above). Moreover, algorithmic rules cannot be violated (and therefore they can also not be followed).

In psychological contexts deontic rules also appear. This kind of rule can be described with modal logic by deontic operators like 'it is obligatory that', 'it is permitted that' 'it is forbidden that' etc. (see von Kutschera, 1973; Hughes & Cresswell, 1984; Ziemba, 1976). Examples are social conventions (Geiger, 1964; but see also Hart, 1961; Lewis, 1975), which could be the subject of a psychological comparison of cultures (Herrmann & Grabowski, 1994, pp. 332ff, 441ff.) A convention of this kind in our cultural context could refer to the fact that you should give your seat to elderly people when travelling longer distances on public transport. Such deontic rules are not per se right or wrong, they can, however, be followed or violated. They say nothing about how individuals internally represent them, how individuals acquire these internal representations, and how they cognitively operate on these representations. In any case these rules, which can be violated, tell us nothing about how such a violation is to be understood as a cognitive process, nor under what circumstances these cognitive processes take place.

Within the field of Linguistics a variety of very different types of rules with varying theoretical status can be found. We have the orthographic rule, that in the German language, adjectives and participles have to be spelt with a capital letter it they are part of a geographical name (Duden rule 198). We already learned the semantic restriction which says that "to ascend", as predicate verb, cannot take "temperature" as its subject. A completely different type of rule is the rewriting rule "NP → DET + N". The first two examples can be interpreted as deontic rules; they can be violated (see also Geiger, 1964). They can also be reconstructed as algorithmic rules, and as such become part of a regulated knowledge system. The last mentioned rule "NP → DET + N" can be easily interpreted as an algorithmic rule, but not as a, breachable, deontic rule. Rules like "NP → DET + N" give information about sentences but not about individuals. If the syntactic structure of a sentence cannot be reconstructed resp. generated by means of such rules, then the sentence can be held to be ungrammatical. Obviously this has nothing to do with whether an individual is allowed to produce a statement with a particular syntactic structure or whether he violates a rule by doing so (see also Herrmann & Grabowski, 1994, pp.332ff.). Under specific additional assumptions all three classes of rules can be reconstructed as propositions with truth value, so that they could at best be used as parts of nomological explanations: "It applies to all English sentences that the verb 'to ascend' does not appear as predicate if the noun 'temperature' is the subject." This statement can be strictly tested – and is obviously false. The proposition "For all sentences applies: whenever a sentence contains an NP, then it also contains DET and N." is also empirically false, unless it is an analytical statement so that the proposition counts as not empirically testable (i.e. of no empirical content). – In order to evaluate the explanatory value of linguistic rules for psychological purposes we first have to decide what kind of linguistic rules we are dealing with and how they should be interpreted regarding their theoretical form.

If we take, for example, the rules for building plural forms in the German language, we can ascertain that the plural "Onkeln" for the singular "Onkel" is a breach of one of these rules (cf. also Duden rule 270; Engel, 1988; Grewendorf, Hamm & Sternefeld, 1987). This observation leaves the question completely open, whether and how these plural rules are internally represented in an individual who says "Onkeln", and how he or she comes to do so (Herrmann & Grabowski, 1994, pp. 38ff.).

One could develop a <u>rule-based symbol manipulation system</u> to reconstruct the internal representation of rules and the cognitive processes that lead to the output "Onkeln". The complex structure of algorithmic rules within this system are not structurally equivalent to grammar rules. The structure of algorithmic rules is a reflection, specific to the model, of grammar rules <u>and</u> the conditional compliance, and violation, of these deontic rules.

It is also possible that the internal representation of plural rules, and the process of compliance and violation, cannot be depicted with help of a rule-based system. One could build a <u>connectionist network</u> which would also produce the output "Onkeln". The required patterns of connective weights in a connectionist network are not the structural equivalents of grammar rules (see Schaub, 1995). (On connectionist modelling of linguistic rules see also Dell, 1988; Schade, 1992.).

<u>Nomological hypotheses</u> can be set up and experimentally tested on the cognitive processes which lead individuals to produce "Onkeln", as well as on the dependence of these processes on specific conditions. With sufficient empirical support for a nomological hypothesis, one could answer, let's say, a "why-question": The behavioral output in question, "Onkeln", can be deductively, nomologically <u>explained</u> under certain methodological conditions by means of the following basic explanatory scheme: "For all p and q follows: 'if p, then q', and (the explanans) p is given. <u>Therefore</u> q (the explanandum) is given." (We have to refrain here from theoretical details of explanation, especially relating to the inductive statistic explanation of events. cf. also, Stegmueller, 1969.)

Rule-based systems and connectionist networks do not serve primarily to answer why-questions. In the first place they are useful for creating variable, model-specific reconstructions of cognitive processes (cf. Thagard, 1988). We have to mention here, that there is a tendency to nomologically load both rule-based systems and connectionist networks, only then can they be falsified. The variable output of the system is modelled as systematically dependent on the variable system input (and on characteristics of the system). Whether the resulting input-output relationship is empirically true, or whether it has to be changed by appropriate "re-programming", can only be decided by empirically testing, experimentally or otherwise, the occurence of (nomological) behavior hypotheses. The cycle of (re-)modelling and (re-) simulation with repeated application of nomological-experimental methods is described as an "experimantally-simulative" method (Rickheit & Strohner, 1993).

It should have become clear, that systems of construction rules (e.g. those for figures) and general grammars cannot per se explain observed behavior. Grammars are nonetheless interpreted every so often as transpersonal ontological structures (for example. Buehler, 1934; on the history of the relevant rationalist doctrines see Eco, 1994). Accordingly, people are deeply influenced by grammar – by the "universal grammar" of language per se, and by the grammar of their "mother tongue". Now and again grammar can even be one of the causes of a person's mentality (cf. for example.,Whorf, 1956). In this respect an attempt is made to explain human behavior to a certain extent as being "caused by grammar". We think this essentialist understanding of grammar, and all similar explanatory strategies, is inadmissable according to all criteria of experience based research. To illustrate, the Broca Area in the left temporal fold of the human brain does not "contain" any grammar; rather it is the area where specific neural structures and functions are located which help to produce observed human (speech) behavior. Speech behavior, its pathogenic changes and theoretically assumed requirements, can be described with the help of (amongst other things) descriptive categories, which have mainly been developed within the field of linguistics. However, the conclusion that grammar "sits" in the Broca Area like some kind of ontological substratum influencing our behavior, makes an almost medieval impression. Linguistic rules and whole grammars are hypothetical constructs – "made by experts". With their help, defined aspects of human behavior can be more or less well described, and – if we think of the parental home or school – also formed by convention. (As is common practice in psychology, we mean by behavior observable introspective statements about "experience" and "knowledge" as well.)

If attempting not only to describe observable speech behavior with the help of grammars, but also to claim grammar to be a condition for behavior – especially as a particular kind of "knowledge" hypothetically induced into people –, so it would seem to us to be recommendable to protect oneself from ontologisation, explanatory short-circuits and pseudo-explanations. When attempting to construe any kind of hypothetical grammatical knowledge as a determinant of behavior, two things appear important: If, for example, a demonstrable grammar rule in the German language "exists", then this fact does not explain any kind of human behavior; it is inadmissable to imply that someone says "die Onkel", because (!) this reflects the construction of the plural form in the German language. And what is described as individual grammar knowledge must not necessarily be internally represented as a set of rules with which the cognitive system works. An individual can demonstrate implied grammar knowledge without applying or following any rules. Compliance with rules can also be conceptualized as learning transfer, habit or analogy construction. The same applies to rule violation. Somebody might say "Onkeln" as an anology because she also says "Runzeln" (gramatically correct plural form of "Runzel" = wrinkle, trans). (For construction of analogies see Klix, 1992, pp. 394ff. and elsewhere.) With the analogous usage of "Onkeln" she might even be following a personal rule which unfortunately doesn't correspond to the language norms laid down by Duden (cf. for example Herrmann & Grabowski, 1994, pp. 38ff. and elsewhere.)

Existing grammars, which are not all strictly transferable amongst themselves (cf. inter al. Chomsky, 1957, 1965, 1981; Weinrich, 1993; Zemb, 1978, 1984) should be seen by psychologists primarily as extremely successful normative, descriptive or hybrid <u>theories</u> and not as ontologized causes of effect. They should understand what tends to be called the "psychological reality" (or rather: "psychic reality") of grammar as a scientifically theoretically precarious quality criterion of grammar theories. The criterion refers to the variable nature of grammar theories for appropriate reconstructing the organisation of observable linguistic behavior and, at the utmost, for cognitive language related functions. Test subjects don't react the way they do in the well-known click experiments (see Fodor & Bever, 1965) because there is a Chomskyian constitutional grammar leading to this behavior; in our opinion this idea is a grotesque essentialist pseudo-explanation. The Chomskyian variety of the theory is obviously much more suitable for appropriately reconstructing the behavior of the test subjects. Under problematical scientifically theoretical conditions, one could, at the utmost, postulate that this constituent grammar – as an empirically rich proposition system! – is empirically corroborated by the behavior of the test subjects in the click-experiments. (Algorithmic or deontic rules with no empirical content cannot be empirically corroborated.)

Concerning the <u>explanatory value of grammars to psychology</u>, so can, as shown above, neither the nature of internal representations, nor the mental processes which run on these representations, nor their individual acquisition, nor the conditions under which variable cognitive processes take place, be nearly appropriately described via the presentation of (systems of) linguistic rules. There can be no mention at all of a deductive explanation of the processes and their conditions. If we reconstruct internal representations of linguistic rules and the processes of their conditional use, compliance or violation, with the help of rule-based knowledge systems, this behavior modelling is still only one of a selection of potential theoretical systemizations. So-called compliance with, or violation of, linguistic rules can also be psychologically described by the fact that the individual himself does not represent or congnitively process any rules. Rules are not parts of connectiontionist networks, these networks do not operate according to rules. The same applies generally to association, and (other) spreading activation theories (cf. for example., Goschke & Koppelberg, 1990; Helm, 1991; Mangold-Allwin, 1993; Spies, 1993) and to the majority of theoretical constructions in general psychology. For example, spread of activitation in connectionist networks can result in output which, linguistically, can be interpreted as compliance with, or violation of, linguistic rules.

Our comments can only touch the surface of the methodological and metatheoretical problems that have to be taken into account at every attempt to mediate between Linguistics and Psychology. Both disciplines have very different traditions of problem definition. This is evidenced by the fact that pychologists are primarily concerned with the problem of human behavior and psychic functions, while linguists are primarily concerned with the problems posed by the systems of individual languages or language itself (Herrmann, 1995, pp. 2ff.). Both disciplines have, as far as we can see, a different understanding of the empirical basis of science. In our opinion, a sufficiently thorough scientific appraisal of the resulting interdisciplinary mediation problems has not yet been made. We summarize our views on the explanatory value of linguistic rules for psychology as follows:

The observable language behavior B_i of an individual can be interpreted as compliance with, or violation of a linguistic deontic rule R_d. For example, the semantic selection restriction which states that "to ascend" as predicate verb cannot be combined with the subject "temperature", can be violated. Or a plural rule can be violated by the plural form "Onkeln". Or an orthographical rule can be complied with by "der Mittlere Osten" (the Middle East). Or the forementioned plural rule can be followed by using the plural form "Onkel".

These linguistic descriptions using the deontic rule R_d, do not impact on the following psychological facts:

(i) Language behavior can be conceptualized by describing and explaining language production completely without assuming internally represented rules and their application (or non-application) (see connectionist, classical learning, and psychodynamic theories). Within the frame of these theories R_d's violation can also be reconstructed. It follows that the conceptual content of psychological theory, which describes and, if need be, explains B_i, makes a rule concept irrelevant.

(ii) Certainly, it is possible to theoretically conceptualize language behavior B_i in a variety of ways which could lead to a description or explanation of language production under considerable implementation of a rule concept.

- If B_i is conceptualized with the help of a model of rule-based symbol manipulation, then, as discussed above, the assumed algorithmic rules R_a and the structure of the linguistic rules R_d are not equivalent. A violation of R_d can also be conceptualized with the help of rule-based symbol processing, whereas this violation can, at the upmost, be modelled with the help of sytems from R_a. The model-specific reconstruction of linguistic rules with the help of systems of algorithmic rules does not necessarily require the theoretical assumption that we are dealing here with the construction of a veritable empirical "Brain" or "Mind" Theory. Therefore, the assumption that individual neural or cognitive processes have the status of rule-based symbol manipulation is not a prerequisite of modelling, even if they are portrayed in this way by the model.
- The materialization of behavior B_i can be conceptualized as follows: the system which produces language has an internal representation of the rules R_d, and operates mentally on this internal representation. This can be construed, inter alia, as follows: the rules R_d are not "explicitly" or declaratively available to the psychophysical system, it doesn't "know" R_d. In this case the system applies the rules R_d, "implicitly", or better put, procedurally (cf. also Perrig, Wippich & Perrig-Chiello, 1993).
- The rule R_d may be declaratively, "explicitly" available to the system as an internal representation which produces B_i. It may operate, so-to-speak, consciously and purposefully on the representation of R_d. In that case the internal rule representative, or R_d, may exist in a different (cognitive) code than the linguistic rule R_d. The internal representative and that which is represented R_d do not have to be structurally equivalent. In this case the cognitive process of violation (or compliance) must, in addition to the assumption of an internal representative of R_d, also be theoretically described and explained, for which purpose, on the other hand, resorting to R_d is not sufficient.

References

Barsalou, L.W. (1992). *Cognitive psychology: An overview for cognitive scientists*. Hillsdale NJ: Lawrence Erlbaum Associates.

Bierwisch, M. & Lang, E. (Hrsg.). (1987). *Grammatikalische und konzeptuelle Aspekte von Dimensionsadjektiven*. Berlin: Akademie-Verlag.

Bierwisch, M. & Schreuder, R. (1992). From concepts to lexical items. *Cognition*, *42*, 23-60.

Bruce, V. & Green, P.R. (1990). *Visual Perception*. Hillsdale NJ: Lawrence Erlbaum Associa-tes.

Buehler, K. (1934). *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Jena: Fischer. (Ungekuerzter Nachdruck 1982, Stuttgart: Fischer.)

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge MA: MIT Press.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

Deese, J. (1962). On the structure of associative meaning. *Psychological Review, 69*, 161-175.

Dell, G.S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language, 27*, 124-142.

Eco, U. (1994). *Die Suche nach der vollkommenen Sprache*. Muenchen: Beck.

Emmert, E. (1881). Groeßenverhaeltnisse der Nachbilder. *Klinische Monatsblaetter fuer Augenheilkunde, 19*, 443-450.

Engel, U. (1988). *Deutsche Grammatik* (2. Aufl.). Heidelberg: J. Groos.

Fellbaum, Ch. (in print). WordNet: Ein semantisches Netz als Bedeutungstheorie. Erscheint in J. Grabowski, G. Harras & Th. Herrmann (Hrsg.), *Bedeutung–Konzepte–Bedeutungskonzepte*. Opladen: Westdeutscher Verlag.

Flores d'Arcais, G.B. & Schreuder, R. (1987). Semantic activation during object naming. *Psychological Research, 49*, 153-159.

Flores d'Arcais, G.B., Schreuder, R. & Glazenborg, G. (1985). Semantic activation during recognition of referential words. *Psychological Research, 47*, 39-49.

Fodor, J.A. & Bever, T.G. (1965). The psychological reactions of the linguistic segments. *Journal of Verbal Learning and Verbal Behavior, 4*, 414-420.

Gadenne, V. (1984). *Theorie und Erfahrung in der psychologischen Forschung*. Tuebingen: Mohr.

Geiger, Th. (1964). *Vorstudien zu einer Soziologie des Rechts*. Neuwied: Luchterhand.

Goldstein, E.B. (1989). *Sensation and perception* (3. Aufl.). Pacific Grove CA: Brooks/Cole Publishing Company.

Goschke, Th. & Koppelberg, D. (1990). Connectionist representation, semantic compositionality, and the instability of concept structure. *Psychological Research, 52*, 253-270.

Graf, R., Herrmann, Th., Grabowski, J. & Schweizer, K. (in print). Konzepte, Woerter und Figuren: Zur Aktivationstheorie multimodaler Repraesentate. Teil II: Grundriß eines Modells der Aktivierung von Konzepten, Woertern und Figuren. Erscheint in J. Grabowski, G. Harras & Th. Herrmann (Hrsg.), *Bedeutung–Konzepte–Bedeutungskonzepte*. Opladen: Westdeutscher Verlag.

Grewendorf, G., Hamm, F. & Sternefeld, W. (1987). *Sprachliches Wissen. Eine Einfuehrung in moderne Theorien der grammatischen Beschreibung*. Frankfurt a.M.: Suhrkamp.

Hart, H.L.A. (1961). *The concept of law*. Oxford: Clarendon Press.

Helm, G. (1991). *Symbolische und konnektionistische Modelle der menschlichen Informationsverarbeitung. Eine kritische Gegenueberstellung*. Berlin: Springer,

Herrmann, Th. (1979). *Psychologie als Problem. Herausforderungen der psychologischen Wissenschaft*. Stuttgart: Klett-Cotta.

Herrmann, Th. (1994a). Forschungsprogramme. In Th. Herrmann & W. Tack (Hrsg.), *Enzyklopaedie der Psychologie. Themenbereich B, Serie I, Bd. 1: Methodologische Grundlagen der Psychologie* (p. 251-294). Goettingen: Hogrefe.

Herrmann, Th. (1994b). Psychologie ohne "Bedeutung"? Zur Wort-Konzept-Relation in der Psychologie. *Sprache & Kognition, 3*, 126-137.

Herrmann, Th. (1995). *Allgemeine Sprachpsychologie: Grundlagen und Probleme* (2. Aufl.). Weinheim: Beltz, Psychologie Verlags Union.

Herrmann, Th. & Grabowski, J. (1994). *Sprechen*. Psychologie der Sprachproduktion. Heidelberg: Spektrum Akademischer Verlag.

Herrmann, Th., Grabowski, J., Schweizer, K. & Graf, R. (in print). Konzepte, Woerter und Figuren: Zur Aktivationstheorie multimodaler Repraesentate. Teil I: Repraesentationstypen. Erscheint in J. Grabowski, G. Harras & Th. Herrmann (Hrsg.), *Bedeutung–Konzepte–Bedeutungskonzepte*. Opladen: Westdeutscher Verlag.

Hick, W.E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology, 4*, 11-26.

Hoffmann, J. (1986). *Die Welt der Begriffe*. Berlin: VEB Deutscher Verlag der Wissenschaften.

Hoffmann, J. (1988). Wird Wissen in Begriffen repraesentiert? *Sprache & Kognition, 7*, 193-204.

Hughes, G. E. & Cressell, M. J. (1984). *A Companion to Modal Logic*. London: Methuen.

Jackendoff, R. S. (1983). *Semantics and Cognition* (Current Studies in Linguistics Series, 8). Cambridge: MIT Press.

Jordan, M. I. (1986). *Serial order: A parallel distributed processing approach* (Technical Report, No. 8604). San Diego: University of California, Institute for Cognitve Science.

Kelter, St. (1994). Kognitive Semantik und Aphasieforschung. Zur Unterscheidung zwischen Bedeutungen und Konzepten. In M. Schwarz (Hrsg.), *Kognitive Semantik/Cognitive Semantics. Ergebnisse, Probleme, Perspektiven* (p. 81-95) (Tuebinger Beitraege zur Linguistik, 395). Tuebingen: Narr.

Kintsch, W. (1974). *The representation of meaning in memory*. Hillsdale: Erlbaum.

Klix, F. (1992). *Die Natur des Verstandes*. Goettingen: Hogrefe.

Kutschera, F. von (1973). *Einfuehrung in die Logik der Normen, Werte und Entscheidungen*. Freiburg: Alber.

Levelt, W.J.M. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.

Lewandowski, Th. (1990). *Linguistisches Woerterbuch* (5., ueberarbeitete Aufl.). Heidelberg: Quelle & Meyer.

Lewis, D. (1975). *Konventionen. Eine sprachpsychologische Abhandlung*. Berlin: de Gruyter.

Mangold-Allwinn, R. (1993). *Flexible Konzepte. Experimente, Modelle, Simulationen*. Frankfurt a. M.: Lang.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: Freeman.

Marr, D. & Hildreth. E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London (Series B). 207*, 187-217.

Massaro, D.W. & Cowan, N. (1993). Information processing models: Microscopes of the mind. *Annual Review of Psychology, 44*, 383-425.

McKay, D.G. (1988). Under what conditions can theoretical psychology survive and prosper? Integrating the rational and empirical epistemologies. *Psychological Review, 95*, 559-565.

Metzger, W. (1954). *Psychologie*. Darmstadt: Steinkopff.

Metzger, W. (1975). *Gesetze des Sehens* (3. Auflage.). Darmstadt: Steinkopff.

Moerk, E.L. (1985). A differential interactive analysis of language teaching and learning. *Discourse Processes, 8*, 113-142.

Morris, Ch. W. (1938). Foundations of the theory of signs. *International Encyclopedia of Unified Science, 2*, No. 2.

Newell, A. (1981). The knowledge level. *AI Magazine, Summer 1981*, 1-33.

Opwis, K. & Luer, G. (1996). Modelle der Repraesentation von Wissen. In D. Albert & K.-H. Stapf (Hrsg), *Enzyklopaedie der Psychologie. Themenbereich C, Serie II, Bd. 4: Gedaechtnis* (p. 337-431). Goettingen: Hogrefe.

Perrig, W., Wippich, W. & Perrig-Chiello, P. (1993). *Unbewußte Informationsverarbeitung*. Bern: Huber.

Popper, K.R. (1935; [3]1969). *Logik der Forschung*. Tuebingen: Mohr.

Prinz, W. (1983). *Wahrnehmung und Taetigkeitssteuerung*. Berlin: Springer.

Rickheit, G. & Strohner, H. (1993). *Grundlagen der kognitiven Sprachverarbeitung*. Tuebingen: Francke.

Rojas, R. (1993). *Theorie der Neuronalen Netze. Eine systematische Einfuehrung*. Berlin: Springer.

Schade, U. (1992). *Konnektionismus: Zur Modellierung der Sprachproduktion*. Opladen: Westdeutscher Verlag.

Schaub, H. (1995). *Selbstorganisation in konnektionistischen und hybriden Modellen von Wahrnehmung und Handeln*. Memorandum Nr. 15. Lehrstuhl Psychologie II. Universitaet Bamberg.

Schlesinger, I. M. (1986). *Do we really think in words?* Working paper No. 18. The Goldie Rotman Center For Cognitive Science. School of Education. The Hebrew University of Jerusalem.

Schwarz, M. (1992). *Kognitive Semantiktheorie und neuropsychologische Realitaet. Repraesentionale und prozedurale Aspekte der semantischen Kompetenz*. Tuebingen: Niemeyer.

Searle, J.R. (1969). *Speech acts*. Cambridge: Cambridge University Press.

Spies, M. (1993). *Unsicheres Wissen: Wahrscheinlichkeit, Fuzzy-Logik, Neuronale Netze und menschliches Denken*. Heidelberg: Spektrum Akademischer Verlag.

Stegmueller, W. (1969). *Wissenschaftliche Erklaerung und Begruendung*. (Probleme und Resultate der Wissenschaftstheorie und Analytischen Philosophie. Band I) Berlin, Heidelberg: Springer.

Tack, W. (1995). Repraesentation menschlichen Wissens. In D. Doerner & E. van der Meer (Hrsg.), *Das Gedaechtnis* (p. 53-74). Goettingen: Hogrefe.

Thagard, P. (1988). *Computational Philosophy of Science*. Cambridge, MA.: MIT Press.

Tulving, E. (1972). Episodic and Semantic Memory. In E. Tulving & W. Donaldson (Hrsg.), *Organization of Memory* (p. 381-403). New York: Academic Press.

Turing, A.M. (1950). Computing machinery and intelligence. *Mind*, *59*, 433-460.

Weinrich, H. (1993). *Textgrammatik der deutschen Sprache*. Mannheim: Dudenverlag.

Whorf, B.L. (1956). *Language, Thought and Reality*. Selected Writings. Cambridge: Technology Press of Massachusetts Institute of Technology.

Zemb, J.-M. (1978). *Vergleichende Grammatik Franzoesisch-Deutsch. Band 1. Comparision de 2 système*. Bibliographisches Institut Mannheim: Dudenverlag.

Zemb, J.-M. (1984). *Vergleichende Grammatik Franzoesisch-Deutsch. Band 2. L'économie de la langue et la jeu de parole*. Bibliographisches Institut Mannheim: Dudenverlag.

Ziemba, Z. (1976). Deontic Logic. In Z. Ziembinski, *Practical Logic* (p. 360 – 430). Dordrecht: D. Reidel.

Zimbardo, Ph.G. (1995). *Psychologie* (6. Aufl.). Berlin/Heidelberg: Springer

Zimmer, A.C. (1989). Gestaltpsychologische Texte – Lektuere fuer eine aktuelle Psychologie. *Gestalt Theory*, *2*, 96-121.

Public Information & Mutual Error(*)

Wolfgang Heydrich
Universität Hamburg/Universität Bielefeld,
SFB 360 "Situierte Künstliche Kommunikatoren",

Hannes Rieser
Universität Bielefeld, SFB 360 "Situierte Künstliche Kommunikatoren"

(*) This manuscript contains two chapters of the more comprehensive paper Heydrich and Rieser (1995).

0 Introduction

The general aim of this paper is to discuss problems of the interrelation between action, communication and belief. More specifically, we are trying to develop empirically adequate descriptions of cooperating agents - descriptions, which can be used to explain, how convictions and beliefs, as well as the dynamics of their constant revision both affect and direct verbal behavior and joint action.

In order to study this empirically, we use the following setting: Instructor (Inst) and constructor (Const) are separated by a screen. Instructor has a block's assembly situated before him which he built up just before.

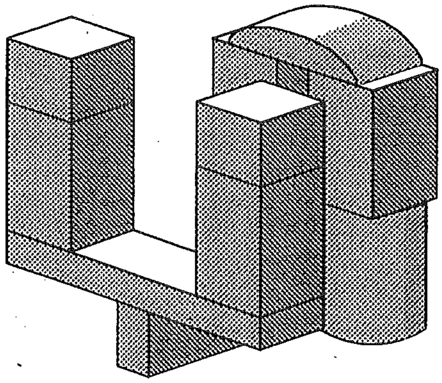


Fig. 1 Instructors block's world

He was instructed to describe it to the constructor. Const in turn was told to build up the same assembly according to Inst's directives. Among the blocks are cubes, rectangular blocks (cuboids), triangular blocks, cylinders and segments of discs of different colour and size. Inst and Const are allowed to freely converse during the task.

Basically, Inst and Const have to solve a problem of cooperation. This problem, however, involves matters of coordination as well. E.g., Inst and Const have to gain awareness and common understanding of kind, number, and position of the blocks they are dealing with. They are already in need of this kind of understanding (a shared frame of reference, a common situation) in order to accomplish such elementary tasks as making transparent to each other which blocks they are talking about and which properties and relations they are attributing to them. The blocks Inst and Const have at their disposal are located on their respective sides of the screen. Inst's blocks (i.e. those which are perceptually accessible to him) are not directly accessible to Const, and Const's blocks are not directly accessible to Inst. In this sense, information about blocks is private in our scenario - at least initially, i.e., as long as only direct perception is concerned. In order to be able to refer to and describe the blocks in a transparent way, private information has to be made - in some way or other - intersubjectively accessible or public. To achieve this, verbal communication suggests itself as the most appropriate means in the given situation.

In this paper, we are dealing with some of the more fundamental aspects of the process of

transforming private (e.g. only perceptually given) information into public (e.g. linguistically mediated) information. We will concentrate on some central features of this process connected with the opposition of *private* and *public* information.
Additional features will be discussed under this perspective, namely

(a) The make-up of discourse patterns and the notion of expectability of moves in an interaction (and, more specifically, the notion of expectability of actions in a conversation).

"Discourse patterns" and "expectability" are important notions for an adequate understanding of the required complexity of agent's beliefs and convictions in the succession of moves in an interaction. Effectively, *only finitely many* and *finitely complex* convictions must be sufficient, since it is hard to see, how empirical agents could possibly sustain beliefs of an arbitrary (let alone infinite) complexity or number. We assume that, basically, it is due to the role of entrenched patterns and established expectations in discourse that a finite number of finitely complex beliefs is sufficient. In this paper we mainly concentrate on the internal structure of the discourse pattern of giving directives, which is of central importance in construction dialogues. Directives, it will be shown, are typically processed as interactive procedures in order to carry out plans. They exhibit a dynamic structure of cycles and stages.

(b) The notion of ratification of actions as a means of establishing consent among discourse participants.

Ratification in discourse consists in turns which are used to affirm or reaffirm information and signal mutual acceptance or agreement among discourse participants. This is a central procedure in discourse constitution, and it is an important one for our topic. Ratification is a means to tag information as mutually accepted and henceforth intersubjectively available. Although we will not go into the micro-analysis of ratifications in this paper, we will emphasise its important role for the dynamics of cycles and stages in the interactive pattern of giving directives.
There is one important aspect of our topic we just point at, but are not going to discuss in detail, namely

(c) The origin of obligations in verbal interaction.

When dealing with the construction dialogues of our sample, we are confronted with a variety of speech-acts, many of them indirect. Typically, Inst's task is to ask, request or order. Typically, Const's task is to accomplish requests and to perform whatever he is asked to do. (This, however, is an oversimplification. There are many passages in our transcripts, where Inst and Const "change rôles" such that Const takes the initiative and Inst reacts.) Clearly, the general distribution of rôles is already implied by the definition of the scenario and the overall purpose of the conversation (the construction task). Given this background, Inst may establish obligations indirectly. E.g., instead of explicitly asking Const to arrange some blocks in a specific manner, Inst may reach the same effect by simply describing the configuration of blocks he intends to be realised. It is well-known that there are more or less general mechanisms involved in procedures like those of using descriptions in the place of requests. We do not have to say much about these mechanisms here, and we leave the topic to the theory of indirect speech-acts.

1 The rise of mutual belief

1.1 Intersubjectively transparent information

Given the separation of the perceptual spaces of Inst and Const in our scenario, we ask: How is it possible for the agents to establish and secure a common situation as a prerequisite for communication and joint action?

It seems obvious that Inst and Const are in need of informational access to the blocks on each side of the screen. The agents have to be aware of at least some of their relevant properties and configurations, in order to be able to characterise them and describe both their positions and the pertinent relations among them. But for each of the two agents it is not enough simply to have informational access to the blocks involved. In order to communicate about the blocks and cooperate with each other, Inst's and Const's access must not be private, it has to be public, i.e. it must be informationally transparent for both participants. Inst and Const do not only need informational access to the blocks, they also need informational access to each others having informational access to the blocks, and - furthermore - they need informational access to each others having informational access to each others having informational access to the blocks, and - furthermore - etc., etc.

In this sense, the concept of mutuality is essential for the notion of intersubjectively transparent information. In order to explain what makes information public, it seems that we need an infinite hierarchy of nested accessibilities.

There is a difficulty here, however. If, on one hand, the notion of *informational access* is construed *psychologically* as *actually having a certain (true) belief*, then this leads to the well-known difficulties of attributing both infinitely many and arbitrary complex beliefs to agents whose mental processing capacity is restricted in all relevant respects. These difficulties have been called "mutual knowledge paradox" by Clark and Marshall (1981).

If, on the other hand, the notion of *informational access* is construed merely *logically*, such that nothing is implied with respect to mental processing capacities of empirical agents, surely no problems arise with respect to questions of complexity and infinity. By this account, however, it becomes quite obscure how public (i.e. transparently accessible) information can possibly influence empirical agents causally.

1.2 Barwise's distinctions and Lewis' connection

Jon Barwise (1989) has stressed the contrast between *psychological* and *logical* explications of epistemic notions of mutuality (*mutual knowledge* or *mutual belief*). According to him, psychological explications aim at the explicandum of agents *having* (i.e. actually disposing of) information, whereas logical explications are concerned with the explicandum of things or situations merely *carrying* (or *supporting*) information. *Carrying* information is a relation closed under logical consequence, *having* information is not. In contradistinction, *having* information is causally relevant for explaining action, *carrying* information is not. The notion of *having information* concerns the psychological theory of knowledge, *carrying information* concerns objective and subject-independent information *proper*. Explicating the notion of *informational access* in terms of *having* information puts it into the framework of psychological epistemology. This account must be distinguished, according to Barwise, from explaning it in terms of logical epistemology (e.g. as some kind of acquaintance with situations (or - more generally - significant objects) supporting (or carrying) information). For Barwise, only the logical notion is worth formal investigation. He writes: "Information travels at the speed of logic, genuine knowledge only travels at the speed of cognition and inference.[...] I think that a theory of information should be part of logic, but I doubt that the theory of knowledge really is" (p. 204).

Barwise is interested in a comparison of different explications (or model-theoretic reconstructions) of mutually accessible information (in the sense of objective information *carried* by situations). In contrast to Barwise (1989), Lewis (1969) is interested in connections between explications of informational access and mutuality in the framework of *logical* vs. the framework of *psychological* epistemology.

We stressed above that the aim of our investigation is to contribute to the description of the underlying regularities of communication and action among empirical persons (here: Inst and Const) in the concrete setting of construction dialogues. We want to explain, in what ways the behavior of our agents depends on - and influences - their respective as well as their common convictions and beliefs. It is clear that a merely logical account of mutuality and transparency of information, which has next to nothing to say about the causal rôles of convictions and beliefs for action and communication, is not of much value for this endeavour.

What is needed instead, is a conception which links a (logically founded) account of intersubjective transparency of public and mutually shared information to a (psychologically relevant) account of the dynamics of belief development and revision among empirical agents. This is why Lewis' integrative appoach is interesting for us. On the one hand, he wants to explain, how it is possible for the members of a group to gain common and mutually transparent convictions without actually processing an infinity of beliefs. On the other hand, he wants to explain which empirical factors constrain this possibility in principle.

We start with Lewis' account (LA) of public information[1].

According to (LA), the information that p is public in group G (in symbols: P_Gp) iff there is a situation s such that the following three conditions obtain:

(LA)	(i)	$\forall x \in G\ R_xHs$
	(ii)	$\forall x, y \in G\ s \models_x R_yHs$
	(iii)	$\forall x \in G\ s \models_x p$

(We write "R_x..." for "x has reason to believe that...", "Hs" for "s holds", and "$s \models_x$..." for "s indicates to x that...".) If for some group G, situation s, and information that p the conditions (i) - (iii) of (LA) obtain, we write LA[G, s, p]. In this case, (i) is the condition of *accessibility* of s, (ii) the condition of *perspicuity* of s, and - in specifying the informational content p - (iii) is the condition of *significance* of s for the members of G. Lewis acknowledges that people will normally differ with respect to background information and inductive standards. Thus, a state of affairs may possibly fail to indicate *to me* that it indicates to everyone whatever it indicates to me. With respect to the justifiability of higher-order expectations, however, Lewis feels licenced to abstract from these inequalities. He

[1] Lewis' own notion is "common knowledge in a population P", the formal notation is ours.

assumes an "ancillary premiss", which might be explicitly formulated like this:

(*) $\forall x, y \in G\ (s \vDash_x p \supset s \vDash_x s \vDash_y p)$.

Additionally, we follow Lewis in eliminating the basic notion of *indication* by reducing it to the two remaining basic notions of *having reason to believe* and *holding* (of states of affairs):

Def.1 $s \vDash_x p := R_x Hs \supset R_x p$.

Lewis uses a counterfactual conditional in his definition[2]. We have simplified this by using material implication instead[3]. Finally, we assume that *reason to believe* satisfies some straightforward logical principles:

(a) $R_x(p \supset q) \supset (R_x p \supset R_x q)$.
(b) $\forall x \in G\ R_y p \supset R_y \forall x \in G\ p$
(c) $(R_x q \wedge R_x p) \supset R_x(q \wedge p)$
(d) $\exists x\ R_y p \supset R_y \exists x\ p$.

Under these assumptions it can be shown[4] that (1) entails (2) (i), (ii), etc., and even (3):

(1) $P_{\{Inst, Const\}} p$
(2) (i) $\forall x \in \{Inst, Const\}\ R_x p$
(ii) $\forall x, y \in \{Inst, Const\}\ R_x R_y p$
(iii) $\forall x, y \in \{Inst, Const\}\ R_x R_y R_x p$
etc.
(3) $P_{\{Inst, Const\}} P_{\{Inst, Const\}} p$

---

[2] "Let us say that A indicates to someone x that... if and only if, if x had reason to believe that A held, x would thereby have reason to believe that..." (Lewis 1969, p. 52). For Lewis' account of counterfactuals cf. Lewis (1973).

[3] This simplification, however, is far from unproblematic: It entails that *every situation* s indicates p to x, provided x has reason to believe that p. This is because $p \supset q$ (e.g. $R_x Hs \supset R_x p$, hence $s \vDash_x p$) is true, whenever q (e.g. $R_x p$) is. To prevent this, we might prefer to use (instead of '$\supset$') an implicative connective '$\rightarrow$', which does not licence the rule of *conditional weakening* (from q to infer $p \rightarrow q$), but still is in accordance with the rules of *detachment* (from p and $p \rightarrow q$ to infer q), *transitivity* (from $p \rightarrow q$ and $q \rightarrow r$ to infer $p \rightarrow r$) and *self-distribution* (from $p \rightarrow (q \rightarrow r)$ to infer $(p \rightarrow q) \rightarrow (p \rightarrow r)$). Hence, some form of strict - or strong - implication (from some modal - or relevant - logic) might do the job. Note, however, that Lewis' own counterfactual conditional (cf. Lewis 1973) is not appropriate, since it does not license transitivity. Using material implication, we get a notion of indication which is rather (perhaps *too*) weak. Note that in the light of Def. 1 we could simplify (*) to

(**) $\forall x,y \in G\ (s \vDash_x p \supset R_x s \vDash_y p)$

which attributes to each member x in G reason to believe that s indicates p to everyone in G if only it indicates p to x himself. For given Def.1 and accessibility of s in G (i.e. $\forall x \in G\ R_x Hs$), (*) and (**) are easily shown to be equivalent. If, however, we strengthen Def.1 by choosing an implicative conditional instead of '$\supset$' this result does no longer obtain and the difference between (*) and (**) turns out to be substantial.

[4] Proofs can be found in the paper Heydrich/Rieser (1995).

1.3 Reasons to believe and actual belief

Having derived all the clauses (2)(i), (ii), (iii), ..., etc., from (LA) (*plus* some background assumptions) we are "still not talking about anyone's actual reasoning or what anyone actually does believe"(Lewis 1969, p. 55). What is needed at each level of the hierarchy of nested *reasons to believe* to arrive at *actual belief* is - according to Lewis - "a sufficient degree of rationality" and, additionally, "reason to ascribe a sufficient degree of rationality to the other" (ibid.).

We noted above that rationality and justified ascription of rationality to others - although perhaps sufficient - are by no means also necessary to reach a level of intersubjective transparency of information sufficient for joint action.

We will, nevertheless, have a closer look here at Lewis' conversion of "iterations of "has reason to believe" to the corresponding iterations of "does believe""(ibid.) We regard Lewis' account as an idealised model illuminating the relation beween the level of public information (as a notion of logical epistemology) and actual mutuality of knowledge (conceived of in terms of psychological epistemology).

The degree of rationality needed for converting *reasons to believe that* p to actual *belief that* p will - in part at least - depend on the conceptual complexity of p. Reading "$Rat_x p$" as "x is rational enough to convert *reasons to believe that* p in *belief that* p" and writing "$B_x p$" short for "x believes that p", we define[5]:

Def.2 $\quad Rat_x p := R_x p \supset B_x p$

Additionally, we accept a schematic generalization of (a), according to which not only single R_x-operators (cf. (a)), but also all finite sequences of R_x-operators distribute over material implication:

(a') $\quad R_{x1} \ldots R_{xn}(p \supset q) \supset (R_{x1} \ldots R_{xn} p \supset R_{x1} \ldots R_{xn} q)$

In order to convert (2)(i) to $\forall x \in \{Inst, Const\}\ B_x p$, it is sufficient to assume:

(4) $\quad \forall x \in \{Inst, Const\}\ Rat_x p$

In order to convert (2)(ii) to $\forall x,y \in \{Inst, Const\}\ B_x B_y p$, however, stronger assumptions are necessary. It is sufficient to assume:

(5) (i) $\quad \forall x, y \in \{Inst, Const\}\ R_x Rat_y p$
(ii) $\quad \forall x, y \in \{Inst, Const\}\ Rat_x B_y p$

According to (5)(i), both Inst and Const have reason to believe that the other is rational enough to convert reasons to believe that p to the actual belief that p. According to (5)(ii) both Inst and Const are rational enough to convert reasons for believing that the other believes that p to the actual belief that the other believes that p.

[5] Here again the use of material implication is not unproblematic (cf. footnote 3). Our definition entails (quite fallaciously, we think) that believing that p implies being rational enough to convert reason to believe that p to belief that p.

In order to convert (2)(iii) to $\forall x, y \in \{Inst, Const\}\ B_xB_yB_xp$, still stronger assumptions are necessary. It is sufficient to assume:

(6) (i) $\forall x, y \in \{Inst, Const\}\ R_xR_yRat_xp$
(ii) $\forall x, y \in \{Inst, Const\}\ R_xRat_yB_xp$
(iii) $\forall x, y \in \{Inst, Const\}\ Rat_xB_yB_xp$

It is obvious from these first steps that "the degrees of rationality we are required to have [...] increase quickly. That is why [beliefs] of only the first few orders are actually formed."(Lewis 1969, p. 56)

1.4 Verbal interaction as a basis for mutuality

The informational content of a situation (i.e. what it indicates to its participants) is public, iff it is informationally transparent (i.e. iff it correctly indicates to the participants that each of them has reason to believe that it obtains). An informationally transparent situation provides a basis for mutually shared knowledge insofar as the infinite supply of information it offers is freely available for its participants and may be exploited *ad libitum.*

Lewis (1969) gives some examples of features rendering situations informationally transparent: "agreement, salience or precedent" (p.52). Clark and Marshall (1981) develop a more comprehensive typology of pertinent situations among which perceptually overt and surveyable ones (type: *physical copresence*) are central. Situations of this type - being the paradigm case of informational transparency - are characterised by the fact that both the situations's participants and all the things they are dealing with are freely perceptible in each relevant respect.

Obviously, our agents' situation during the construction task is far from being of this type. For Inst Const's blocks are hidden behind the screen, and so are Inst's blocks for Const. Verbal interaction provides the only medium for making available information about the blocks on one's own side of the screen and to get information about the blocks on the other side. Since the set-up of the situation does not - on mere perceptual grounds - provide a basis for mutually shared knowledge, the participants have to face the task to further *develop their situation* in the sense of *establishing features of informational transparency.* The general goal of Inst and Const at each stage of the construction process is to develop their common situation such that the configurations of blocks on each side of the screen are sufficiently akin. If the agents managed to establish this as public (hence intersubjectively transparent) information this would be an ideal substitute for a perceptually overt and surveyable situation of the type physical copresence. Inst and Const accomplish their task by means of conversational procedures, mainly procedures of *ratification.* Ratification has the function of making sure that some piece of information can henceforth be considered as mutually accepted and agreed upon by all discourse participants. Very often, we find explicit turns of ratification in our data, but also non-intervention may - under certain conditions - contribute by default to the procedure of ratification. In any case, ratification makes information public. As a mechanism of processing verbally articulated content, ratification is a means (perhaps the most prominent one) to *constitute* an informationally transparent situation where the perceptual surrounding fails to guarantee informational transparency.

Take the utterance

4-10 I: also unten steht ein grünes langgezogenes Rechteck hochkant well, on the ground stands a green longish rectangle upright

and let "α_1" stand for the information articulated here. Then only after α_1 has been processed and ratified, Inst and Const have managed to establish a common situation which is informationally transparent and where the configurations of blocks on each side of the screen are indicating α_1 to both agents. Under these circumstances α_1 has become public information for Inst and Const and may be exploited by each of them in order to establish beliefs (about the blocks and (possibly nested) beliefs of the other one). Given that these beliefs exhibit certain features of (limited) mutuality they may become causally relevant for joint action and will lead to the next steps of construction and instruction.

Let us have a closer look at what is going on.

2. Coordination and belief

2.1 Calling blocks *rectangles*

The theory discussed so far will now be applied to the initial passage of a transcript recording part of a blocks world session. The transcript shows the verbal interaction going on during a constructor's building a see-saw according to the directives of an instructor. The instructor focusses the built-up see-saw placed before him (cf. Fig. 1). The agents are separated by a screen, they can freely converse, however.

4-10 I: also unten steht *ein grünes langgezogenes Rechteck* hochkant
well, on the ground stands *a green longish rectangle* upright
4-11 C: hochkant ja
upright, yes
4-12 I: und vorne . wobei also die lange Seite nach oben zeigt nich
and at the front . the long side is pointing upwards right
4-13 | die kurze Seite ne' also hochkant jetzt das
the short side right' hence upright next the

4-14 C: | hochkant jaja längs
upright yes
4-15 hoch
vertical
4-16 I: und vorne drauf liegt *ein grünes eh n blaues Rechteck* wobei
and at the front there is *a green, sorry blue rectangle* with
4-17 die +s+ breite +s+ . Seite nach oben zeigt, . das legste da drauf,
the wide side pointing upwards, . put this one on the other one.
4-18 C: ja
yes.

Example from a task-oriented dialogue

We do not want to go into the details of the transcript here. Notice, however, that blocks are called *rectangles* by the instructor. He first describes in line 4-10 a green flat block standing on a thin long side, but does not specify the exact position of the block. A blue flat block of the same shape (see lines 4-16/4-17) is placed at the front of the green block's narrow side pointing upwards. The blue block forms the seesaw's beam (cf. Fig.1). In line 4-12 the

instructor first starts to describe the next block, which is the blue one, but then he adds a repair concerning the green block mentioned first. He needs the repair's more elaborate description of surfaces involved in order to introduce the position of the blue block. Fig. 1 shows the block's world on the instructor's side.
Fig. 2 depicts the block focussed by him before he produces the speech act 4.10.

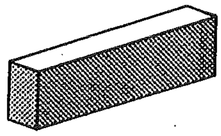


Fig. 2 Inst's focussed block before 4-10

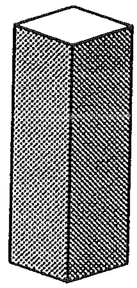


Fig. 3 Const's first manipulated block

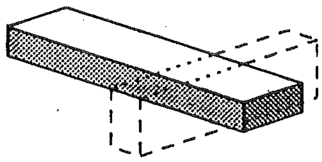


Fig. 4 Inst's focussed block before 4-12

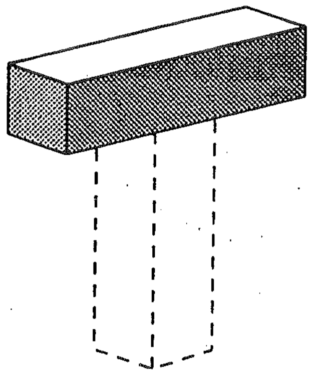


Fig. 5 Const's second block

The block selected and manipulated by the constructor is shown in Fig. 3. Inst has the constellation in Fig. 4 in mind before he starts to produce 4-12. Const puts a block on his first one which does not match the instructor's intentions.
Before we enter upon our discussion of the agents' beliefs and intentions, we want to say a few words about the succession of actions in a task-oriented dialogue.

2.2 Private *vs* public information: intuitions

The biggest problem Inst and Const face is the coordination of information. They have different sources of information indicating different things to them. These different information sources are on the one hand the private situations they are in and on the other hand the situations common to both of them. The private situations are mainly based on the visual contact with the respective working spaces, common situations arise due to verbal interaction.

In order to get on with the task, the opposition private *vs* public information has to be levelled after every one of the agents' task-relevant actions. What can count as public information is determined by the content the utterance tokens transport and by the ratifications of the agents, hence, by the things they acknowledge as intersubjectively valid. Since the private *vs* public opposition exists, the agent's current information situation is clearly asymmetric: At the start, Inst has the whole blocks' world before him. Therefore he may design a plan which action to demand from Const and which speech act to produce in order to achieve the task result necessary. Const can only learn from Inst's speech act how Inst's situation looks like and how his own situation, therefore, should turn out to be. Here Inst has privileged access to the information needed. Const, on the other hand, has privileged access to the result of his action. Not until after Const's action report will Inst know what the situation on Const's side is. Hence we have an overall adjustment of private information and public information, i.e. the different levels of information get balanced.

Before Inst and Const can further proceed with their task, a suitable niveau of mutual belief must be established: Const must believe that the situation demanded by Inst is of the "Inst-situation type". Inst must believe this as well. After Const's action, Inst must believe that the situation asked for was realised on Const's side. Const must, of course, believe the same thing. Hence both of them must believe that the situations match on both sides. In other words: there has to exist cumulative information leading to the level of mutual belief necessary to keep the whole matter going.

Fig. 7 shows how types of situations and actions of agents are related to each other. In this figure we have two realisations of a schema of the following sort: Inst looks at his private situation, focusses an object there, builds up an intention according to his plan concerning task solution, and produces a suitable *directive*. Const answers the directive by *acting* in the way expected. Then he *confirms* the result of his action. Inst shows *approval* of Const's *confirmation* and, consequently, is prepared to start the next utterance. Such a pattern of verbal and non-verbal actions is fairly prototypical with respect to task-oriented dialogue. It embodies one directive and the ensuing consecutive actions. Then the next directive can be started. In addition to the directive-action-confirmation-acceptance schema we have a repair sequence 3 in the data. Natural discourses contain various patterns of similar sorts e.g. description of objects, defense of/attack of arguments or wording out alternative routes of actions.

We model the directive-action-confirmation-acceptance schema as a cycle which has a stage-wise structure. Each cycle capturing the manipulation of one object has essentially four stages. It starts with Inst's directive concerning the block to be manipulated (stage 1), followed by an action of Const's (stage 2), Const's confirmation of his action (stage 3), and, finally, Inst's approval of Const's confirmation of the action carried out (step 4). Stage 0 is preparatory. It consists only of Inst's viewing his scene. Cycles such as this one can be iterated several times, depending on the global task at hand.

Fig. 7 contains information in addition to cycles and stages, namely, information concerning Inst's private situations, Const's private situations, a classification of the actions observed and, finally, the tokens coming from the agent's verbal activities. Numerical indices used in the description have two parameters, separated by a dot. The first indicates the current cycle and the second the current stage. For example, $s_{1.0}$ is a private situation of Inst's in the first cycle at stage 0 and $s_{1.4}$ is a common situation of Inst's and Const's in the first cycle at stage 4.

Cycl	Stage	Priv. Sit. Inst	Common Sit.	Priv. Sit. Const.	Actions	Tokens
1	0	$s_{1.0}$			Inst's perception	
1	1		$s_{1.1}$		Inst's directive 1	4-10 I: also unten steht *ein grünes langgezogenes Rechteck* hochkant well, on the ground stands *a green longish rectangle* upright
1	2			$s_{1.2}$	Const's action 1	
1	3		$s_{1.3}$		Const's confirmation of action 1	4-11 C: hochkant ja upright, yes
1	4		$s_{1.4}$		Inst's acceptance of Const's action 1	
2	0	$s_{2.0}$			Inst's perception	
2	1		$s_{2.1}$		Inst's directive 2	4-12 I: und vorne . and at the front .
3	1		$s_{3.1}$		Inst's self-repair of directive 1	4-12 I: wobei also die lange Seite nach oben zeigt nich and at the front . the long side is pointing upwards right 4-13 ¦ die kurze Seite ne' also hochkant jetzt das the short side right' hence upright next the

3	2		$s_{3.2}$		Const's confirmation of Const's action 1 and his repetition of Const's confirmation of action 1	4-14 C: | hochkant jaja längs upright yes 4-15 hoch vertical
2	1		$s_{2.1}$		Inst's resumption of his own directive 2	4-16 I: und vorne drauf liegt *ein grünes eh n blaues Rechteck* wobei and at the front there is *a green, sorry blue rectangle* with 4-17 die +s+ breite +s+ . Seite nach oben zeigt, . das legste da drauf, the wide side pointing upwards, . put this one on the other one.
2	2			$s_{2.2}$	Const's action 2	
2	3		$s_{2.3}$		Const's confirmation of own action 2	4-18 C: ja yes.
2	4		$s_{2.4}$		Inst's approval of Const's action 2	

Fig. 7 Schema of observable situations and actions

The directive-action-confirmation-acceptance cycle exhibited by the example is fairly commonplace. However, shortly after the start of the second cycle Inst comes up with a repair interrupting the normal goings-on. A repair cycle (cf. cycle 3) consisting of two steps emerges: First we have Inst's attempt at repairing his previous directive. Then Const closes this repair-*gestalt* by a confirmation of his (Const's) last action (step 2 in cycle 1). Unfortunately, he does not show the optimal reaction to Inst's attempt at repair. After Const's confirmation Inst regards the repair-*gestalt* as closed, resumes his second directive and cycle 2 proceeds in the normal fashion.

2.3 Private *vs* public information: formal explication

Fig. 7 represents only a rough outline of the things actually going on. It does not yet capture explicitly all the information concerning the agents' beliefs and intentions. According to a more fine-grained perspective, we find at each stage of a cycle (e.g. directive, action, confirmation of action, and approval) a sequence of two cognitive states, an action and the resulting situation as shown in Fig. 8:

Belief ⇒ Intention ⇒ Action ⇒ Situation
╙ else .....

Fig. 8 Schema of a sequence of cognitive states and the resulting action and situation

Applying the schema set up in Fig. 8 to the example presented in Fig. 7, we get the following more elaborate variant:

(1) Inst believes his situation $s_{1.0}$ to be such-and-such. Consequently, he builds up an intention leading to the afore-said directive. The directive in turn engenders a common situation $s_{1.1}$ to the effect that Inst and Const have reason to believe that Inst's situation $s_{1.0}$ indicates to Inst such-and-such. If Inst could not acquire the initial belief, he most probably had doubts concerning the information indicated by his situation. This could have happened if, say, he had observed some weird Escherian object there instead of the block's world.

Assuming that things happen in the right way, then, considering Const's side, we will have the following:

(2) Given that Const believes that it is public information that Inst's situation indicates to Inst such-and-such, he will intend to make it happen that his own situation indicates to him such-and-such, i.e. indicates to him roughly the content of Inst's directive. Const will carry out the directive thereby producing a situation $s_{1.2}$. $s_{1.2}$ will indicate such-and-such to him. Of course, if he had not indulged in the belief about Inst's situation $s_{1.0}$ indicating such-and-such to Inst in the first place, matters would have turned out differently.

Furthermore,

(3) if Const has acquired the additional belief that his own situation indicates such and such to him, he will intend to confirm that his situation $s_{1.2}$ indicates such-and-such to him, and in doing so provide the basis $s_{1.3}$ for common information that his situation $s_{1.2}$ indicates such and such to him.

After all that we are still not done, although we are swiftly approaching the end of the cycle:

(4) Assuming that Inst believes that it is public information that $s_{1.2}$ indicates such and such to Const and that $s_{1.0}$ indicates such and such to Inst, he will react in the following way: He

will form the intention to accept that Inst's situation $s_{1.0}$ and Const's situation $s_{1.2}$ indicate the same thing to both of them. Then he will word his acceptance. This in turn will lay the foundation $s_{1.4}$ for the public information that the situations on both sides indicate the same thing.

In order to achieve a more fine-grained modelling of cycles embodying the information described in Fig. 8, we have to introduce some notational conventions provided below. The values of the subindices chosen have to be generalised in the obvious way, e.g. "Const" could be used instead of "Inst".

$B_{Inst}p$	Inst believes that p
$Hs_{i.j}$	situation s in cycle i and stage j holds
$s_{i.j} \models_{Inst} \alpha_i$	situation s in cycle i and stage j indicates to Inst that α_i
$A \Rightarrow B$	if A then B
$A \Rightarrow B$ ╙ C	if A then B else C
$I_{Inst}p$	Inst intends that p
$!_{Inst}p$	Inst issues the directive that p
$Do_{Inst}p$	Inst does p
$P_{\{Inst,Const\}}p$	it is public information among Inst and Const that p
$R_{Inst}p$	Inst has reason to believe that p
$?_{Inst}p$	Inst calls into question that p
$Conf_{Inst}p$	Inst confirms that p
$Acc_{Inst}p$	Inst accepts that p
LA[{Inst,Const}, $s_{i.1}$, $s_{i.0} \models_{Inst} \alpha_i$]	Lewis' account of public information with slots for participants ({Inst,Const}), the situation involved ($s_{i.1}$), and the content believed ($s_{i.0} \models_{Inst} \alpha_i$)

Fig.9 below shows how a more fine-grained explanation can be set up by a series of decision trees. It delineates a recursive process. The idea expressed therein is as follows: Idealising somewhat, we may assume that Inst has a plan π of the task to be accomplished, already neatly cut up into steps. If Inst starts with i = 1, no value for k will be selected and the antecedent

$$\text{for all } k\ (1 \leq k < i)\ Hs_{k.4}\ \&\ B_{Inst}P_{\{Inst,Const\}}(s_{k.0} \models_{Inst} \alpha_k\ \&\ s_{k.2} \models_{Const} \alpha_k)$$

will turn out trivially true. This leads to carrying out the consequence

$$Hs_{i.0},\ LA[Inst,\ s_{i.0},\ \alpha_i],\ I,\ II,\ III,\ \text{and IV}.$$

Observe that the following short-cut will be used: In the four clauses below we always have a transition involving four issues:

Firstly, the fact that a certain situation holds plus the agent's belief concerning situations,

secondly, the agent's intention to accomplish something,

thirdly, the action accomplished,

fourthly, the fact that some situation evolves indicating public information in the sense of (LA).

Suppose that Inst has a plan π of task-relevant construction-steps $\alpha_1,...,\alpha_n$. Then the following holds:

For all i $(1 \leq i \leq n)$: If for all k $(1 \leq k < i)$ $Hs_{k.4}$ & $B_{Inst}P_{\{Inst,Const\}}(s_{k.0} \vDash_{Inst} \alpha_k$ & $s_{k.2} \vDash_{Const} \alpha_k)$ then

$Hs_{i.0}$, $LA[Inst,\ s_{i.0},\ \alpha_i]$, I, II, III, and IV, where:

(I) $Hs_{i.0}$ & $B_{Inst}s_{i.0} \vDash_{Inst} \alpha_i \Rightarrow\ !_{Inst}Do_{Const}s_{i.2} \vDash_{Const} \alpha_i \Rightarrow Hs_{i.1}$ & $LA[\{Inst,Const\},\ s_{i.1},\ s_{i.0} \vDash_{Inst} \alpha_i]$

‖

╙ $?_{Inst}s_{i.0} \vDash_{Inst} \alpha_i$

(II) $Hs_{i.1}$ & $B_{Const}P_{\{Inst,Const\}}s_{i.0} \vDash_{Inst} \alpha_i \Rightarrow Do_{Const}s_{i.2} \vDash_{Const} \alpha_i \Rightarrow Hs_{i.2}$ & $LA[Const,\ s_{i.2},\ \alpha_i]$

‖

╙ $?_{Const}P_{\{Inst,Const\}}s_{i.0} \vDash_{Inst} \alpha_i$

(III) $Hs_{i.2}$ & $B_{Const}(P_{\{Inst,Const\}}s_{i.0} \vDash_{Inst} \alpha_i$ & $s_{i.2} \vDash_{Const} \alpha_i) \Rightarrow Conf_{Const}s_{i.2} \vDash_{Const} \alpha_i \Rightarrow Hs_{i.3}$ & $LA[\{Inst,\ Const\},\ s_{i.3},\ s_{i.2} \vDash_{Const} \alpha_i]$

‖

╙ $?_{Const}(P_{\{Inst,Const\}}s_{i.0} \vDash_{Inst} \alpha_i$ & $s_{i.2} \vDash_{Const} \alpha_i)$

(IV) $Hs_{i.3}$ & $B_{Inst}(P_{\{Inst,Const\}}s_{i.2} \vDash_{Const} \alpha_i$ & $s_{i.0} \vDash_{Inst} \alpha_i) \Rightarrow Acc_{Inst}(s_{i.2} \vDash_{Const} \alpha_i$ & $s_{i.0} \vDash_{Inst} \alpha_i) \Rightarrow Hs_{i.4}$ & $LA[\{Inst,Const\},\ s_{i.4},\ s_{i.2} \vDash_{Const} \alpha_i$ & $s_{i.0} \vDash_{Inst} \alpha_i]$

‖

╙ $?_{Inst}(P_{\{Inst,Const\}}s_{i.2} \vDash_{Const} \alpha_i$ & $s_{i.0} \vDash_{Inst} \alpha_i)$

Fig.9 Schema for the interaction of belief, intention, action, public information, and change

E.g.

$$Hs_{i.0} \;\&\; B_{Inst}s_{i.0} \vDash_{Inst} \alpha_i \Rightarrow \underline{I_{Inst}!_{Inst}Do_{Const}s_{i.2} \vDash_{Const} \alpha_i} \Rightarrow !_{Inst}Do_{Const}s_{i.2} \vDash_{Const} \alpha_i \Rightarrow Hs_{i.1} \;\&\; LA[\{Inst, Const\},\ s_{i.1},\ s_{i.0} \vDash_{Inst} \alpha_i]$$

$$\Vert$$

$$\llcorner ?_{Inst}s_{i.0} \vDash_{Inst} \alpha_i$$

Since we do not make much of the intention-part in this section (cf. however section 4), we omit it for reasons of greater perspicuity. Intentions will be mentioned in the description of Fig.9 , however.

Line (I) says: If situation $s_{i.0}$ holds and Inst believes that it indicates α_i to him, then he will intend to issue a directive demanding that Const bring about that Const's situation, $s_{i.2}$, indicate α_i to Const. This in turn will lead to his production of a *directive* of the appropriate sort. As a consequence of that the ground $s_{i.1}$ is laid for the public information among Inst and Const that Inst's situation, $s_{i.0}$, indicates α_i to Inst.

Line (II) says the following: Given that $s_{i.1}$ holds and Const believes that it is public information among Inst and Const that the situation $s_{i.0}$ indicates α_i to Inst, he will form an intention eventually leading to an appropriate *action.* A suitable action result presupposed, $s_{i.2}$ obtains and the situation $s_{i.2}$ is accessible, perspicuous and significant with respect to α_i.

Line (III) comes up with the following suggestion concerning Const's side: If $s_{i.2}$ holds and Const believes it is public information for Inst and Const that the situation $s_{i.0}$ indicates α_i to Inst and believes that $s_{i.2}$ indicates α_i to Const then he will needs intend to confirm that his situation indicates α_i to Inst and then also *confirm* this very fact. *Confirmation* in turn will yield $s_{i.3}$. This is the ground for the public information that $s_{i.2}$ indicates α_i to Const.

Line (IV) models Inst's response. If $s_{i.3}$ holds and Inst believes that it is public information among Inst and Const that $s_{i.2}$ indicates α_i to Const and believes that $s_{i.0}$ indicates α_i to Inst then Inst will intend to *accept* that the situations $s_{i.2}$ and $s_{i.0}$ respectively indicate α_i to Inst and Const and he will accept that α_i is indicated to the respective agents in the relevant situations. This move eventually yields the public information that the situations $s_{i.2}$ and $s_{i.0}$ respectively indicate α_i to Inst and Const.
The different else-conditions obtain, if the antecedents turn out to be false.

2.4 Private *vs* public information: modelling Lewis' account

As we pointed out above, a Lewisian account of epistemic mutuality is concerned with the mediation of

(a) information as an objective feature of situations (its role within logical epistemology) and

(b) information as a source of agent's attitudes, intentions and actions (its role within psychological epistemology).

Both roles may be combined, we argued, provided the agents share some significant situation as participants which is accessible and perspicuous for them (in the sense of LA(i) and (ii)). In this case, the situation's informational content (its significance in the sense of LA(iii)) is intersubjectively transparent: its participants mutually have reason to believe in the situation's informational content, and - moreover - they mutually have reason to believe that its content is public information among them. It is up to the agents to exploit subjectively transparent information. In doing so, they will acquire convictions and beliefs leading to further action. The degree of higher order mutuality in these beliefs will in general depend

(a) on the agents' sophistication in transforming (objectively ascribable) reason to believe into (subjectively operative) actual belief and
(b) on the agents' readiness to attribute such sophistication to their co-agents.

The pattern of interaction schematised in Fig.9 is about making privately retrievable information public, and it concerns a task-oriented way of mediating subjective and objective features of communication. The subjective features deal with aspects of psychological epistemology (beliefs about private situations of perception (on Inst's part) and awareness of constructional activity (on Const's part)). The objective features concern aspects of logical epistemology (information encoded in utterances, i.e. overtly recognizable speech-act events). Our schema implements a Lewisian account of epistemic mutuality in a dynamic and procedural way. It is intended to exhibit a cooperative mechanism of rule-governed interaction according to which Inst and Const are concerned with the establishment of accessible and perspicuous situations licensing intersubjectively transparent information - information that can be exploited by (co-)agents in order to acquire beliefs which, again, lead to action establishing new accessible and perspicuous situations, etc.. In this way, Inst and Const may work through all relevant construction steps of Inst's plan and, in connection with each such step, through the associated moves of directive, action, confirmation and acceptance, which are considered to be the basic units of the interactive pattern of giving and following directives. To each relevant construction step α_i within plan π we assign a cycle i of four successive stages (I - IV) corresponding to the moves.

Associated with each such move are preconditions and consequences. Both of them require that certain situations hold which determine the position of the move within the pattern of interaction. The situations are conceived of as the direct result of performing some move. They characterise how far things have been developed at cycle i and allow Inst and Const to exploit information and form beliefs leading to intentionally performing the next move. Not until the fourth move of acceptance is performed, will a situation obtain which provides the ground for the public information among Inst and Const that the construction step α_i is realised on both sides of the screen (in the sense that α_i is indicated by $s_{i.0}$ - Inst's private situation of perception - as well as $s_{i.2}$ - Const's private situation of constructional activity). After the moves due to the four stages have been accomplished successfully, things may (and have to) repeat themselves at the next cycle (associated with construction step α_{i+1}). And so forth, until the last step α_n of Inst's plan π is reached and has been processed successfully. In order to illustrate how the schema of Fig.9 is supposed to work through all the stages and cycles, let us assume that everything went fine up to step α_i such that for all k ($1 \leq k < i$) the situation $s_{k.4}$ resulting from the kth move of acceptance actually holds and Inst has processed enough information to arrive at the belief that it is public information among him and Const that α_k is realised on both sides of the screen.

Given plan π and given the configuration of blocks before him, Inst will now consider the next step. Let us assume that a situation $s_{i.0}$ obtains which is accessible and perspicuous for

Inst and which indicates α_i to him. In other words, we assume

(7) $Hs_{i.0}$ and
(8) $LA[Inst, s_{i.0}, \alpha_i]$.

$s_{i.0}$ is the initial situation for the sequence of moves between Inst and Const that constitute the stages of cycle i. It is the situation of Inst's perceptional awareness of (and focussing on) a specific configuration of blocks at his side of the screen. Since Inst is $s_{i.0}$'s only participant, the situation is private in character: whatever information $s_{i.0}$ licenses, it will indicate it to nobody else but Inst.

(7) is the first conjunct of the antecedent in stage (I) of cycle i which formulates the precondition of Inst's move of giving the directive that Const should realise α_i on his (Const's) side of the screen. The second conjunct of the antecedent in stage (I) is

(9) $B_{Inst}s_{i.0} \models_{Inst} \alpha_i$.

Hence, we require that Inst should be aware of the fact that $s_{i.0}$ indicates α_i to him, in order to give his directive. Since $s_{i.0}$ is accessible and perspicuous for Inst, it is not hard for him to arrive at this awareness. (9) can be derived from (8), assumption (*) (formulated in 1.3) and

(10) $Rat_{Inst}s_{i.0} \models_{Inst} \alpha_i$,

i.e. the supposition that Inst is rational enough to proceed from *reason to believe* that his private perceptual situation $s_{i.0}$ indicates α_i to him to the corresponding *actual belief*.[6] Note that (9) is a simple belief about what $s_{i.0}$ indicates to Inst. No higher-order belief is required[7]. Note further that (8) and assumption (*) are sufficient to derive $R_{Inst}\alpha_i$, $R_{Inst}s_{i.0} \models_{Inst} \alpha_i$, $P_{\{Inst\}}\alpha_i$ and $P_{\{Inst\}}s_{i.0} \models_{Inst} \alpha_i$, with arbitrary additional prefixes of the operators R_{Inst} or $P_{\{Inst\}}$.

According to clause (I) we stipulate that (7) and (9) are jointly sufficient for Inst's (intention to give the) directive that Const should realise α_i on his (Const's) side of the screen, i.e. that he should act such that his private situation of constructional activity $s_{i.2}$ indicates α_i to him. Giving this directive will in turn lead to a situation $s_{i.1}$. Since directives are socially overt speech-acts, $s_{i.1}$ is common and not private. In contrast to $s_{i.0}$ (with Inst as its single participant), $s_{i.1}$ is shared by Inst and Const. Being accessible and perspicuous for both, it makes the content of Inst's belief in (9) (i.e. $s_{i.0} \models_{Inst} \alpha_i$) public among Inst and Const. In other words, it will lead to the consequences

6 (8) implies $R_{Inst}Hs_{i.0}$ and $s_{i.0} \models_{Inst} \alpha_i$. From the latter we obtain $s_{i.0} \models_{Inst} s_{i.0} \models_{Inst} \alpha_i$ by (*). Using Def. 1 and detachment we arrive at $R_{Inst}s_{i.0} \models_{Inst} \alpha_i$. (10) and Def. 2 lead immediately to (9).

7 Alternatively, one might think of requiring that Inst believes that it is public information for him that $s_{i.0}$ indicates α_i to him (i.e $B_{Inst}P_{\{Inst\}}s_{i.0} \models_{Inst} \alpha_i$ instead of (9)).Likewise - no higher-order belief or mutuality hierarchy being involved here - this is a simple belief about $s_{i.0}$: the belief that $s_{i.0}$'s indicating a_i to Inst is indicated by some (otherwise unspecified) accessible and perspicuous situation to him. Without a compelling argument, we prefer the version without the $P_{\{Inst\}}$-operator. Note that the two versions are logically independent as long as we do not posit additional constraints on B, R and H.

(11) $Hs_{i.1}$ and
(12) $LA[\{Inst,Const\}, s_{i.1}, s_{i.0} \models_{Inst} \alpha_i]$.

(11) in turn is the first conjunct of the antecedent at the next stage (II) of the cycle i which formulates the precondition of Const's move of acting according to Inst's directive. The second conjunct is

(13) $B_{Const}P_{\{Inst,Const\}}s_{i.0} \models_{Inst} \alpha_i$.

Thus, we count among the preconditions of Const's move of constructional activity his awareness of the fact that it is public information for him and Inst that things are arranged according to α_i on Inst's side of the screen (cf. (13)), i.e. the fact that there is some accessible and perspicuous situation indicating $s_{i.0} \models_{Inst} a_i$ to him. Since $s_{i.1}$ is such a situation and Const is among its participants, it is not hard for him to arrive at this awareness. We may derive (13) from (12), assumption (*), (a) - (d) (cf. 1.3) and the essential assumption

(14) $Rat_{Const}P_{\{Inst,Const\}}s_{i.0} \models_{Inst} \alpha_i$,

i.e. the supposition that Const is rational enough to proceed from *reason to believe* that it is public information among Inst and Const that $s_{i.0} \models_{Inst} \alpha_i$ to the corresponding actual belief.[8] Note that believing that $s_{i.0} \models_{Inst} \alpha_i$ is public information is belief in the existence of an accessible and perspicuous situation indicating $s_{i.0} \models_{Inst} \alpha_i$ to all its participants. It is not belief about belief of others, and especially it is not belief about other's belief about one's own belief. We are dealing with first-order belief here. No mutuality hierarchy of beliefs is at stake. No paradox threatens. Note further that (12) and assumption (*) are sufficient to derive $R_{Const}s_{i.0} \models_{Inst} \alpha_i$, $R_{Inst}s_{i.0} \models_{Inst} \alpha_i$, and $P_{\{Inst, Const\}}s_{i.0} \models_{Inst} \alpha_i$, with arbitrary additional prefixes of the operators R_{Inst}, R_{Const} and $P_{\{Inst, Const\}}$. Hence mutuality hierarchies result. But these hierarchies only depend on the notions of having-reason-to-believe-that and being-public-information-that. They are not based on the notion of believing-that. Only the realm of logical epistemology is concerned, not the realm of psychological epistemology, which is causally relevant for intention and action. Therefore, again, no paradox threatens. According to clause (II) we stipulate that (11) and (13) are jointly sufficient for Const's (intending to) act along the lines of Inst's directive, i.e. he will realise α_i on his side of the screen. As a consequence, a situation $s_{i.2}$ obtains which is accessible and perspicuous for Const alone and which indicates α_i to him. In other words, we have

(15) $Hs_{i.2}$ and
(16) $LA[Const, s_{i.2}, \alpha_i]$.

$s_{i.2}$ is the situation of Const's constructional activity of building a specific configuration of blocks at his side of the screen. Since Const is $s_{i.2}$'s only participant, it is (in contrast to $s_{i.1}$ but like $s_{i.0}$) private in character: whatever information $s_{i.2}$ licenses, $s_{i.2}$ will indicate it to nobody else but Const.

[8] It is a simple step of existential generalisation from (12) to $P_{\{Inst,Const\}}s_{i.0} \models_{Inst} \alpha_i$. As indicated in (1.3), we may iterate the $P_{\{Inst,Const\}}$-operator to obtain $P_{\{Inst,Const\}}P_{\{Inst,Const\}}s_{i.0} \models_{Inst} \alpha_i$ and proceed from this to $R_{Const}P_{\{Inst,Const\}}s_{i.0} \models_{Inst} \alpha_i$. As made clear in 1.4 the rationality assumption (14) and Def. 2 will lead from this to (13).

(15) is again the first conjunct of the antecedent of the next stage (III) at the cycle i which formulates the precondition of Const's move of confirming what he has realised on his side of the screen. In order to argue for the truth of the second conjunct

(17) $B_{Const}(P_{\{Inst,Const\}}s_{i.0} \vDash_{Inst} \alpha_i \ \& \ s_{i.2} \vDash_{Const} \alpha_i)$

we observe its equivalence with

(17') (13) & $B_{Const} s_{i.2} \vDash_{Const} \alpha_i$

by assuming distribution of the B_x-operator over conjunction:

(e) $B_x(p \ \& \ q) \equiv B_x p \ \& \ B_x q$.

Since (13) has already been derived (essentiality by the rationality assumption (14)), it remains to be shown how to derive

(18) $B_{Const} s_{i.2} \vDash_{Const} \alpha_i$

from (16). Stipulating that (18) is entailed by the antecedent of stage (III), we require that Const should be aware of the fact that $s_{i.2}$ indicates a_i to him, in order to confirm his constructional activity. Since $s_{i.2}$ is accessible and perspicuous for Const, it is not hard for him to arrive at this awareness. Besides assumption (*), the requisite rationality assumption is

(19) $Rat_{Const} s_{i.2} \vDash_{Const} \alpha_i$,

i.e. the supposition that Const is rational enough to proceed from *reason to believe* that his private situation of activity $s_{i.2}$ indicates α_i to him to the corresponding *actual belief*.[9] Note that (18) is a simple belief about what $s_{i.2}$ indicates to Const. No higher-order belief is required. Note further that (16) and assumption (*) are sufficient to derive $R_{Const} a_i$, $R_{Const} s_{i.2} \vDash_{Const} a_i$, $P_{\{Const\}} a_i$ and $P_{\{Const\}} s_{i.2} \vDash_{Const} a_i$, with arbitrary additional prefixes of the operators R_{Const} and $P_{\{Const\}}$. (The relation between $s_{i.2}$, Const and the informational content $s_{i.2} \vDash_{Const} a_i$ in the antecedent at stage (III) is strictly parallel to the relation between $s_{i.0}$, Inst and the informational content $s_{i.0} \vDash_{Inst} a_i$ in the antecedent at stage (I). This is because both $s_{i.2}$ and $s_{i.0}$ are private situations, not common ones)[10].

According to clause (III) we stipulate that (15) and (17) are jointly sufficient for Const's (intending to) confirm that α_i is realised on his side of the screen. As a consequence, the situation $s_{i.3}$ obtains. Since confirmations are socially overt speech-acts, $s_{i.3}$ is common and not private. Like $s_{i.1}$, but in contrast to $s_{i.0}$ and $s_{i.2}$ (with Inst and Const respectively as their

---

[9] (16) implies $R_{Const} H s_{i.2}$ and $s_{i.2} \vDash_{Const} \alpha_i$. From the latter we obtain $s_{i.2} \vDash_{Const} s_{i.2} \vDash_{Const} \alpha_i$ by (*). Using Def.1 and detachment we arrive at $R_{Const} s_{i.2} \vDash_{Const} \alpha_i$. (19) and Def. 2 lead immediately to (18).

[10] In analogy to the alternative account for stage (I) suggested in footnote 7 one might think about inserting a $P_{\{Const\}}$-operator immediately before $s_{i.2} \vDash_{Const} \alpha_i$ in the antecedent of stage (III). Here, like there, we prefer the version without the additional operator.

single participants), $s_{i.3}$ is shared by Inst and Const. Being accessible and perspicuous for both, it makes the content of Const's belief (18) (i.e. $s_{i.2} \models_{Const} a_i$) public among Inst and Const.

In other words we have

(20) $Hs_{i.3}$ and
(21) $LA[\{Inst,Const\}, s_{i.3}, s_{i.2} \models_{Const} \alpha_i]$.

Again, (20) is the first conjunct in the antecedent of the next stage (IV) in the cycle i formulating the precondition of Inst's move of acceptance. In order to argue for the truth of the second conjunct

(22) $B_{Inst}(P_{\{Inst,Const\}} s_{i.2} \models_{Const} \alpha_i \ \& \ s_{i.0} \models_{Inst} \alpha_i)$

we observe its equivalence with

(22') $B_{Inst} P_{\{Inst,Const\}} s_{i.2} \models_{Const} \alpha_i$ & (9)

due to (e).
Since (9) has already been derived (essentially by means of the rationality assumption (10)) it remains to be shown how to derive

(23) $B_{Inst} P_{\{Inst,Const\}} s_{i.2} \models_{Const} \alpha_i$

from (21). With (23) we count among the preconditions of Inst's move of acceptance his awareness of the fact that it is public information for him and Const that things are arranged according to α_i on Const's side of the screen, i.e. the fact that there is some accessible and perspicuous situation indicating $s_{i.2} \models_{Const} \alpha_i$ to him. Since $s_{i.3}$ is such a situation and Inst is among its participants, it is not hard for him to arrive at this awareness. Besides assumption (*) and (a) - (d) (cf. 1.3) the requisite rationality assumption is

(24) $Rat_{Inst} P_{\{Inst,Const\}} s_{i.2} \models_{Const} \alpha_i$,

i.e. the supposition that Inst is rational enough to proceed from *reason to believe* that it is public information among Inst and Const that $s_{i.2} \models_{Const} \alpha_i$ to the corresponding *actual belief*.[11] Note that believing that $s_{i.2} \models_{Const} \alpha_i$ is public information is belief in the existence of an accessible and perspicuous situation indicating $s_{i.2} \models_{Const} \alpha_i$ to all its participants. Such belief is first-order. Note further, that (21) and assumption (*) are sufficient to derive $R_{Const} s_{i.2} \models_{Const} \alpha_i$, $R_{Inst} s_{i.2} \models_{Const} \alpha_i$, and $P_{\{Inst, Const\}} s_{i.2} \models_{Const} \alpha_i$, with arbitrary additional prefixes of the operators R_{Inst}, R_{Const} and $P_{\{Inst, Const\}}$. Like in the case of the informational content of situation $s_{i.1}$ (cf. (12) *vs* (13)), mutuality hierarchies arise in connection with the present case of situation $s_{i.3}$ (cf. (21) *vs* (23)) only with respect to notions of logical epistemology (having-reason-to-believe-that and being-public-information-that). They do not arise with

[11] Existential generalisation leads from (21) to $P_{\{Inst,Const\}} s_{i.2} \models_{Const} \alpha_i$. By iteration of the $P_{\{Inst,Const\}}$-operator (cf. 1.3), we obtain $P_{\{Inst,Const\}} P_{\{Inst,Const\}} s_{i.2} \models_{Const} \alpha_i$ and proceed from here as shown in 1.3 to $R_{Inst} P_{\{Inst,Const\}} s_{i.2} \models_{Const} \alpha_i$. (24) and Def. 2 lead immediately to (23).

respect to the notion of believing-that from the realm of psychological epistemology, which is causally relevant for intention and action. Therefore, again no paradox threatens. (The relation between $s_{i.3}$, Inst and the informational content $s_{i.2} \models_{Const} \alpha_i$ in the antecedent at stage (IV) is strictly parallel to the relation between $s_{i.1}$, Const and the informational content $s_{i.0} \models_{Inst} \alpha_i$ in the antecedent at stage (II). This is because both $s_{i.3}$ and $s_{i.1}$ are common situations, not private ones.)

According to clause (IV) we stipulate that (20) and (22) are jointly sufficient for Inst's (intending to) accept that α_i is realised on both sides of the screen. As a consequence of this acceptance (as a socially overt speech-act), a common situation $s_{i.4}$ holds (*common* in the sense of being accessible and perspicuous for *both* Inst and Const) indicating to all its participants that Const's private situation of activity indicates α_i to Const (i.e. $s_{i.2} \models_{Const} \alpha_i$) and Inst's private situation of perception indicates α_i to Inst (i.e. $s_{i.0} \models_{Inst} \alpha_i$). In other words, we have:

(25) $Hs_{i.4}$ and
(26) $LA[\{Inst,Const\}, s_{i.4}, s_{i.2} \models_{Const} \alpha_i \ \& \ s_{i.0} \models_{Inst} \alpha_i]$.

This concludes the processing of the four stages at cycle i associated with step α_i.

Note that we have presupposed in this exposition that (27) and (28) hold for all k ($1 \leq k < i$):

(27) $Hs_{k.4}$ and
(28) $B_{Inst}P_{\{Inst,Const\}}(s_{k.0} \models_{Inst} \alpha_k \ \& \ s_{k.2} \models_{Const} \alpha_k)$.

In order, now, to go through the four stages of the next cycle i+1 (associated with step α_{i+1}) in the same fashion, we must make sure that (27) and (28) hold as well for k = i. Because of (25) this is trivially true for (27), and we can show that it is true of (28) as well by assuming the rationality condition

(29) $Rat_{Inst}P_{\{Inst,Const\}}(s_{i.0} \models_{Inst} \alpha_i \ \& \ s_{i.2} \models_{Const} \alpha_i)$

i.e. the supposition that Inst is rational enough to proceed from *reason to believe* that it is public information among Inst and Const that α_i is realised on both sides of the screen ($s_{i.0} \models_{Inst} \alpha_i \ \& \ s_{i.2} \models_{Const} \alpha_i$) to the corresponding *actual belief*.[12] Like in the cases of the informational content of the situations $s_{i.0}$ (cf. (8) *vs* (9)), $s_{i.1}$ (cf. (12) *vs* (13)), $s_{i.2}$ (cf. (16) *vs* (18)) and $s_{i.3}$ (cf. (21) *vs* (23)) arbitrary iterations of operators (e.g. mutuality hierarchies) arise in the case of $s_{i.4}$ (cf. (26) *vs* (28)) only with respect to the notions of having-reason-to-believe-that and being-public-information-that. They do not arise with respect to the notion of believing-that, since higher-order beliefs quickly require extreme sophistication in transforming reasons to believe into actual belief. Additionally, they require extreme confidence in attributing such sophistication to co-agents. (Actually, the derivation of higher-order beliefs is in need of additional rationality assumptions at each new level.) In the antecedents of the four stages of our schema (including the antecedent of their

---

[12] Existential generalisation leads from (26) to $P_{\{Inst,Const\}}(s_{i.0} \models_{Inst} \alpha_i \ \& \ s_{i.2} \models_{Const} \alpha_i)$. By iteration of the $P_{\{Inst,Const\}}$-operator (cf. 1.3) we obtain $P_{\{Inst,Const\}}P_{\{Inst,Const\}}(s_{i.0} \models_{Inst} \alpha_i \ \& \ s_{i.2} \models_{Const} \alpha_i)$ and proceed from here, as shown in 1.3 to $R_{Inst}P_{\{Inst,Const\}}(s_{i.0} \models_{Inst} \alpha_i \ \& \ s_{i.2} \models_{Const} \alpha_i)$. (29) and Def. 2 lead immediately to (28) with k = i.

conjunction (cf. (28)) only first-order beliefs are required. This stipulation might be seen as a tentative hypothesis: we suppose that agents engaged in firmly entrenched pattern of interaction (like giving and following directives) need not develop higher-order beliefs (about beliefs of their co-agent).[13]

Opting for first-order beliefs in the preconditions is, in a sense, the strongest choice. We do not deny that, in general, this hypothesis might be too restrictive. It is very well feasible that - especially with respect to certain less firmly entrenched patterns of interaction - other-directed beliefs and ascriptions of self-directed belief to others turn out to be indispensible. In such cases, one might rely on empirical arguments to the effect that missing higher-order beliefs cause step-backs, side sequences or repairs in the interactional pattern. In general, we hypothesise that the degree of higher-order beliefs required is inversely linked to the degree of entrenchement the pattern of interaction has gained. As soon as social conventions stronly license expectabilities, the agents are, so to say, exonerated from sustaining complicated higher-order convictions (concerning beliefs and convictions of others) about the next steps of their interaction. Assuming that the interactive pattern of giving and following directives is deeply entrenched, we suppose that - instead of sustaining higher-order beliefs - it is sufficient for Inst and Const to establish accessible and perspicuous situations encoding public information. The content of these situations may, in principle, be exploited *ad libitum* but, in practice, it will be expoited only to a very limited extent. Accessible and perspicuous significant situations provide infinite informational resources (in the sense of logical epistemology). They license epistemic mutuality hierarchies based on reasons to believe. They also, however, provide finite informational resources (in the sense of psychological epistemology). Thus, they lead to cognitive states of agents effectively causing their intentions and actions.

Bibliography

Aczel, P., (1988), *On Non-well-founded Sets.* CSLI Lecture Notes Number 14

Barwise, J., (1988), Three views of common knowledge. In M. Vardi (ed.), *Theoretical reasoning about knowledge II.* Los Altos: Morgan Kaufman, 365-379

Barwise, J. (ed.), (1989), *The Situation in Logic.* CSLI Lecture Notes Number 17

Barwise, J. (1989), On the Model Theory of Common Knowledge. In J. Barwise (ed.), *The Situation in Logic,* 201-221

Barwise, J. & Etchemendy, J. (1987), *The Liar.* An Essay on Truth and Circularity, OUP, New York

Clark, H.H. & Marshall, C. R. (1981), 'Definite reference and mutual knowledge', in A.K. Joshi et al. (eds.), *Elements of Discourse Understanding.* CUP, Cambridge, 11-63

Heydrich, W. & Rieser, H. (1995), *Public information and mutual knowledge*, Report 95/12, SFB 360, Situierte Künstliche Kommunikatoren, Univ. Bielefeld

Lewis, D. K. (1969), *Convention.* A Philosophical Study. Harvard Univ. Press, Cambridge, Mass.

Lewis, D. K. (1973), *Counterfactuals.* Blackwell, Oxford

---

[13] Note that the stipulation of any finite degree of belief (the assumption of any finite iteration of belief-operators in the antecedent parts of (I) - (IV)) would do the job of avoiding the embedding paradox of mutuality (i.e. the paradox of ascribing to finite minds infintely many and/or infinitely complex intentional states as causes for overt behaviour).

THE TWO SIDES OF A COIN

Content and Form in Natural Language Generation

Wolfgang Hoeppner

Gerhard-Mercator University Duisburg
FB3 • Computational Linguistics
47048 Duisburg
email: hoeppner@unidui.uni-duisburg.de

ABSTRACT

In this paper we investigate the architecture of natural language generation. Modular approaches in Artificial Intelligence and Psycholinguistics are briefly introduced. We then pick out one domain of discourse, namely the presentation of route descriptions. This leads to the notion of 'interdependency', which we claim is a proper mode of interaction between the modules responsible for content selection and form generation. We follow up earlier debates on these issues and argue for a communicative view on the text generation process.

1. INTRODUCTION

For several centuries the generation of linguistic utterances has been a neglected area by those disciplines which would be supposed to work on this issue. Theoretical Linguistics discarded any commitments to the actual processing of language by inventing the notion of an 'idealized speaker/hearer', thus throwing not only pragmatics out of the realm of Linguistics proper, but also by inventing a sort of artificial objective for scientific research (we don't have to quote Chomsky here, do we?). Psycholinguistics has been engaged in numerous lucid experiments, and still is, to find out, what really is at the bottom (and brain) of the human linguistic faculties. These experiments are often conducted with test-persons who exhibit a certain strange behaviour with regard to linguistic abilities. This is a similar approach as in Neuroscience which tries to investigate the functioning of the brain by comparing deficient brains with 'regular' ones. Both, Psycholinguistics, as well as Theoretical Linguistics, have not devoted much attention to language production. An account of an abstract system of linguistic faculties has been the goal of the latter one; understanding linguistic utterances has been the main research program of the former discipline because this human feature is more easily accessible by empirical methods.

Research in Computational Linguistics, and Artificial Intelligence (the CL&AI approach we will call it subsequently) has shed new light on matters to be studied. By the upcome of powerful computational facilties, new methods arose which enabled researchers to investigate human capabilities from a different angle: simulation of behaviour produces a scientifically similar result as in the natural sciences, e.g. physics. Experiments can be conducted which shed light on the object of inspection by way of equalizing the properties of the domain under observation with the designed model. In language generation, CL&AI have been rather reluctant to tackle these issues. The starting enterprise might be seen in workshop at Stanford in 1981 which was documented in (Mann 1982). We do not want to argue about the prospects then envisaged, which by now should have been established if the participants had been able to foresee future developments. Clearly they could not, and so today one might consent that at least the main problems in generation are known.
We rather want to discuss issues which are at the center of a process description of natural language generation: the division of labour between conceptual layers and linguistic layers, and, in particular, the intricate mode of interaction among these layers.

We will do this by first sketching the classical view of CL&AI at language generation, which is, basically, a bipartite seperation between processes which determine the content of utterances, and those which determine their linguistic form. We then discuss a special area of application, namely the generation of route descriptions. This finally gives rise to some more general statements about language generation and the proper treatment of Semantics (and Pragmatics, if this is something different) in a practical theory of Linguistics.

2. PROCESSING MODELS FOR LANGUAGE PRODUCTION

In the early stages of computational models of generation, i.e. in the 80s, the task to be performed was broken down into two subtasks: the first was to determine which content should be verbalized, and the successive process would then transform ,bits of content' (propositions, to keep to a more scientific terminology) into linguistic units (phrases). One of the subcomponents of the second step was called ,realization', and was dedicated to produce a well-formed surface utterance. Fig. 1 gives a sketch of this breakdown of the generation task. The first step was often called the ,What-to-say' component, the second one the ,,How to say it' component (cf. McKeown 1985, Busemann, Novak 1993). By importing military terminology, which perhaps is not too far fetched, these two stages were often called the ,strategic component' and the ,tactical component'.

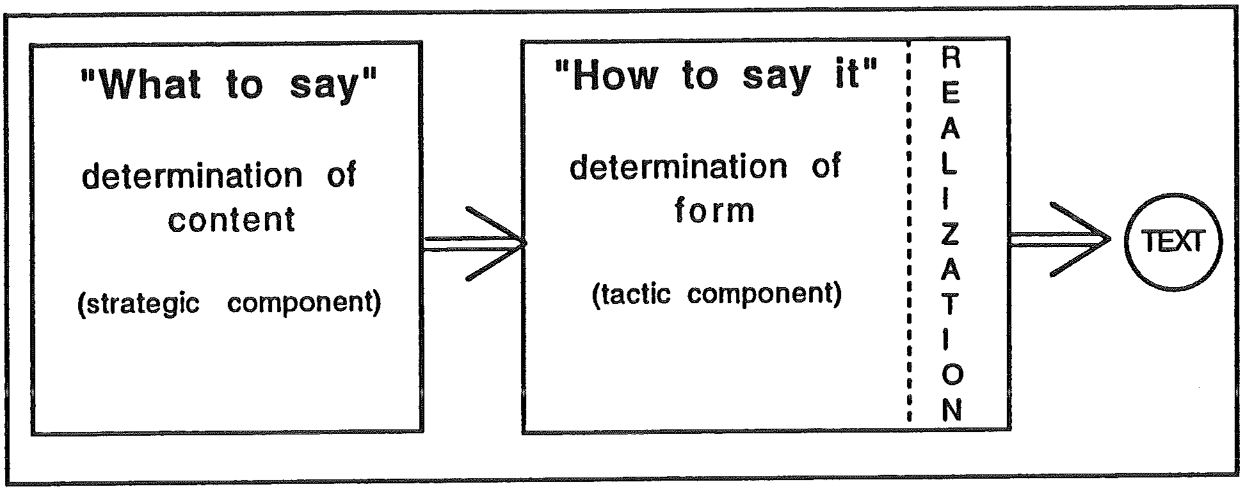


Fig. 1: The classical view on generation

It should not be left unmentioned that this simple view of the generation process soon found critics. Hovy was one of the most prominent ones who argued for a third component, the ‚why to say something' part (Hovy 1988). He thus introduced a pragmatic aspect to the generation process. This was acknowledged by some, ignored by others. Nowadays there seem to exist two approaches to language generation: one is concerned with the linguistic issues (grammars, lexicon, morphological mechanisms, etc.), the other one is constructing components for applications in restricted domains, e.g. explanation facilities for expert systems (cf. Paris 1993).
A more general approach to different architectural designs has been investigated and compared by (De Smedt et al. 1996) in a recent paper.

If we look at research from other disciplines, namely those of Psycholinguistics, we find a model which is similar to the classical view on language generation briefly sketched above. Levelt (1989) gives an architecture of human language produduction which is more elaborate than the CL&AI approach but which makes basically the same distinction between content and form. Fig. 2 gives a simplified version of the way he breaks down the task of generating natural language utterances into three independent components. The ‚conceptualizer' produces a mental message, the *preverbal message* as it is called which is then transformed by the ‚formalizator' into a *phonetic plan*. This is input to the ‚articulator' which finally produces *overt speech*. Acoustics has until recently been out of the realm of CL&AI approaches, and so the classical view seems to match nicely with psycholinguistic results. And, Computational Linguistics approaches have adopted these results for their architecture decisions (e.g. the SYNPHONICS project, cf. Abb et al. 1996).

The main deficiency in such approaches is the reluctance with which they treat pragmatic issues. Levelt uses *discourse model, situation knowledge, encyclopedia etc.* as the input of the conceptualizer without taking much care about what ‚discourse model', ‚situation knowledge', ‚encyclopedia' and especially ‚etc.' might be. A further characteristic may be highlighted, namely, that the final results of the generation process are sentences, not utterances, or even texts. This, though in accordance with Theoretical Linguistics, is not what people really do, nor with what computers should do. Human beings produce texts, in a broad variety of course, but they do not produce ‚well-formed sentences', as defined by a specific theory. And computers should not only produce well-formed sentences, but rather well-formed output in the sense that it might be appropriate for, and helpful to human users.

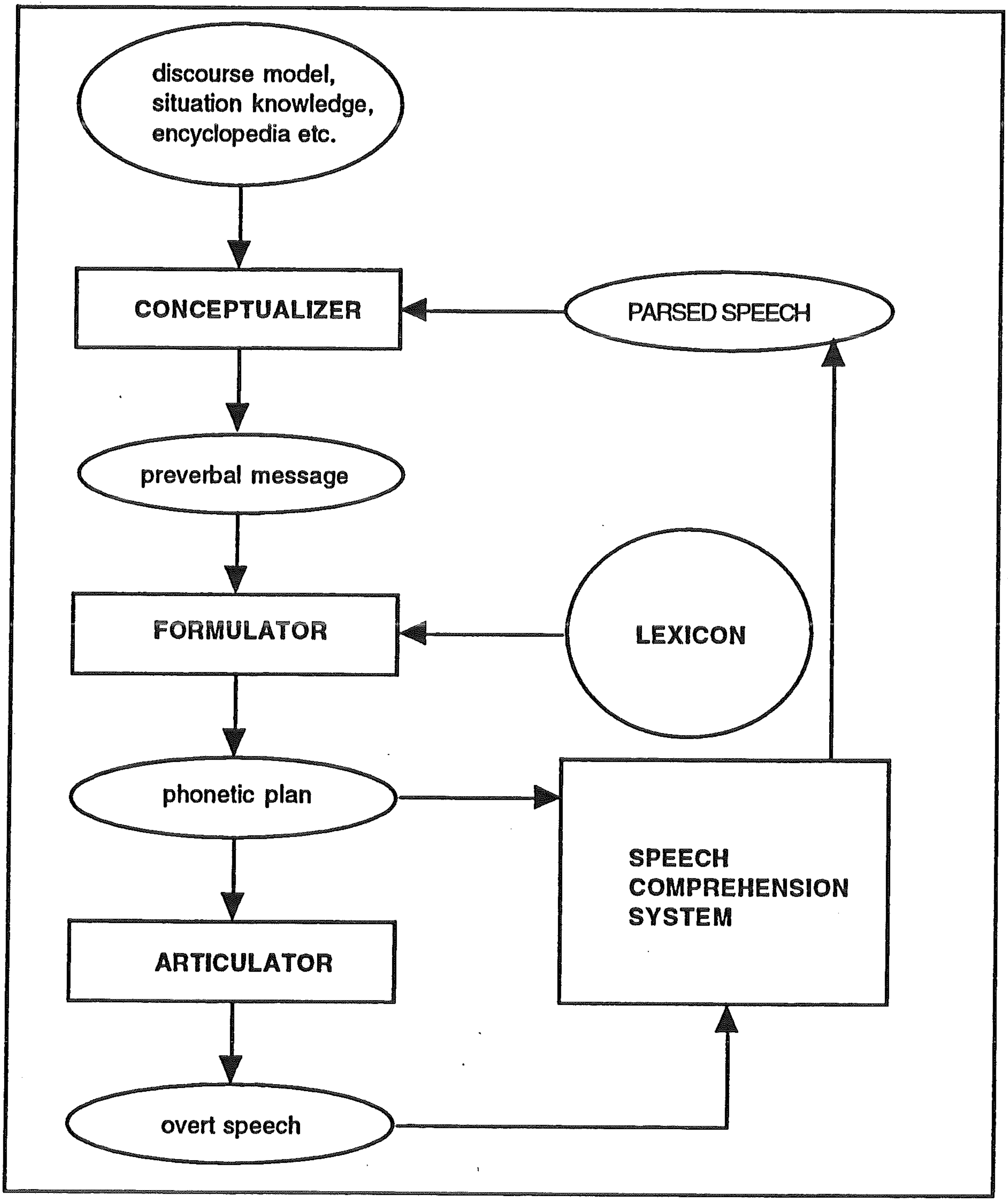


Fig. 2: A broad sketch of Levelt's architecture

Having thus discussed, in a rather theoretical way, the different approaches to generation, we will now turn to a specific domain of discourse. This shall shed light on the position we are proposing in this paper.

3. Generating Route Descriptions

As mentioned above, texts have been a neglected area in Linguistics research since the upcoming of formal theories which were and still are devoted to sentences. However, properties of texts have been studied by empirical linguists (e.g. Grimes 1975), as well as Computational Linguists (e.g. Mann, Thompson 1988). the latter ones being mostly concerned with generation issues rather than with those of analysis.
Among these, route descriptions have been a domain which has been investigated rather intensively. Empirical work has been conducted by (Klein 1979, 1982, 1983, Wunderlich, Reinelt 1982), computational models have been proposed by others (cf. Carstensen 1988, Müller 1989, Rau, Schweitzer 1987, Ruhrberg, Rutz 1990). The general framework in these research efforts has been to establish a process model of the human language production capability. The result is most often similar to the model of Levelt which we briefly introduced in the preceding section: for the specific task at hand there is a component which determines the route to be described, followed by a component which transforms this ‚content' into a sequence of linguistic sentences. Habel (1988) presents this approach on a general level.

In this chapter we would like to propose a different view while exploring the pragmatic conditions of route descriptions. This is not meant as a further contribution to the specific communicative task, but rather as a new perspective on the generation process which, as it will turn out is a hearer-oriented perspective.

3. 1 The specific situation

When we look at the communicative constraints for route descriptions we have to deal with a number of different and often incompatible requirements (cf. Hoeppner et al. 1990).

Some of these are:

- the route should be a short one (both in space and time)
- the description should be easy to memorize by the hearer
- the description, or rather its mental counterparts, should easily be mapped onto reality.

These constraints belong to different sources of knowledge. They refer to properties of the domain of discourse (a representation of geographic facts), to linguistic properties (the complexity, and thus the structure of texts), and to cognitive properties, i.e. mental representations of verbal utterances. The speaker has to decide on all of these constraints, and usually does so without being aware of it.

To clarify these issues let's take a look at a concrete example. Fig. 3 contains a rather broad sketch of the street map of Duisburg, the area between the central main station and the university. If someone at the station came up with a question such as ‚How can I get to the University?' this or a similar representation would serve as a basis for generating a linguistic description of a route. Of course, there is a variety of different routes which can be taken, those which make up the minimal spatial distance, those which prefer a simple description without too many turns, those which would suggest public transport, etc. Choice among these parameters is not an easy task because it involves reasoning about the hearer's understanding and memorizing capabilities.

Most of the computational approaches mentioned above have made use of Dijkstra's algorithm for finding shortest paths in graphs (Dijkstra 1959). From a theoretical point of view this is a solid foundation. However, the algorithm relies on an ‚evaluation function' which finally yields a numeric value. This method might be adequate for a formal approach, where the costs of traversing each transition are known and computable. This is hardly possible for the task of finding adequate routes for the purpose of giving a verbal description, because there are a number of restrictions, which are not expressible in numeric terms. Properties of real-world traffic situations have been investigated for the purpose of finding optimal routes by (Mertens 1991). In his implementation spatial distances are mapped onto temporal constraints, thus allowing an elegant integration of public transport facilities. Here we have a single metric, namely a temporal one which allows the use of graph algorithms.

For a pedestrian unaquainted with the geographic specifics of Duisburg, the route depicted with a double line in fig. 3 would be a good solution, for a car driver a different route would be more appropriate. How might texts look which describe this physical environment? Fig 4a and 4b give some possibilities describing the same matter of affairs. Although these are very different descriptions of the same route, it is not easy to pin down these differences to a representation of the geographic setting. The text of 4b might sound a bit verbose, containing too many unnecessary details for the purpose at hand. On the other hand, the description of 4a is rather short as compared to the distance one has to walk. The hearer might easily be led astraiy when remarkable sites were not mentioned in the description.

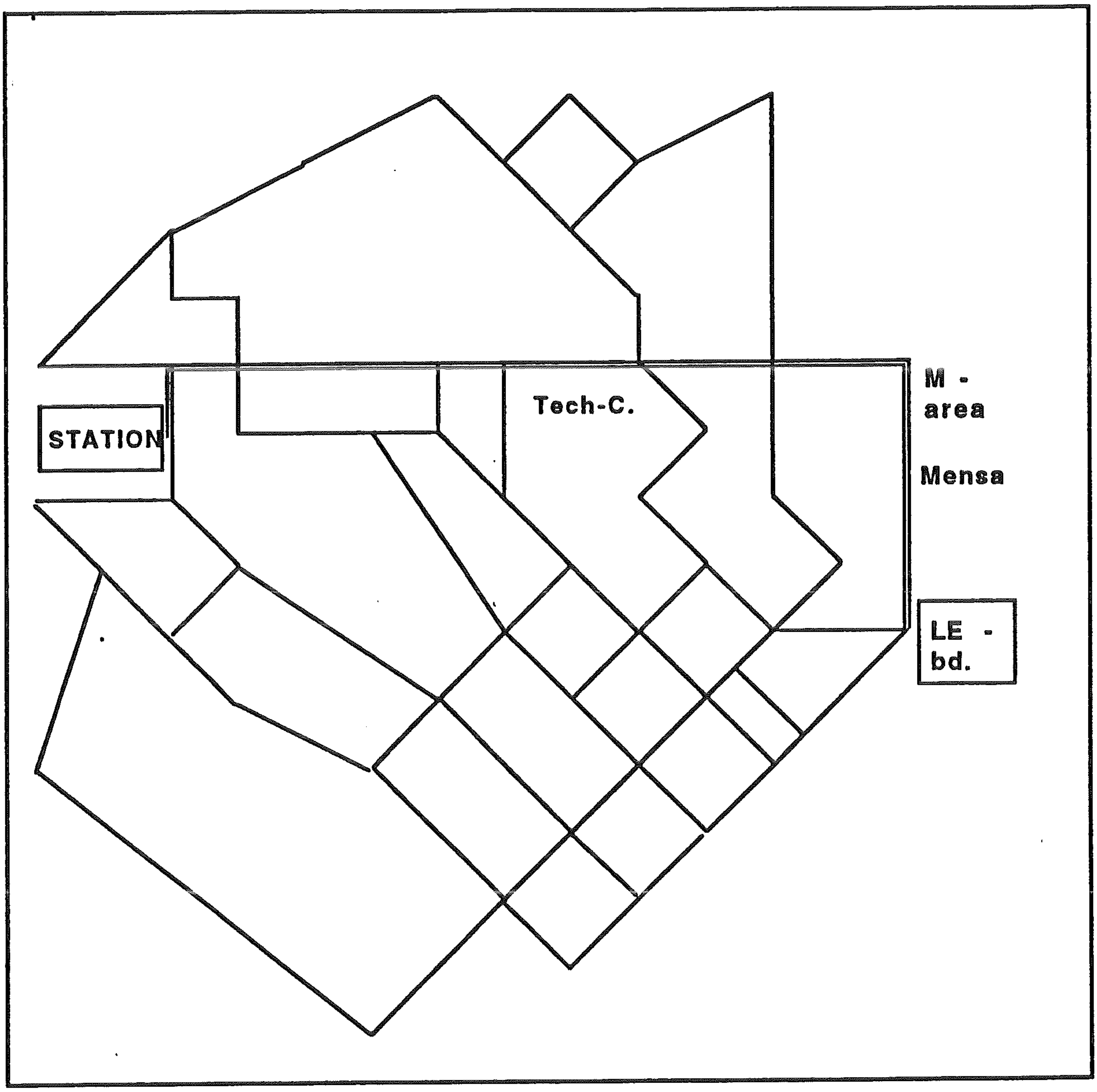


Fig. 3: Sketch of (parts of) Duisburg

You'll best take the Mülheimer Straße and go straight ahead until you reach the Lotharstraße. Then turn right and straight ahead to the Forsthausweg. And there you are.

Fig 4a: A scarce description of the route from Fig. 3

Leave the main station through the east entrance and turn left. There you hit a main street, the Mülheimer Straße. It is called that way because it leads to Mülheim. There turn right and go straight ahead for quite a while. When you reach the Lutherplatz, you will notice on the left hand a futuristic building. This is the Technology Centre, which opened in 1991. You still go straight ahead - the street ascending slightly - until you reach the Lotharstraße, where you turn right. On the opposite side, you will notice the so-called 'Keksdosen' (i.e. cookie jars), which are named according to their appearances. There are located the natural sciences. Along the Mensa (the blue rotunda) and a tennis court you'll reach the Forsthausweg where you turn left. On the right side, in the bend, there is the LE-building, where you can find the principal and the chancellor of the University.

Fig 4b: A rather verbose decription of the route from Fig. 3

To shift to a more general level, it is not a mere point of mapping the conceptual structure of a speaker to his (or her) linguistic utterances, but it is a matter of matching the speaker's representation of conceptual entities with those attributed to the hearer, however contested these attributions might be. This has been noted before (the most prominent quotation is, of course (Grice 1975), and therefore this is one of the most frequently quoted articles in the Natural Language Generation literature). The problem lies in the way such factors can be acquainted in man-machine communication. This pragmatics based approach to the modularization of the generation process we have called 'interdependency' (cf. Hoeppner 1995): of course linguistic form is dependent on a content to be verbalized, nobody would doubt that. But content selection is also guided by the way a route is to be presented through linguistic means, if one adopts a hearer-centered perspective.

We have proposed a solution in this specific domain through a refinement of the basic data of a map representation and a layered architecture of the necessary representational prerequisites (cf. Carstensen et al. 1991). Such approaches can be found in many research efforts, which have to cope with the intricacies of specific domains. A layered approach is capable to deal with granularity problems, and seems thus to be a more promising alternative to the uniform approach formal logical calculi tend to provide.

3.2 Criticizing Critics

The pragmatic approach outlined above (for more details cf. Hoeppner et al. 1990) has been the starting point of a discussion which we would like to continue here. Carstensen (cf. Carstensen 1992) has tried to mediate betweeen the strictly modular view of other researchers and the approach of radical pragmaticism in Natural Language Generation which we ourselves like to subscribe to. While acknowledging arguments in favour of an ‚Interdependency hypothesis', Carstensen introduced ‚conceptual knowledge' as a mediator between spatial and linguistic knowledge. This maneuver still clings to the assumptions of Theoretical Linguistics, which regards the appropriateness of utterances solely to the linguistic abilities of the speaker.

What belongs to the abilities of a speaker is, however, also the ability to judge the cognitive capabilties of the recipient, and the communicative factors which in every-day language use determine the linguistic form of utterances. In the domain of route descriptions it would not be a good decision to describe elements of the route in terms of features, which are not observable while making use of the description (e.g. due do darkness). And neither would it be a successful strategy to talk extensively about issues which do not contribute to the specific task at hand. Thus, talking about a specific building might be appropriate in certain settings, it might not be in others where the intention of the recipient is directed towards a different purpose.

To illustrate this with a particular example (cf. fig. 3 and 4b): the mentioning of the Technology Centre in a description of a route from Duisburg's main station to the University is a helpful (and thus cooperative as computer generated texts should be) with respect to the task at hand. The route is rather long, takes about 20 to 30 minutes to walk. And, for the purpose of reassurance, mentioning this building is a useful item to include into the description. BUT: this is no property of the geographic environment, nor is it one of linguistic competence. Rather, it is crucial to determine the functions of utterances, because they determine what is said and what is left unsaid.

Our critique to the position of (Carstensen 1992) and others engaged in the field of natural language generation can be summarized as follows:
There is no sound solution to the task of language production without taking into account the recipient as a serious and necessary device. It might be compelling to model only the part of a speaker, it might turn out to be a simpler process model. But doing so one is simply ignoring the fact that any production of language is dominated by a more or less concise conjecture of the potential or even real recipient(s). Throwing out particulars of the domain under inspection has never been a method for successively coping with these phenomena, at least not in the long run. Theoretical Lingusitics has been obliged to such an endeavour, but Computational Linguistics should not fall into the same pit.

4. CONCLUSION

We have argued for the integration of a hearer-centered approach in the generation of linguistic utterances, be it in the realm of Theoretical Linguistics or Computational Linguistics. It is the communicative function of linguistic utterances, not the properties of a linguistic system which is the prevaling feature of human language. The linguistic system, the regularities of a specific language of course are worth while for thorough investigations. But linguistic knowledge treated as such, i.e. as a closed environment, does not seem to have any resemblance to ‚real-world' matters. Of course, one might study such artificial domains – mathematicians have done this ever since, and, they have successfully done so – there is no need for any science engaged in natural domains, i.e. those who work on topics of everyday life, to establish a simplified version of what is being studied.

For natural language generation this means that the communicative function should be the point of departure. We do not claim that this is a brand new insight into the matter: Wittgenstein among others already has dealt with these issues. (cf. Wittgenstein 1960). These issues, however, have mostly been discussed in meta-theoretical debates (Yorick Wilks, for instance, has put forward such topics from ‚philosophy of mind' considerations during a summer school in 1990 dealing with symbolic and subsymbolic computation). But they are of utmost importance to process models of the human linguistic capabilties.

At last, it is a matter of what one wants to investigate, what one wants to clarify using scientific methods. If it is human behaviour, then notions of ‚intention', ‚goal-directed action' and the like should be taken seriously. If it is only some part of human behaviour one has to convincingly claim that that part can be studied on its own. In language production this does not seem to be the case, not even in any restricted area. So, why not treat the object domain in a way which tries to account for all its features?

REFERENCES

Abb, B., Günther, C., Herweg, M., Lebeth, K., Maienborn, C., Schopp, A.: Incremental grammatical Encoding - An Outline of the SYNPHONICS Formulator. In: Adorni, Zock (eds) 1996, 277-299

Adorni, G., Zock, M. (eds): Trends in Natural Language Generation. An Artificial Intelligence Perspective. Berlin: Springer 1996

Busemann, St., Novak, H.-J.: Generierung natürlicher Sprache. In: G. Görz (ed): Einführung in die Künstliche Intelligenz. Bonn: Addison Wesley 1993, 499-558

Carstensen, K.-U.: WebS - ein System zur Generierung von Wegbeschreibungen. IBM Deutschland. LILOG-Memo 11, Stuttgart, September 1988

Carstensen, K.-U.: Finding Adequate Routes for the Generation of Route Descriptions. In: G. Görz (ed): KONVENS 92. Berlin: Springer 1992, 309-318

Carstensen, M., Groß-Hardt, M., Hoeppner, W., Kölzer, A., Masermann, U., Rhein, U.: Repräsentationskonzepte für die Generierung von Wegauskünften im Projekt KOPW. Univ. Koblenz-Landau, Fachberichte Informatik 21, 1991

De Smedt, K., Horacek, H., Zock, M.: Architectures for Natural Language Generation: Problems and Perspectives. In: Adorni, Zock (eds) 1996, 17-46

Dijkstra, E.W.: A Note on Two Problems in Connexion with Graphs. Numerische Mathematik 1, 1959, 269-271

Freksa, C, Habel, C.: Repräsentation und Verarbeitung räumlichen Wissens. Berlin: Springer 1990

Grice, H.P.: Logic and Conversation. In: P.Cole, J.L.Morgan (eds): Speech Acts. New York: Academic 1975, 41-58

Grimes, J.E.: The Thread of Discourse. The Hague: Mouton 1975

Habel, C.: Prozedurale Aspekte der Wegplanung und Wegbeschreibung. In: Schnelle, H., Rickheit, G. (eds): Sprache in Mensch und Computer. Wiesbaden: Westdeutscher Verlag 1988, 107-133

Hoeppner, W.: Modularity in Natural Language Generation. Proceedings of the 4th International Colloquium on Cognitive Science, 1995, 115-121

Hoeppner, W., Carstensen, M., Rhein, U.: Wegauskünfte: die Interdependenz von Such- und Beschreibungsprozessen. In: Freksa, Habel (eds) 1990, 221-234

Hovy, E.H.: Generating Natural Language under Pragmatic Constraints. Hillsdale: Erlbaum 1988

Jarvella, R.J., Klein, W. (eds): Speech, Place, and Action. Studies in Deixis and Related Topics. New York: Wiley 1982

Klein, W.: Wegausjünfte. Zeitschrift für Literaturwissenschaft und Linguistik 9, 1979, 9-57

Klein, W.: Local Deixis in Route Directions. In: Jarvella, Klein (eds) 1982, 161-182

Klein, W.: Deixis and Spatial Orientation in Route Directions. In: H.L. Pick, L.P. Acredolo (eds): Spatial Orientation. Theory, Research, and Application. New York: Plenum 1983, 283-311

Levelt, W.J.M.: Speaking. From Intention to Articulation. Cambridge: MIT Press 1989

Mann, W.C.: Text Generation. Americam Journal of Computational Linguistics, 8, 2, 1982, 62-69

Mann, W.C., Thompson, S.A.: Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. Text 8,3 1988, 243-281

McKeown, K.R.: Text Generation. Using Discourse Strategies and Focus Constraints to Generate Natural Language Text. Cambridge: University Press 1985

Mertens, A.; Räumliche und zeitliche Aspekte in Wegauskunftssystemen. Universität Koblenz-Landau, Diplomarbeit, 1991

Müller, S.: CITYGUIDE - Wegauskünfte aus dem Computer. Univ. des Saarlandes, Sonderforschungsbereich 314, Memo Nr. 37, 1989

Paris, C.L.: User Modeling in Text Generation. London: Pinter 1993

Rau, A., Schweitzer, M.: Untersuchungen zur Verwendung script-artiger Wissensrepräsentationen und Story Grammars bei der Generierung von Wegbeschreibungen. Duisburg: L.A.U.D. Bericht 158, Januar 1987

Ruhrberg, P., Rutz, H.: Räumliches Wissen und Semantik im Kontext der Generierung von Wegbeschreibungen. In: Freksa, Habel (eds) 1990, 235-249

Wittgenstein, L.: Philosophische Untersuchungen (1953). In: Schriften 1, Frankfurt/M.: Suhrkamp 1960, 279-544

Wunderlich, D., Reinelt, R.: How to get there from here. In: Jarvella, Klein (eds) 1982, 183-201