From Conceptualization to Formulation in Generating Spatial Descriptions

Christian Ebert Daniel Glatz Martin Jansche
Ralf Meyer-Klabunde Robert Porzel

{chris,daniel,martin,ralf,robert}@pc03.idf.uni-heidelberg.de

Ruprecht-Karls Universität Heidelberg
Computerlinguistik
Karlstraße 2
69117 Heidelberg

SFB 245 "Language and Situation"
Tp A7 "Implicit Information"
Plöck 55
69117 Heidelberg

1 Introduction

Architectures of the language production system are usually subdivided into three processing modules, which are usually termed conceptualizer, formulator, and articulator (LEVELT 1989). Generally, these three subsystems compute specific representations as input for the next subsystem. Within the conceptualizer content selection, preparation, and linearization takes place. The linearized conceptual representation—the "preverbal message", in LEVELT's terms—is handed on to the formulator. Planning the preverbal message is crucially influenced by the speaker's beliefs of what the addressee beliefs, what their goals, plans, and intentions might be. Additional knowledge sources a speaker considers while planning the message are the spatial representation, the discourse protocol which keeps track of what has been said, additional situative beliefs, and encyclopedic knowledge. The formulator take the preverbal message as input and computes a phonetic plan as output by means of grammatical and phonological encoding. Finally, the phonetic plan triggers the articulator.

Although the processes and representations that have been assigned to the conceptualizer are decisively driving the whole production process, relatively little is known about them, compared with the number of works concerning the formulator and articulator. Existing insights into conceptualization processes have also often been investigated independent of each other. Therefore, a comprehensive modeling of conceptualization processes in language production can shed some additional light on these 'hidden' processes which are accessible only indirectly, e. g. through linguistic data. Hence, in modeling the cognitive mechanisms for conceptualizing, an interaction of the investigated factors must be worked out, additional assumptions about specific structures and processes have to be postulated where necessary, and the adequacy of the model must be tested by means of an implementation.

In planning the content of spatial descriptions several decisions have to be drawn, concerning e. g. the points of view the speaker should take, the strategies for describing the object constellation, or the function of an object as reference or primary object, which will be discussed in detail below.

We will proceed as follows. After presenting the problems that appear in the generation of spatial descriptions, we will provide some empirical results with respect to these problems. Based on this discussion we will describe the current state of our computational model of the production of spatial descriptions and the hypotheses which have been derived from the model.

Ideally, the hypotheses we have formulated while elaborating the computational model and the hypotheses generated by the model should be tested by means of empirical studies again, which is currently a pie in the sky.

2 Background

The linearization processes involved in constructing preverbal messages for spatial descriptions comprise

the selection of a linearization strategy for the individual localizations. Three kinds of linearization strategies are identifiable as classified by CARROLL (1993). *Global* strategies employ a frame of reference that superimposes a spatial structure that encompasses the entity under description as a whole (e. g., a Cartesian coordinate system). *Point-by-point* strategies employ spatial structures that are anchored at the individual parts which make up the entity under description. *Imaginary tour* strategies employ a fictive observer that is placed and moved within the scenario; the individual objects are localized in reference to that fictive observer.

the selection of a point of view for the individual localizations. Spatial descriptions are given from a specific point of view, either the speaker's own (egocentric) or from a different point of view (allocentric), which can be that of the listener, an object from the spatial representation, or a fictive observer. Producing allocentric localizations may require mental rotations by the speaker from her own point of view to that of the listener or a third entity.

the sequential ordering of the individual localizations. The order of the individual localizations depends on the linearization strategy.

the construction of localization units. Localization units are propositions which establish a spatial relation, depending on the chosen point of view, between one or more primary object(s) and one or more reference object(s).

2.1 Selection of the Linearization Strategy

A speaker's selection of appropriate linearization strategies to be employed for spatial localizations has been linked to features of the spatial representation. The selection of a linearization strategy is influenced by the spatial constellation of the objects and the kind of scenario represented. Each linearization strategy can be regarded to exhibit individual preferences, that determine which kind of spatial constellation is well suited for which kind of strategy. An imaginary walk strategy, for example, would not be well suited for star-shaped scenarios as it would involve guiding the fictive observer back and forth between a central reference object and the peripheral primary objects. Redundant employments of objects as reference objects should be avoided for imaginary walks. Analogously, a global strategy would only be well suited if the objects to be described exhibit a suitable structure. It is rather unlikely that a speaker will guide a fictive observer, performing an imaginary walk, through a set of object such as cups and books lying on a table, whereas geographic localizations of countries on the globe are well suited for global linearization strategies. The speaker's acquisition of a spatial representation of the scenario, i. e. the cognition phase, may also influence the selection of a linearization strategy. If the spatial representation, e. g., was obtained through a video presentation featuring a tour through the spatial scenario, as in BUHL's (1995) study, an imaginary walk becomes a likely strategy for a later description.

2.2 Selection of the Point of View

BUHL (1995) concluded from her study that a speaker's internal representation of a spatial configuration does not provide 'point of view'-free information so that all spatial relations that can exist between the objects are equally available to the speaker. It is rather bound to the viewpoint the speaker held during the construction of that representation. In BUHL's study subjects learned the localization of the buildings of a toy village by means of a video showing a specific route through the village connecting two target objects, a church and a fountain. When the communicative task was to describe the route along the church and fountain from the direction opposite that previously shown, an overwhelming number of subjects produced an imaginary walk containing a localization of the fountain with respect to the church that was given from the viewpoint established during the cognition 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 from the acquisition phase coincided, although this resulted in a longer tour through the village. BUHL concludes that the point of view from the cognition phase is the dominant one. This is in agreement with PINKER's (1988) and SHOBER's (1993) assignment of a special status to the egocentric viewpoint among all possible points of view.

HERRMANN and GRABOWSKI (1994, ch. 3) are able to show that there are at least two 'pragmatic' reasons for a speaker to abandon her egocentric point of view in favor of the listener's point of view. The speaker estimates the listener's cognitive competence as low. A typical case would be talking to a child whose cognitive competence is deemed insufficient for building an adequate mental model of the scenario from a description given from a point of view other than its own. Additionally, a speaker might estimate the social status of the listener as sufficiently higher to perform the mental rotation for him as a means of being polite.

2.3 The Sequential Ordering of Localizations

One of the central problems in modeling the production of spatial descriptions (and implementing such a model) concerns the way in which the individual localizations should be ordered. Employing only minimal constraints, such as localizing every object only once, each spatial constellation of n objects can be sequenced with at least $\frac{n!^2}{n}$ different possibilities. For a detailed account of this problem see PORZEL ET AL. (1996).

In addition to the aforementioned constraints concerning the linearization strategies' preferences, LEVELT (1982) proposes some principles for relieving the working memory while describing spatial networks. These principles state that the linearization sequence should whenever possible proceed along nodes that have direct connections (principle of connectivity), that a return should go to the last node in the waiting line (stack principle), and that an alternative path should keep the resulting memory load minimal (minimal-load principle).

2.4 Constructing the Localization Units

Regarding the selection of objects as reference and primary objects for individual localization units, the speaker's assessment of the identifiability of the individual objects for the listener, the saliencies of the objects represented, and the path the speaker intends to follow through the scenario constitute important factors for the construction of localization units.

The individual saliencies of the objects represented constrain the construction of localization units insofar as the saliency of the reference object, for example, must be sufficiently high for localizing its primary object. The statement *There is a cathedral behind the bicycle* would violate that constraint. The bicycle would therefore not be considered as a suitable reference object for the cathedral, unless extraordinary circumstances have increased its saliency sufficiently.

Furthermore, as mentioned above, only objects deemed identifiable for the listener by the speaker will be considered and selected as reference objects in localization units. The assumed identifiability depends on the speaker's beliefs of the knowledge the listener has of the spatial scenario, what he can know via visual accessibility, or infer from the previous discourse.

3 From Conceptualization to Formulation in ParOLE

Based on the factors outlined above that influence the planning of the content of spatial descriptions, we are implementing the generation system ParOLE (an acronym standing for "partner-oriented linguistic explication"). Its primary task is to produce spatial descriptions of various scenarios. For now the focus lies on spatial descriptions of village settings. ParOLE possesses a spatial representation of the object constellation at hand that is congruent to the findings from the studies on the influence of the cognition phase. Once ParOLE has obtained a specific spatial representation, it will start to generate spatial descriptions based on that spatial representation in accordance to a specific communicative task and in regards to a specific listener.

3.1 Conceptualization

Until now the focus in implementing the ParOLE system lies on modeling the conceptual processes responsible for the generation of preverbal messages for spatial descriptions. The preverbal message contains the information necessary for encoding by the formulator module. In order to enable the formulator to retrieve the appropriate lexical items, to construct syntactic structures, and to produce a coherent text, the preverbal message must supply information concerning the objects to be localized, the sequential order in which they will be localized, the point of view from which the localizations will occur, the overall strategy for the localizations, and the informational structuring of the individual propositions.

ParOLE is provided with a modular architecture so that it is congruent to the organization of psycholinguistic models of language production. The modules are a spatial representation, the partner model representing assumptions about those properties of its listener that are relevant to producing spatial descriptions, a discourse model, and encyclopedic knowledge. The partner model will be updated as new information is introduced into the discourse.

In ParOLE the so-called *quaestio* (KLEIN & VON STUTTERHEIM 1987) is used as the heuristic device to handle the speaker's intentions and to set certain constraints for the produced discourse and the build-up of the modular knowledge stores. According to the quaestio-model, every text may be interpreted as a complex communicative answer to an abstract textual question, the *quaestio*. The quaestio imposes constraints on the textual structure, both on global and local levels.

The Partner Model ParOLE's partner model is a simple knowledge source which provides the system with information concerning its listener that is relevant to the communicative situation. In case there is no specific information at hand the system relies on default settings representing a generic listener in the sense of DELL and BROWN (1991). The information the partner model contains concerns the system's estimation of his cognitive competence and his social status. The listener's knowledge of the scenario becomes important for the system in terms of knowing which objects are identifiable for him as he has prior knowledge of their location or existence, his alleged cognitive competence and social status as well as his position within or absence from the spatial scenario. The listener can be placed on several positions or be absent from the scenario. This information is supplied by the spatial representation. His cognitive competence and his social status can each be high or low. These properties and their corresponding values are supplied by the partner model via frames.

The Discourse Model ParOLE's discourse model influences the produced text insofar as it contains the information concerning which objects can be referred to by means of anaphora or for the employment of certain information- structural features, such as topicness, and syntactic structures which are sensitive for information-structural features, e. g. left dislocation. It also supplies information concerning which objects are identifiable as they have been already introduced into the discourse at a previous point.

Encyclopedic Knowledge The general encylcopedic knowledge also enables the system to make assumptions whether the listener might be able to identify certain objects as a result of world knowledge assumed to be mutual as in (1).

(1) There is a church behind the town hall.
The church is green, but the steeple is blue.

(2) Behind the town hall is the church.

If the listener was to know that a small village was described, the statement (2) would also be a result of such inferences.

The Spatial Representation We employ a constraint-based approach in processing the relevant knowledge. Spatial objects are represented as atomic entities possessing certain properties. Each object has a set of properties, such as its saliency, position, intrinsic orientation, size and color, that are represented in corresponding frames assigned to the individual objects.

ParOLE is supplied with information concerning the properties of the objects represented which is continuously updated as some of these properties, e. g. the saliency, are subject to changes as the discourse progresses. This information stems in part from the spatial representation, e. g. information concerning the intrinsic orientations of the objects, and from the additional knowledge sources, e. g. information concerning the identifiability of the individual objects supplied by the partner model or information concerning the referential status of the discourse referents supplied by the discourse model.

Figure 1 illustrates a spatial constellation of four object positions P1 through P4 as seen from the system's point of view. In one setting the four objects—in addition to the system's self— are a town hall (th1),

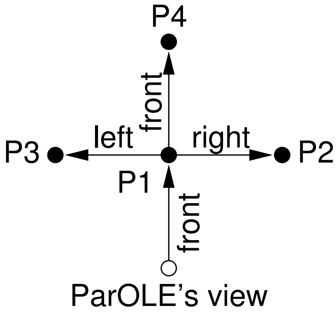

Figure 1: A spatial constellation from the system's point of view

a church (ch1), a fire station (fs1), and a bus stop (bs1). A sample collection of relevant information, as supplied by the individual knowledge stores, concerning the objects at hand which represents the state of affairs at the beginning of the spatial localization task looks as follows:

property	bs1	ch1	fs1	th1	*knowledge store*
position	P1	P2	P3	P4	spatial representation
facing	P4	P1	P1	P1	spatial representation
saliency	50	90	80	100	encyclopedic knowledge
identifiable	no	no	no	no	partner model

The *position*-property tells ParOLE where in its world the individual objects are located. The *facing*-property denotes the position of each object's intrinsic front. The intrinsic orientation of the objects is represented in the spatial representation, as facing a certain direction is one of the central properties of an object, and the system may and will employ the intrinsic sides of buildings or objects for localizations when it decides to localize from their point of view.

The *saliency*-property, represented by a numeric value, changes as the discourse progresses and, as the focus of attention shifts, additional saliency points are bestowed upon individual objects. The scenario presented above features a listener who is not physically present. As the listener can be placed on various positions this can distribute additional saliency due to proximity.

The *identifiability*-property bases on the speaker's estimation of the listener's ability to identify the objects at hand. In case an object is identifiable it can be employed as a reference object for localizing a primary object. An object that has been localized as a primary object becomes identifiable and can therefore be considered for employment as a reference object from then on.

Binary relations are e. g. *neighboring* or *being in front of*. Note that a proposition such as $\text{leftof}'(\text{bs1}, \text{ch1})$ represents the spatial relation that is valid between the primary object 'bs1' (the bus- stop) and the reference object 'ch1' (the church) from the system's point of view only. It is independent of the intrinsic orientations of the objects. Only in those cases where an object is facing the same direction as the system's self does, its intrinsic orientation coincide with the relation embedded in the spatial representation.

Finding the Best Linearization Finding the best linearization with respect to a given spatial representation, partner model, and quaestio could be viewed and implemented as a combinatorial optimization task.

An exhaustive search is, however, neither cognitively plausible nor computationally feasible, as mentioned above, due to the enormous amount of possible sequences for each non-trivial spatial setting.

Therefore, additional constraints imposed e. g. by the linearization strategies' preferences, the distribution of saliency values within the spatial representation, and the quaestio have to be included. But even then it will not be possible to reduce the set of possible linearizations to a reasonable size that would allow for an exhaustive search to be performed. This means that the best linearization sequence must be found by making local decisions only, which is also in congruence with the proposed incremental processing mentioned earlier.

With local decisions that base on the constraints and principles outlined above the system will produce linearizations that avoid redundant sequences for imaginary walks, it will find paths by recourse to the saliency values, general spatial constellation of the scenario, the point of view embedded in the spatial representation, features of the addressee, etc. There is no guarantee that ParOLE will find the globally optimal linearization in each case, but it will find one that is cognitively adequate with respect to existing empirical data and it will, moreover, do so in a cognitively plausible way.

3.2 Formulation

Preverbal messages generated by the conceptualizer are represented as feature structures as e. g. the following example for a single sentence within a larger description:

(3)
$$\begin{bmatrix} \text{ROLE} & \text{located} \\ \text{PrimaryObject} & \begin{bmatrix} \text{CONCEPT} & \text{town\_hall} \\ \text{IDENTIFIABLE} & \text{no} \\ \text{INSTANCE} & \boxed{1} \end{bmatrix} \\ \text{PLACE} & \begin{bmatrix} \text{RELATION} & \text{infront} \\ \text{ReferenceObject} & \begin{bmatrix} \text{CONCEPT} & \text{bus\_stop} \\ \text{IDENTIFIABLE} & \text{yes} \\ \text{INSTANCE} & \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

This representation serves as input for the formulator module. The formulator will map ROLE features onto verbal phrases and CONCEPT features onto noun phrases. These noun phrases will be definite or indefinite depending on their identifiability. The INSTANCE feature bears a discourse marker for each object enabling, e. g., anaphoric reference by recourse to the same discourse marker. The outer features are determined by the type of the ROLE feature. Here, the ROLE expresses a spatial relation between a primary object and a PLACE. The PLACE in (3) is the location in front of (given by RELATION) the reference object bus_stop.

However, a preverbal message such as (3) does not completely determine the syntactic structure of the corresponding sentence. Without any further constraints the following two sentences would be generated.

(4)	Vor	der Bushaltestelle	befindet sich	ein Rathaus.
	in front of	the bus stop	there is	a town hall
(5)	Ein Rathaus	befindet sich	vor	der Bushaltestelle.
	a town hall	there is	in front of	the bus stop

Though containing the same propositional content and the same reference and primary objects, both sentences are not mutually interchangeable in the context of their appearance, as the following example shows:

(6)	Links von	der Kirche	ist	eine Bushaltestelle.
	left of	the church	is	a bus stop
	Vor	der Bushaltestelle	befindet sich	ein Rathaus.
	in front of	the bus stop	there is	a town hall

?? Ein Rathaus	befindet sich	vor	der Bushaltestelle.
a town hall	there is	in front of	the bus stop

Both sentences differ with respect to their constituent order (and the position of the nucleus accent). The selection of these linguistic means is driven mainly by discourse-relational information such as the degree of familiarity obtained by the individual discourse referents or the structuring of the proposition by means of information structural categories.

We assume that formulator-specific constraints determine the final shape of the single sentences. In order to block the inappropriate verbalization in the context (6) given above, one has to rely on the instantiation of the identifiability feature and a *precedence constraint*, which moves phrases with identified discourse referents to the left of phrases with non-identified referents.

Acknowledgments

This research is funded by the German Science Foundation (DFG) in an interdisciplinary research program, the Sonderforschungsbereich 245 "Language and Situation," project A7 "Implicit Information."

References

BUHL, H. M. (1995) Erwerbssituation, mentale Repräsentation und sprachliche Lokalisationen: Blickpunktinformation als Bestandteil der Raumrepräsentation. Submitted.

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

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/NJ: Lawrence Erlbaum; 105–129.

HERRMANN, T. & J. GRABOWSKI (1994) *Sprechen: Psychologie der Sprachproduktion*. Heidelberg: Spektrum.

KLEIN, W. & VON STUTTERHEIM, CH. (1987) Quaestio und referentielle Bewegung in Erzählungen. *Linguistische Berichte, 109*; 163–183.

LEVELT, W. J. M. (1982) Linearization in Describing Spatial Networks. In: S. PETERS & E. SAARINEN (eds.) *Processes, Beliefs, and Questions*. Dordrecht: Reidel; 199–220.

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

PINKER, S. (1988) A Computational Theory of the Mental Imagery Medium. In: Denis, M. et al. (eds.) *Cognitive and neuropsychological approaches to mental imagery*. Amsterdam: Martinus Nijhoff; 17–32.

PORZEL, R., M. JANSCHE & R. MEYER- KLABUNDE (1996) Generating Spatial Descriptions from a Cognitive Point of View. To appear in: P. OLIVIER (ed.) *Proceedings of the ECAI Workshop on the Representation and Processing of Spatial Descriptions*.

SHOBER, M. F. (1993) Spatial Perspective-Taking in Conversation. *Cognition, 47*; 1–24.