Multidisciplinary Approaches to Language Production

Trends in Linguistics

Studies and Monographs 157

Editors

Walter Bisang
Hans Henrich Hock
Werner Winter

Mouton de Gruyter
Berlin · New York

Multidisciplinary Approaches to Language Production

Edited by

Thomas Pechmann
Christopher Habel

Mouton de Gruyter
Berlin · New York

Mouton de Gruyter (formerly Mouton, The Hague)
is a Division of Walter de Gruyter GmbH & Co. KG, Berlin.

♾ Printed on acid-free paper which falls within the guidelines
of the ANSI to ensure permanence and durability.

ISBN 3-11-017840-0

Bibliographic information published by Die Deutsche Bibliothek

Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie;
detailed bibliographic data is available in the Internet at <http://dnb.ddb.de>.

Cover design: Christopher Schneider, Berlin.
Printed in Germany.

Preface

The present volume reflects and summarizes a six-year-long interdisciplinary scientific journey into many different regions of the language production research landscape. In 1996 the Deutsche Forschungsgemeinschaft DFG (German Research Council) approved our grant proposal for an umbrella project dedicated to one of the major stepchildren of psycholinguistics: the generation of spoken and written language. We started in 1997 with about 20 individual projects spread over 17 German universities, covering such diverse areas as cognitive psychology, psycholinguistics, linguistics, phonetics and phonology, computer science, neuropsychology and neurology, the study of written and of sign language. The results and findings of the individual projects were discussed at annual meetings that took place at different German universities. In addition, we had four smaller conferences on topics like *Processes of conceptualization in language production: The role of perspectivization, Time and timing in language production, The mental lexicon in language production: Normal and deviant processes, and Processing of syntactic gender in language production*. In order to prevent a national and continental bias we invited guest speakers and discussants from all over the world, among them some of the most distinguished specialists in the field. We also received advice and help from a panel of supervisors who were critical, as they had to be, but always remained constructive and supportive. Our thanks go to Ria de Bleser, Veronika Ehrich, Manfred Bierwisch, John Nerbonne, Manfred Pinkal, Gert Rickheit, and Gerhard Strube. Of the many guests we had the pleasure to host during the past six years, three colleagues stand out who escorted us on more than one annual conference and became part of the whole project. We are very grateful to Kay Bock, Merrill Garrett, and Gerard Kempen for being with us for many years. Our final thanks go to Dieter Zerbst and Andreas Opitz for their substantial help in quarreling with the intricacies of WORD-formatting, to Birgit Sievert and Wolfgang Konwitschny from Mouton Publishers for their editorial support, and, last but not least, to the Deutsche Forschungsgemeinschaft and especially to Susanne Anschütz and Roswitha Müller, who laid the financial foundation for this project.

Thomas Pechmann & Christopher Habel
Leipzig and Hamburg, February 2004

Contents

Introduction

Merrill F. Garrett

The research reported in this book arises from a cooperative research program by groups of investigators from multiple German universities and research institutes. In 1996, these groups proposed to the DFG an ambitious scientific enterprise. They proposed to investigate several different aspects of human language production, and to do so within the general framework set out by W.J.M. Levelt in his landmark 1989 book *Speaking: From Intention to Articulation*. The integrative treatment of language production processes offered in that book held out the promise of coordinated and mutually constraining efforts by multiple research teams. The proposal was approved and in 1997 the groups began, with Thomas Pechmann of the University of Leipzig and Christopher Habel of the University of Hamburg as coordinators, a sustained collaborative program that has this book as one of its outcomes.

Levelt's book set the theoretical stage for systematic empirical studies of a staggering array of specific information processing problems that must be 'solved' by the real-time mental engines of human speaking. The scope of this process extends from the most rarified realms of abstract human thought to the physical minutia of articulation. The broad gauge decomposition of production falls into three broad classes:

Conceptualization yields a message level representation that reflects communicative intention and controls language level processing. *Formulation* incorporates lexical, phrasal and phonological integration of sentence form. *Articulation* implements the phonetic formulae resulting from formulation processes to yield a pronounced form. Each of these has multiple component processes. The whole is an orchestration of interlinked subsystems that map from one extreme to the other. The papers in this volume tap into many different aspects of that spectrum.

Several of the papers tackle problems at the earliest stages of language generation: the real-time elaboration of a specific pre-verbal message that is distilled from conceptual and perceptual information. These processes are among the most difficult to turn into manageable empirical targets. Some noteworthy efforts are represented in this collection. These combine computational modeling with data from various language production tasks to offer quite detailed accounts of one or another aspect of conceptual to preverbal message development, and in some cases, aspects of message to language formulation.

The chapter "Incremental generation of interconnected preverbal messages") by Guhe, Habel, and Tschander reports modeling for generation of message structures from conceptual/perceptual inputs. The early stages of this process are explored using data from real-time descriptions of complex scenes whose apposite description requires object individuation and ordering in a discourse record. Their incremental planning approach to the real-time descriptive problems is profitably complemented by the chapter by Gardent, Manuélian, Striegnitz, and Amoia ("Generating definite descriptions: Non-incrementality, inference and data."), which compares incremental and non-incremental algorithms for generating referring expressions. They explore ways of extracting the communicatively relevant elements from representations of discourse structures for incorporation into sentence planning. The chapter by Harbusch and Woch ("Integrated natural Language Generation with Schema-tree adjoining grammars) treats both conceptualization and formulation problems from the perspective of computational linguistic systems for language generation. They use a common representational format to facilitate computational integration for conceptualization, message level microplanning, and sentence generation.

Other chapters address message level processes that are more immediately linked to formulation level issues and reflect modes that have surfaced in linguistic theory in a variety of ways. The chapter by Klabunde & Glatz ("On the production of focus") reports investigation and modeling of focus assignment, but with treatment situated in real-time language production activity. Focus implicates discourse level structures that surface in identifiable features of phrasal ordering and accent assignment in sentence formulation, and these afford empirical leverage for evaluation of modeling approaches.

The chapter by Tappe, Härtl, and Olsen ("Thematic Information, Argument Structure, and Discourse Adaptation in Language Production") asks how thematic information is represented in the interface between conceptual and sentence level processes. They propose a *thematic processor* as a component of language production models and paper empirically investigate its function in production tasks. The paper by Kempen and Harbusch ("A corpus study into word order variation in German subordinate clauses: Animacy affects linearization independently of grammatical function.")

describes a corpus analysis of the effects of animacy on linearization in sentence planning. They relate distributional evidence for the ordering of lexical and phrasal structures, and suggest contrasts between models in which animacy exerts both conceptual level and sentence level effects.

These chapters just enumerated provide some distinctive approaches to systems involved in distilling conceptual information specific enough to guide sentence formulation. Another chapter puts a reverse spin on this. Carroll, Stutterheim, and Nüse ("The *language and thought* debate: a psycholinguistic approach") revisit classic questions of how language may shape conceptualization. Their cross-linguistic comparisons of English, German, Dutch and three Romance languages French, Italian, and Spanish examine language specific factors as influences on the nature of the preverbal message level of language production.

A different kind of language comparison underpins the chapter by Leuninger, Hohenberger, Waleschkowski, Menges, and Happ ("The impact of modality on language production: Evidence from slips of the tongue and hand"). This work provides a multilevel comparison of signing (DGS) and spoken language performance. Experimentally induced DGS signing error patterns are compared with those for spoken speech errors. They report impressive similarity of error typology, combined with some differences in the distribution of error types that are understandable in terms of differing modality constraints on expression. Sign language research has contributed substantially to our understanding of both linguistic and psycholinguistic theory. This paper makes an important new contribution by documenting detailed parallels of error mechanisms for signed and spoken language.

Several papers address aspects of sentence formulation and related lexical processes, with special attention to the time course for elaboration of different classes of structure. Several different methodologies are represented ranging across reaction time studies of picture naming, electrophysiology, aphasic disorders, and brain imaging. Together these several papers give new perspectives on the relations among the major classes of lexical structure involved in sentence formulation.

The chapter by Pechmann and Zerbst ("Syntactic constraints on lexical selection in language production") summarizes their studies using the picture-word interference paradigm. This method has been a mainstay of lexical retrieval studies, and the paper reports an extensive application of it to identify and temporally titrate semantic, syntactic, and phonological effects during lexical retrieval. Methods that yield reliable syntactic effects are reported and are then evaluated in the same time frames as the established effects for semantic and phonological variables. They report significant overlapping of the activation profile for the three information classes and suggest a need to modify strictly sequential architectures in the direction of cascading models to accommodate this.

The chapter by Blanken, Kulke, Biedermann, Bormann, Dittmann & Wallesch ("The Dissolution of Word Production in Aphasia: Implications for Normal Functions") addresses similar issues in lexical retrieval with data from on the language performance of brain-damaged patients. They also argue for a modification of the strict two stage feed-forward model in order to accommodate regularities of phonologically related word substitutions and to explain phonological influences on syntactically constrained word substitutions in patient populations. The chapter by Schade ("The benefits of local connectionist production") also calls attention to indications for some departure from strict feed-forward staged modeling. His aim is to stress the value of comparisons of different modeling approaches to identify the strengths and limitations of each. In aid of this, he evaluates various target processes in production from the perspective of interactive activation models and compares this with the multi-level staged architectures to find points of productive empirical contrast.

Two other papers report work with brain-based measures and these complement the studies just described in the preceding paragraph in revealing ways. The chapter by Münte et al. ("Electrophysiological studies of speech production") They used several ERP measures, (including evaluation of lateralized readiness potentials in a "two-choice go/nogo" procedure) to apply electrophysiological methods to production study. Their experimental studies reveal stable differences across several different tasks in the relative timing for retrieval of conceptual/semantic, syntactic, and phonological information during language production, with ordering as given. This evidence indicates that the information types become available to different time courses, but does not constrain the choice of sequential vs cascading models. The chapter by Dogil et al. ("Brain dynamics induced by language production") reports outcomes of innovative applications of brain imaging tools for the study of component systems in language generation. They report evidence both for significant localization effects for semantic and syntactic systems, and for different types of brain activity associated with processes that occur in common areas of the brain. They also report reflections of specific prosodic factors and their association with semantic and syntactic processes. The overall picture of brain activity associated with language processing, both production and comprehension, is more complex than most such characterizations as well as implicating both segregated and interactive processes for the component systems.

Several additional papers in this collection treat detailed features of the formulation process relating to morphology and phonology. The chapter by Bölte, Zwitserlood, and Dohmes ("Morphology in production research: Evidence for decomposed representations") asks how the productive potential of morphological structure is embedded in models of language generation. Are compositional morphological processes at work in on-line process as opposed to retrieval of static stored information about whole word forms? The work reported here used a variation on the picture-word interference paradigm to compare different types of morphological structure (derivation, inflection and compounding) and to argue for decompositional morphemic structure at the word form level. Janssen, Bordag, and Pechmann's chapter ("Inflectional frames:...") focuses on a narrower issue of morphology, namely the processing of inflection based on stored inflectional templates. Experimental results for German (as compared to Dutch and English) suggest the use of such frames is language dependent. The chapter by Weingarten, Nottbusch and Will ("Morphemes, syllables and words in written word production") examines the projection of production questions into the domain of written word production. Timing constraints during typing and their relation to lexical structure reveal evidence for sublexical structures in the implementation of written language. Still another exploration of the relation between language and other cognitive systems is Hamm and Bredenkamp 's report ("Working Memory and slips of the tongue.") of working memory mechanisms in sound exchange errors. Experimentally induced errors are tested for susceptibility to phonological and semantic influences to examine the role of central executive and memory processes in the specialized domain of language generation

Readers of these papers will not find one voice, but they will find a better than usual opportunity to follow the different threads of the discussions. One of the admirable features of this DFG supported enterprise is the extent to which the investigators maintained a collaborative exchange over the life of the grant. They met each year in stimulating gatherings in which the progress – and frustrations – of each working group were shared with the collective. I was privileged to attend some of those meetings, and can testify to the lively discussion and interactions they sustained. The papers in this volume, as well as other published results of the research program demonstrate the effectiveness of program design and process. The strategy adopted of using a common framework for setting their diverse problems was a central factor in this. The common framework makes for good communication and enhances the potential for coordination of results. Reading and understanding the several papers in this volume is facilitated by the use of common terminology and distinctions. It does not, however, straightjacket the critical evaluation of results and their interpretation. The outcomes of the projects do not always conform to the empirical predictions derived from the theoretical starting points in Levelt's framework. How could they? Some of the problem domains Levelt addressed in 1989 were quite undeveloped, while others, though better explored, were still open to multiple interpretations. It is the virtue of the coordinated research program underlying this book that it afforded common ground for communication and collaboration while supporting diversity in the scientific elaboration of the several problem domains that comprise the overall effort.

Incremental generation of interconnected preverbal messages

Markus Guhe, Christopher Habel, and Ladina Tschander

1. Introduction

When producing verbal messages a speaker has to integrate different knowledge sources, among others perceptually based knowledge provided by current observation of the environment and knowledge about systematic relations in the external world. Furthermore, during this first phase of language production, called *conceptualization* or *conceptual preparation*, the speaker has to select knowledge to be communicated and has to decide which aspects will be presented as being in the foreground of the message. The result of conceptualization, the *preverbal message*, is a complex conceptual structure, which – during later phases of production – is realized as spoken or written language (Levelt 1989, 1999).

During the last two decades, the later phases have been in the focus of research on language production. Lexical access, grammatical, morpho-phonological, and phonetic encoding as well as articulation have been investigated from different scientific perspectives: psycholinguistics and neurolinguistics as well as theoretical and computational linguistics. However, research on conceptualization has been done primarily in the field of computational linguistics (Levelt 1999: 89).

In this paper we take a cognitive modeling approach. We are concerned with the conceptual preparation in the task of describing *on-going* events.[1] Speakers, who observe what is going on in their environment and communicate what they currently see, have to decide which objects and which events they will verbalize. Additionally, they have to do this in real time. In other words, during language production the speaker has to synchronize a stream of knowledge provided by perception and processes of building up complex conceptual structures which are the basis of further linguistic encoding. To solve this problem of a temporally shifted synchronization we propose an *incremental conceptualizer* (INC), which follows the general principle of incremental architectures widely accepted for language production (Kempen and Hoenkamp 1987; Levelt 1989). Incremental models are often depicted as a cascade of processes: like water in a water cascade splashing down from one level to the next, information is splashing down from one process to the next. Unlike water, however, the stream of information in language production is not continuous but consists of pieces of knowledge, called *increments*. In other words, the process of conceptualization presupposes *conceptual building blocks*, which are manipulated and combined to new and more complex conceptual structures. These conceptual increments are used to generate preverbal messages, which, in turn, are processed by later components of language production.

The conceptualizer can combine the conceptual material in different ways, which give rise to a variety of utterances appropriate to describe observed situations and objects. In Guhe, Habel, and Tschander (2003a, 2003b), we describe how INC generates complex conceptual structures in an incremental fashion, which can function as preverbal messages to verbalize *A plane turns to the left* after observing a corresponding movement. In the present paper, we focus on the question of how multiple preverbal messages can be combined resulting in coherent discourse.

In particular, we consider two ways of establishing coherence in discourse: coreference and discourse relations. The first depends on the fact that preverbal messages contain conceptual entities that refer to perceived objects. So, if two or more conceptual increments in (different) preverbal messages contain references to the same object, their coreferentiality has to be encoded in the preverbal messages in question. On the linguistic surface this can be realized by pronouns or other anaphoric devices. The second way of establishing coherence is related to conceptual entities representing situations of the perceived environment, in particular events. In online descriptions of events – which are in the focus of the present paper – temporal relations between the observed situations lead to a specific realization of discourse relations: *temporal connectives* (Kehler 2002; Asher and Lascarides 2003). While the first type of interconnections is between the parts of preverbal messages, these connectives link whole preverbal messages.

At first sight, coherence of discourse (coherence in the stream of preverbal messages) on the one hand and the incremental character of the conceptualization process on the other hand seem to contradict each other. In the following sections we will dissolve these opposing requirements of *coherent structures* vs. *incremental processing*. In section 2 we describe the verbalization task to be solved by our computational conceptualizer INC, which is a model in the sense of cognitive modeling (Guhe, Habel, and

Tschander 2003b). In particular, we discuss some key phenomena of coreference and temporally induced coherence, which occur in producing online verbalizations of events. Section 3 presents the representational basis of INC. Firstly, we give a more detailed analysis of relational temporal concepts, which are used in the incremental construction of conceptual structures. Secondly, we exemplify the representational formalism *referential nets* by presenting the INC-internal representations for some verbalizations discussed in section 2. Section 4 addresses the architecture of INC and section 5 describes how coreference and temporal connectives are established in preverbal messages in the INC conception.

2. Coherent descriptions of simultaneously occurring motion events

2.1. The overall setting

We use a complex scenario to exemplify the conceptualization process and the architecture INC. In this scenario the speaker observes the taxiing of aircraft – the movements of aircraft from the terminal to the assigned runways and vice versa. Beyond exemplification, the function of this scenario is twofold. On the one hand, it serves as the world of the computational realization of INC, i.e. it is the domain for testing our approach by a computational model (Guhe, Habel, and Tappe 2000; Guhe and Habel 2001; Guhe, Habel, and Tschander 2003a, b). On the other hand, visualizations of the dynamic world described below are presented to participants in verbalization studies.[2]

Figure 1a depicts the spatial environment of the observed events that are to be verbalized. This background, which can be seen as the visual field of an observer, consists of an maneuvering area containing two parallel runways (identified as *runway 1* and *runway 2*) and a connecting way (*connection*). Figure 1b illustrates the events discussed in the following. Two planes are moving on these runways. They enter the scene, i.e. their movements become observable, at different points of time, indicated in the figure by *time-stamps* (t_1 and t_3). All time-stamps depicted in figure 1b give the reader information of the temporal order of the events. The events are presented dynamically. Furthermore, these time-stamps correlate with distinguished positions of the planes during their movements that subdivide the scene into different situations. For example, the appearance of a plane in the visual field of the observer always is a distinguished position. Likewise, the locations where the direction of motion changes or a movement ends constitute new situations.

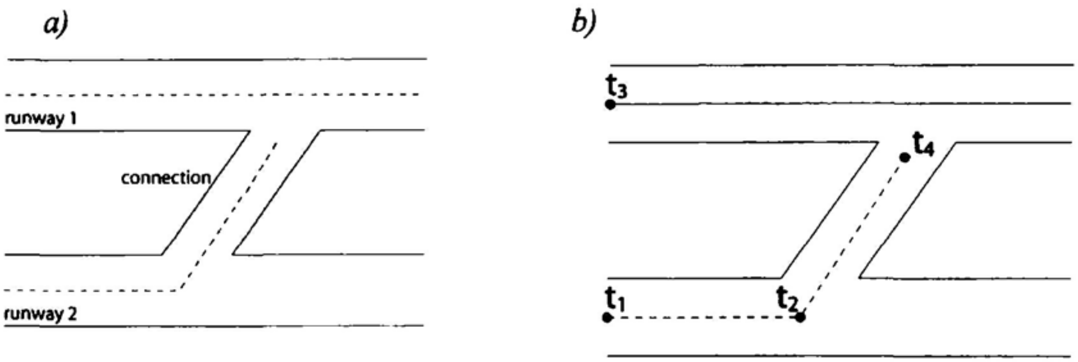


Figure 1. Paths of motion of two planes moving on a maneuvering area (a) and the temporal structure of the events (b)

In the verbalization study, the participants are asked to describe the motion events while they are observed. Analogously, the INC simulation uses temporally ordered snapshots of the scene as input (section 4.1).[3] When INC starts to prepare a preverbal message it builds up simple conceptual representations corresponding to individuated events. The individuation of events is performed by a perceptual pre-processing unit (PPU), which we will not describe here in detail.

The simple conceptual representations, which already contain temporal information, are used to build up complex conceptual structures (sections 3 and 4.1). For example, a movement on runway 2, which passes into a movement on the connection (see figure 2), leads to a conceptual structure, which also contains a complex event corresponding to the whole movement. For constructing such complex, hierarchical representations the temporal relations between events play an important role (see section 3.1).

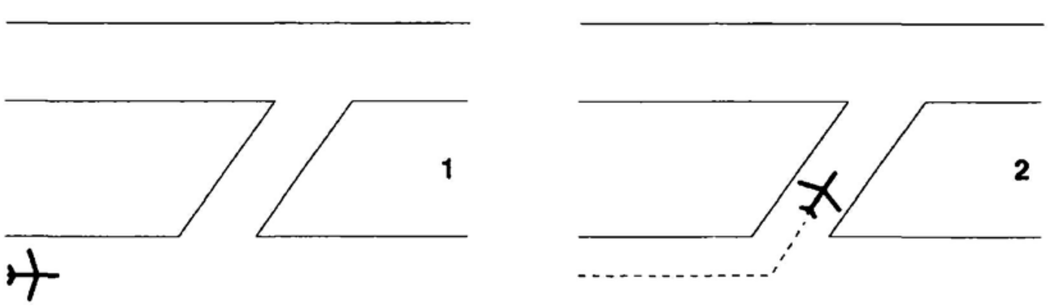


Figure 2. Starting phases of the example scene with only one plane

INC's communicative goal is to inform a person about what is happening. Thus, the moving objects have to be identified verbally, and a sequence of utterances is employed to describe the complex scene – if the verbal description does not consist of a one-sentence summary. The bearer of motion is, therefore, mentioned several times in the discourse. Consequently, the generation process has to be able to express such coreferences. Furthermore, since there are two bearers of motion of the same type later on in the scene, they have to be distinguished. In section 2.2, we look closer at the linguistic means providing coreference information. Online descriptions of sequences of events lead mostly to sequences of utterances, where both sequences coincide. Describing the events in the order in which they occur is a standard verbalization strategy. However, if the situation contains simultaneous events, e.g. two planes in motion on different runways, INC must decide in which order it will describe the events. In doing so, however, it must make explicit the temporal order of the verbalized events in the preverbal messages. This provides a conceptual connection that can be used for producing coherent discourse (section 2.3 and 3.1).

2.2. Coreferential expressions

Let us start with the first and second phase of the scene, which are depicted in figure 2. The first phase can be described by verbalizations as given in (1) and (2). The generation of utterances of this kind is described in detail in Guhe, Habel and Tschander (2003a). In the second phase of the scene the plane turns into the way connecting the two runways. To verbalize this movement, an object already mentioned has to be described again. This is facilitated by the fact that the moving entity of phases 1 and 2 is represented by one entity in the conceptual representation (see section 3.2).

(1) *Ein Flugzeug erscheint.*
'A plane appears.'[4]

(2) *Ein Flugzeug fährt geradeaus.*
'A plane is moving straight on.'

The simplest way of expressing such a coreference is repetition. Assume that the plane is identified by its name as in (3)a. The simple repetition of the name sounds quite unnatural. Similarly, a definite noun phrase referring to the same referent is an unusual kind of expressing coreference, shown in (3)b. However, if the noun phrase is specified, for example with a prepositional phrase locating the object as in (4), the coreference can be verbalized by a repetition.

(3) a. *? Lufthansa 112 fährt auf Runway 2. Bei der Gabelung biegt Lufthansa 112 nach links ab.*
'? Lufthansa 112 is moving on runway 2. At the branch, Lufthansa 112 turns off to the left.'
b. *? Auf Runway 2 fährt ein Flugzeug. Das Flugzeug biegt auf die Verbindung ab. Das Flugzeug fährt weiter.*
'? On runway 2, a plane is moving. The plane turns off into the connection. The plane is moving on.'

(4) *Das Flugzeug auf Runway 2 fährt. Bei der Gabelung biegt das Flugzeug nach links ab.*
'The plane on runway 2 is moving. At the branch, *the plane* turns off to the left.'

The phenomenon of coreference is tightly connected to that of pronominal anaphora. The use of anaphoric expressions seems to be the most natural way to express conceptual coreference.[5] Referents that are antecedents for anaphora are usually introduced via indefinite noun phrases or mentioned by definite noun phrases. However, in our scenario the first mentioning of the object in motion can also be expressed by a definite noun phrase, as in (4)–(6), because the conceptual representation contains only one plane. To mention the plane again without giving a simple repetition, other noun phrases expressing designations related to the concept *plane* can be chosen. The noun phrase *die Maschine* 'the vehicle' in (5) is an example of constituting coreference by the use of nouns that are semantically related to the concept *plane.* However, coreference is most often expressed by the use of pronouns as in (6).

(5) *Das Flugzeug fährt auf Runway 2. Bei der Gabelung biegt die Maschine nach links ab.*
'The plane is moving on runway 2. At the branch, the vehicle turns off to the left.'

(6) *Das Flugzeug fährt auf Runway 2. Bei der Gabelung biegt es nach links ab.*
'The plane is moving on runway 2. At the branch, it turns off to the left.'

Webber et al. (to appear) regard expressions like the ones given in (5) and (6) as discourse anaphora. They distinguish three groups of discourse anaphora according to the relation between the denotations of the anaphora and their referents. The most obvious relation is the one in which the denotations of the anaphora and the referents are *identical* as in the examples given above.[6] Expressions that are associated via their meaning to *parts* of referents are the second kind of discourse anaphora. They do not refer to the whole conceptual entity but only to associated parts.[7] However, the use of this kind of anaphora is restricted in the description of a moving object. For example, it is not enough to mention a part of the plane for verbalizing that the whole plane turns to the left (7).

(7) **Das Flugzeug fährt auf Runway 2. Bei der Gabelung biegen die Räder nach links ab.*
*'The plane is moving on runway 2. At the branch, the wheels turn off to the left.'

In comparison to anaphora expressing the reference to the same conceptual entity Webber et al. (to appear) give a third class of anaphoric expressions. These anaphora denote a referent that is *related* to an already given referent. Consider (8) in which the discourse anaphor *das/ein rote(s)* 'the/a red one' establishes a contrast relation with an already given entity. The anaphor cannot be linked to the same referent (8)a; instead it refers to another referent of the same kind (8)b.

(8) a. **Das Flugzeug fährt auf Runway 2. Bei der Gabelung biegt das rote nach links ab.*
*'The plane is moving on runway 2. At the branch, the red one turns off to the left.'
b. *Das Flugzeug biegt nach links ab. Ein rotes erscheint auf Runway 1.*
'The plane turns to the left. A red one is appearing on runway 1.'

So, these anaphora do not denote the referent given by the antecedent. They convey a function in their meaning[8] that is applied to either the referent of the antecedent or an associate of the referent of the antecedent. Since these anaphora often establish a contrast relation, several conceptual entities of the same type must be available. This is given in our scenario when a second plane appears (see figure 3).

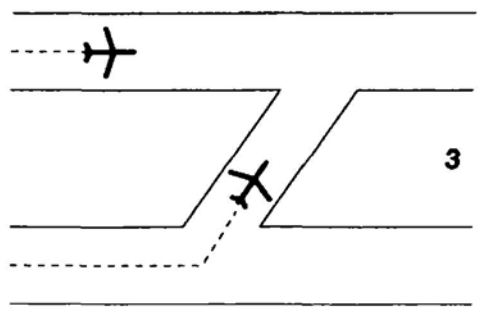


Figure 3. Third phase of the example scene, where another plane appears

Consider (9), which introduces this second plane appearing in phase 3 of the scene. The NP *another plane* constitutes an anaphoric relation to a referent already given in the discourse, i.e. the plane moving on the connection.

(9) *Das Flugzeug biegt nach links ab. Ein anderes Flugzeug erscheint auf Runway 1.*
'The plane turns to the left. Another plane is appearing on runway 1.'

In comparison to the relations given in examples (1) to (7), the anaphoric expression in (9) establishes that the two planes are different objects, i.e. the anaphora have different referents. So, the lexically given function of *another* or *other* says that an entity of the same type as the one referred to is already given in the discourse, and the referents are not identical.

These considerations on the way of interpreting anaphora result in conditions that have to be considered in generating anaphora. This is one way to interrelate preverbal messages. If an object is referred to by a definite noun phrase, another mentioning of the same object should not be given by the same definite noun phrase but by a pronoun or a noun phrase semantically related to the referent. Additionally, if there are two objects of the same type the generation of coreferential expressions must be blocked. It must be linguistically marked that they are two different entities. This can be done by using *ein anderes/andere* 'other/another' or by attributing a count adjective such as *zweites* 'second'.

After presenting how conceptual coreference is related to anaphora, we will now focus on the temporal relation between events – another way to establish coherent discourse.

2.3. Temporal connectives as means to establish coherent discourse

Figure 4 statically depicts the dynamic situation in which two planes are in motion (phases 3 and 4 of the scene). INC must build up conceptual structures that explicitly contain temporal relations among the co-occurring events (see section 3.1). Temporal ordering at the level of conceptual structures can differ from the order of verbalizations, in other words, an event that occurred later than another one can be verbalized earlier. Thus, the temporal relations connecting preverbal messages are the conceptual basis for linguistic markers (temporal connectives) that establish discourse relations. The linguistic markers themselves are computed in later stages of language production.

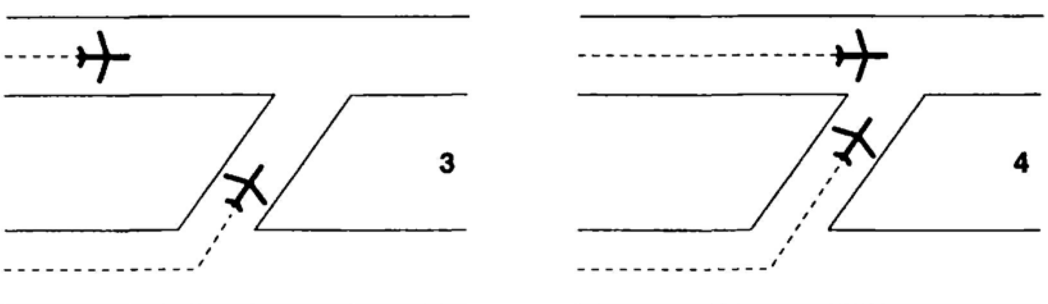


Figure 4. Phases of the example scene with two planes in motion

Speakers plan their contributions with the goal that hearers build up a particular representation. For this reason, and because this is the perspective of the majority of research on discourse relations (see for example Lascarides and Asher 1993; Asher and Lascarides 2003; Kehler 2002) we change to the comprehension perspective for the time being.

We start with discourse (10 a-d), which describes the example scene without using linguistic markers. (The e_x denote the events that are mentioned in the discourse.) In particular, this discourse does not contain any explicit information about temporal relations between the events in question. The only connections between the underlying preverbal messages are chains of coreferences in (10a), (10b), and (10d) or the denial of coreference in (10c). These combinations establish weak relations between the mentioned situations. They do not say explicitly how situations as a whole are related.

(10) a. *Ein Flugzeug fährt auf Runway 2.* (e_1)
'A plane is moving on runway 2.'
b. *Es biegt auf die Verbindung ab.* (e_2)
'It turns off onto the connection.'
c. *Ein anderes Flugzeug fährt auf Runway 1.* (e_3)
'Another plane is moving on runway 1.'
d. *Das erste Flugzeug stoppt.* (e_4)
'The first plane stops.'

Since the relations between the events described in the discourse are not marked explicitly, e.g. by using *dann, danach, nachdem, davor, vorher* (German counterparts to *then, after, before*), the hearer of the discourse has to infer appropriate relations during comprehension.[9] However, since they are not given explicitly, the correct interpretation of the utterance cannot be guaranteed.

A plausible interpretation of a narrative discourse, which does not contain explicit information about the temporal ordering of events, is that the order of mentioning in the discourse corresponds to the temporal order of the events. This interpretation strategy follows the conception of the *order of mention contract* (Clark and Clark 1977), which can be seen as an instantiation of Grice's maxim of manner (Grice 1975).

Asher and Lascarides (2003) augment this idea using the theoretical conception of discourse relation in the framework of SDRT (*segmented discourse representation theory*): in narrative texts the relation NARRATION can be used as standard assumption to relate two successive propositional discourse entities. NARRATION implies that the situations described are temporally ordered or that a weak causal relation is established between the situations.

As explained in Lascarides and Asher (1993) the temporal information conveyed by the textual order is represented by two rules: a *nonmonotonic rule* for NARRATION that interacts with an *axiom of narration*. As a consequence, it is inferable that if the relation NARRATION holds between two discourse segments, then the situation described in the first discourse segment occurs before the one described in the second segment. So, in the simplest case each of the discourse segments given in (10) is related by

NARRATION to its preceding segment, and the inferred temporal sequence in which the mentioned events occurred is $e_1 \prec e_2 \prec e_3 \prec e_4$. This relation can be given explicitly by inserting *dann* 'then' as in (11).

(11) *Ein Flugzeug fährt auf Runway 2 (e_1). Dann biegt es auf die Verbindung ab (e_2). Dann erscheint ein anderes Flugzeug (e_3). Dann stoppt das erste Flugzeug (e_4).*
'A plane is moving on runway 2 (e_1). Then it turns off onto the connection (e_2). Then another plane is moving on runway 1 (e_3). Then the first plane stops (e_4).'

Switching back to the production view, some consequences of using such inferences are obvious. Speakers wanting to describe a perceived scene properly must consider the interpretation strategy described above. If there are two events occurring simultaneously, a discourse without explicit temporal connectives can lead the hearer to incorrect assumptions about the situation. In other words, giving information about the temporal relations explicitly by using temporal connectives blocks the default reasoning. This means, especially the relations between (10)b, (10)c, and (10)d have to be given explicitly.

Other discourse markers relating events are the temporal connectives *nachdem* 'after' and *bevor* 'before'. In (12), e_2 is related to e_3 by *nachdem*. Due to the semantics of *nachdem* and the use of past perfect the statement says that the event e_2 is bounded and e_3 does not overlap with e_2.

(12) *Nachdem es auf die Verbindung abgebogen ist* (e_2), *fährt ein anderes Flugzeug auf Runway 1* (e_3).
'After it has turned off onto the connection, another plane is moving on runway 1.'

Just like *nachdem* 'after', *bevor* 'before' relates segments temporally. In comparison to the semantics of *nachdem*, which demands that the related event is terminated, the semantics of *bevor* requires that one event (e_3 in (13)) has started earlier than the other one (e_4 in (13)). Furthermore, the semantics of *bevor* does not imply that the mentioned event is terminated. So, (13) says that the plane is moving towards a specific location but has not reached it yet. Therefore, the mentioned movement is on-going.

(13) *Bevor das Flugzeug auf Runway 1 die Verbindung erreicht* (e_3), *stoppt das Flugzeug auf der Verbindung* (e_4).
'Before the plane on runway 1 reaches the connection, the plane on the connection stops.'

If on-going events are to be verbalized, e.g. simultaneous movements of both planes, another type of temporal connective has to be selected. In comparison to the semantics of *nachdem* 'after' and *bevor* 'before', which demand that the events are bounded (or complete), the temporal connector *während* 'while' is used for unbounded events, i.e. the events are not complete.[10] Its basic temporal meaning can be seen as asserting that the situations related via *während* possess a significant temporal overlap (see section 3.1).

Looking at example (14), the mentioned events in (14) as well as in (14) are connected by *während* 'while'. In the case of (14), two uncompleted motion events are described, which happen at the same time but at different locations on the maneuvering area: the mentioned events overlap. In comparison to this, the first event of (14) is a completed event whereas the second one is an uncompleted one. The established temporal relation is that the completed event is contained by the uncompleted one. Nevertheless, both temporal relations (overlap and contain) are expressed by the same linguistic marker.

(14) a. *Ein anderes Flugzeug fährt auf Runway 1, während das erste Flugzeug auf der Verbindung fährt.*
'Another plane is moving on runway 1 *while* the first one is moving on the connection.'

b. *Das erste Flugzeug stoppt, während das Flugzeug auf Runway 1 weiterfährt.*
'The first plane stops while the plane on runway 1 is moving on.'

Putting all this together, we propose the discourse given in (15) as an appropriate verbalization of the observed scene. It describes the movements of the planes as a sequence of events that happen on the maneuvring area. The segments of the discourse are connected by two interrelated chains of coreferences, each concerning one of the objects in motion. Additionally, discourse relations realized by temporal connectives establish further coherence of the description. Note that since e_2 is terminated and nothing else is said about the movement of the first plane, it has to be inferred that the plane's movement on the connection is continuing in order to relate the statements given in (15) and (15).

(15) a. *Ein Flugzeug fährt auf Runway 2* (e_1).
'A plane is moving on runway 2.'
b. *Nachdem es auf die Verbindung abgebogen ist* (e_2),
'After it has turned off onto the connection,'
c. *fährt ein anderes Flugzeug auf Runway 1* (e_3).
'another plane moves on runway 1.'
d. *Dann stoppt das erste Flugzeug* (e_4),
'Then the first plane stops,'
e. *während das Flugzeug auf Runway 1 weiterfährt* (e_5).
'while the plane on runway 1 is moving on.'

Before showing how INC generates a coherent discourse in section 5 we now first present the underlying representations and describe INC in detail.

3. INC's conceptual representations of motion events

3.1. Representing temporal relations

Let us summarize a main result of section 2. The explicit representation of temporal relations between events in conceptualization is fundamental for producing interconnected preverbal messages. Thus, we start our discussion of conceptual representations in INC by focusing on the temporal aspects of event conceptualization; the presentation of the representational formalism in general is postponed to section 3.2.

As described above, the perceptual pre-processing unit of INC provides conceptual building blocks, which can be seen as *simple event conceptualization* (see section 4). From our information processing perspective conceptualization consists of processes that construct and transform conceptual structures. Thus, in the context of the conceptualizer we will use *event* and *situation* instead of *event conceptualization* and *situation conceptualization* for referring to conceptual entities. Hence, we will not mention the term *conceptualization* explicitly, but we will always assume that a conceptual representation is internal to a system/human.

The conceptual representations built up by INC are realized with the representational formalism *referential nets* (Habel 1986, 1987), which is described in more detail in section 3.2. Referential nets were developed to model cognition-motivated conceptual and linguistic processes, especially representations that change over time.[10] For this section it is enough to know that in referential nets entities are represented as *referential objects* (refOs), and all information stored in a referential net is associated with refOs.

In the discussion of temporal relations among situations we abbreviate representations of situations by their referential net counterparts. Thus, in observing the exemplifying scene depicted in figure 2, the visual perception of the plane moving on runway 2 leads to a situation refO, for example *s1*, in the INC-representation.[12] Let us now focus on INC's current conceptual representation at a time *t* between t_1, the time the plane enters the scene, and t_2, the time the plane changes its orientation (both given as time-stamps in figure 1). Since the movement is on-going, the situation *s1* is attributed with the property *–complete*. Later on, at time *t* between t_2 and t_3, the time the second plane enters the scene, the perceptual pre-processor has provided information about the change of orientation in the movement of the plane under observation. This change of orientation closes *s1* and opens a new situation, which corresponds to the movement on the connection. This leads to the following modifications of the conceptual representation: firstly, the completeness-attribute of *s1* changes to *+complete*; secondly, a new situation refO *s2*, attributed with *–complete*, is introduced.

At this stage of observing and conceptual processing two situations are introduced in the conceptual representation of INC. On the basis of the pre-processor's stream of information the temporal relation between *s1* and *s2* can be determined as: *s1 meets s2*. *meets* is one of Allen's topological relations for characterizing the structure of temporal relations between situations (Allen 1984). Its meaning can be described in a semi-formal manner as: *s1* is before *s2*, and there is no situation between *s1* and *s2*.

At *t3* the second plane enters the scene. At a time *t* between t_3 and t_4 two simultaneous movements with two different protagonists are observed. During this phase the conceptual representation is updated to the following state: firstly, a new situation refO *s3*, attributed with *–complete*, is introduced. Note that the completeness-attribute of *s2* does not change. Secondly, the temporal relation between *s2* and *s3* has to be determined. At this stage of processing, three relations of Allen's topological relation system could be ascribed.

(a) *s2 contains s3*, '*s3* is temporally properly included in *s2*'

s2: xxxxxxxx
s3: xxx

(b) *s2 finished_by s3*, '*s2* starts before *s3*, and both situations end simultaneously'

s2: xxxxxxxx
s3: xxxxx

(c) *s2 overlapped_by s3*, '*s2* and *s3*, and there is no situation between *s2* and *s3*'

s2: xxxxxxxx
s3: xxxxxxxx

Case (a) is the continuation that takes place in the exemplifying scene, (c) would be the case, if the second plane stopped to give way for the first plane, and (b) would, for example, describe a crash of the two planes. Since it is not known yet, how *s2* and *s3* will go on, INC can only assume that the following disjunction holds:

s2 contains s3 $\vee$ *s2 finished_by s3* $\vee$ *s2 overlapped_by s3*

The last phase we discuss in this subsection commences with t_4, i.e. it starts with the stop of the first plane. This leads to the following modification of the conceptual representation: firstly, the completeness-attribute of *s3* changes to *+complete*; secondly, a decision between the relational alternatives (a) to (c) is possible. Since (a) and (b) can now be rejected, *s2 overlapped_by s3* must hold. Thirdly, a new situation in which the plane is standing at the location where it stopped is introduced.

The conceptual computations described above only concern the completeness-attribute of the situations and the temporal relations between current situations. Beyond this, Allen's framework – also often called Allen's calculus of topological relations – can be used, to propagate knowledge about the temporal relations among the situations of the scene. For example, from *s2 overlapped_by s3* and *s1 meets s2* it is possible to infer to *s1 before s3*, i.e. *s1* is before *s3*, and *s1* is not immediately succeeded by *s3*.

Let us now focus on the linguistic potential of the representations discussed above. Firstly, the dynamic ascription of completeness-attributes is the basis to provide information about aspect in the preverbal message if it is required in the target language. Secondly, some disjunctions of temporal relations are not only complexes of temporal relations, but they also possess the status of indeterminate temporal relations for their own. For example, for many conceptualization tasks – this also holds for many comprehension tasks – the difference between *s before s'* and *s overlaps s'* is not relevant. Often the difference is not observable. In other words, assuming an indeterminate temporal relation, which we call *earlier* here, can be characterized via *before* and *overlaps*, but it has the advantage of being cognitively more efficient than a mere disjunction. Thirdly, the temporal relations as used by INC consider the asymmetry of the two related situations, e.g. the same temporal constellation can be expressed by *s before s'* and *s' after s*. This asymmetry is employed in the section 5.3 where we describe how a situation refO to be verbalized can be connected to others that precede or succeed this one in the sequence of incremental preverbal messages.

3.2. Representing conceptual structures in referential nets

After we laid out how situations are temporally connected on the conceptual level in the previous section, we now discuss how this representational frame is complemented by object and spatial knowledge so that we obtain a complete conceptual representation of the observed scene. Figure 5 depicts a referential net representing the example scene after all events were observed.

Since the conceptual representation is built up incrementally, elements change over time, described in section 4.2. There are three different sorts of entities in the conceptual representation:

- *object*: Objects are concrete real-world objects, e.g. a plane or a runway.
- *spatial entity*: A spatial entity is an abstract entity representing locations, directions, or paths of motion.
- *situation*: Situations cover all situation types as state, event, and process (Bach 1986).

Note that the paths of motion are not visually persistent, i.e. paths are no real-world tracks like runways. Paths are linear, directed, bounded entities, cf. Eschenbach et al. (2000). A path has two distinguished points: the *starting point* precedes every other point of the path; the *final point* is preceded by every other point of the path.

We do not distinguish the different situation types as such, i.e. we do not use the notions process, state, and event, because on the conceptual level this distinction is problematic.[13] This is especially true in an incrementally working model like INC. One problem is that the situation type can change according to what has been observed. For example, when the first plane enters the scene the conceptual representation contains no starting or endpoint of the movement. Consequently, the situation type is *process*. As soon as the plane stops moving, however, the situation type changes to *event*, because it now contains a culmination point. Another problem is that the situation type is often just a matter of which view on the same situation is used. For example, when the movement of the first plane on runway 2 is verbalized it can be described in different ways:

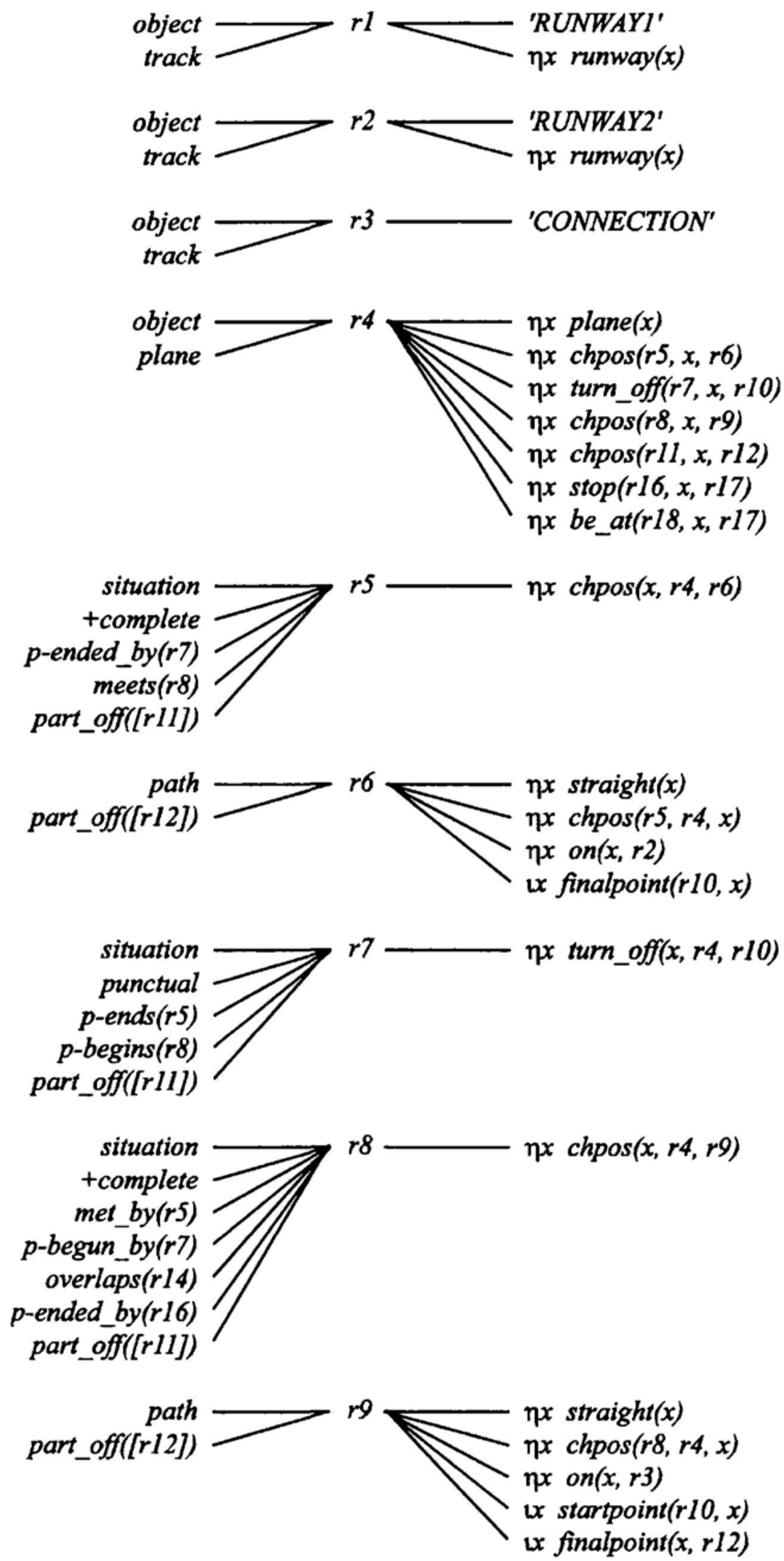
(16) a. *Das Flugzeug fuhr in Richtung der Verbindung.*
'The plane was moving towards the connection.'
b. *Das Flugzeug fuhr bis zur Verbindung.*
'The plane moved to the connection.'

In (16)a the emphasis is on the *process* of moving, while in (16)b it is on the culmination point of the movement, and, therefore, on the *event* character of the movement. For convenience we allow that in extension to section 3.1 situations can not only be extended but also punctual.

As mentioned in the previous section referential nets contain refOs, which are specified by:

- a term
- their sort
- attributes
- designations

Terms serve to refer to refOs, e.g. *r1*, *r2*, etc. The sort of a refO locates the entity that can be referred to by the refO term in a sort hierarchy. Formally, the sort hierarchy is a lattice. As we already argued, there are three kinds of entities in the domain of motion events: *object*, *spatial_entity*, and *situation*. While objects and situations are not further specified by their sort, spatial entities can be of the (sub-)sort *path* or *location*. Path refOs represent the trajectories along which an object moves, locations stand for prominent points like points where an object changes is orientation.


object
track
r1
'RUNWAY1'
ηx runway(x)
object
track
r2
'RUNWAY2'
ηx runway(x)
object
track
r3
'CONNECTION'
object
plane
r4
ηx plane(x)
ηx chpos(r5, x, r6)
ηx turn_off(r7, x, r10)
ηx chpos(r8, x, r9)
ηx chpos(r11, x, r12)
ηx stop(r16, x, r17)
ηx be_at(r18, x, r17)
situation
+complete
p-ended_by(r7)
meets(r8)
part_off([r11])
r5
ηx chpos(x, r4, r6)
path
part_off([r12])
r6
ηx straight(x)
ηx chpos(r5, r4, x)
ηx on(x, r2)
ιx finalpoint(r10, x)
situation
punctual
p-ends(r5)
p-begins(r8)
part_off([r11])
r7
ηx turn_off(x, r4, r10)
situation
+complete
met_by(r5)
p-begun_by(r7)
overlaps(r14)
p-ended_by(r16)
part_off([r11])
r8
ηx chpos(x, r4, r9)
path
part_off([r12])
r9
ηx straight(x)
ηx chpos(r8, r4, x)
ηx on(x, r3)
ιx startpoint(r10, x)
ιx finalpoint(x, r12)

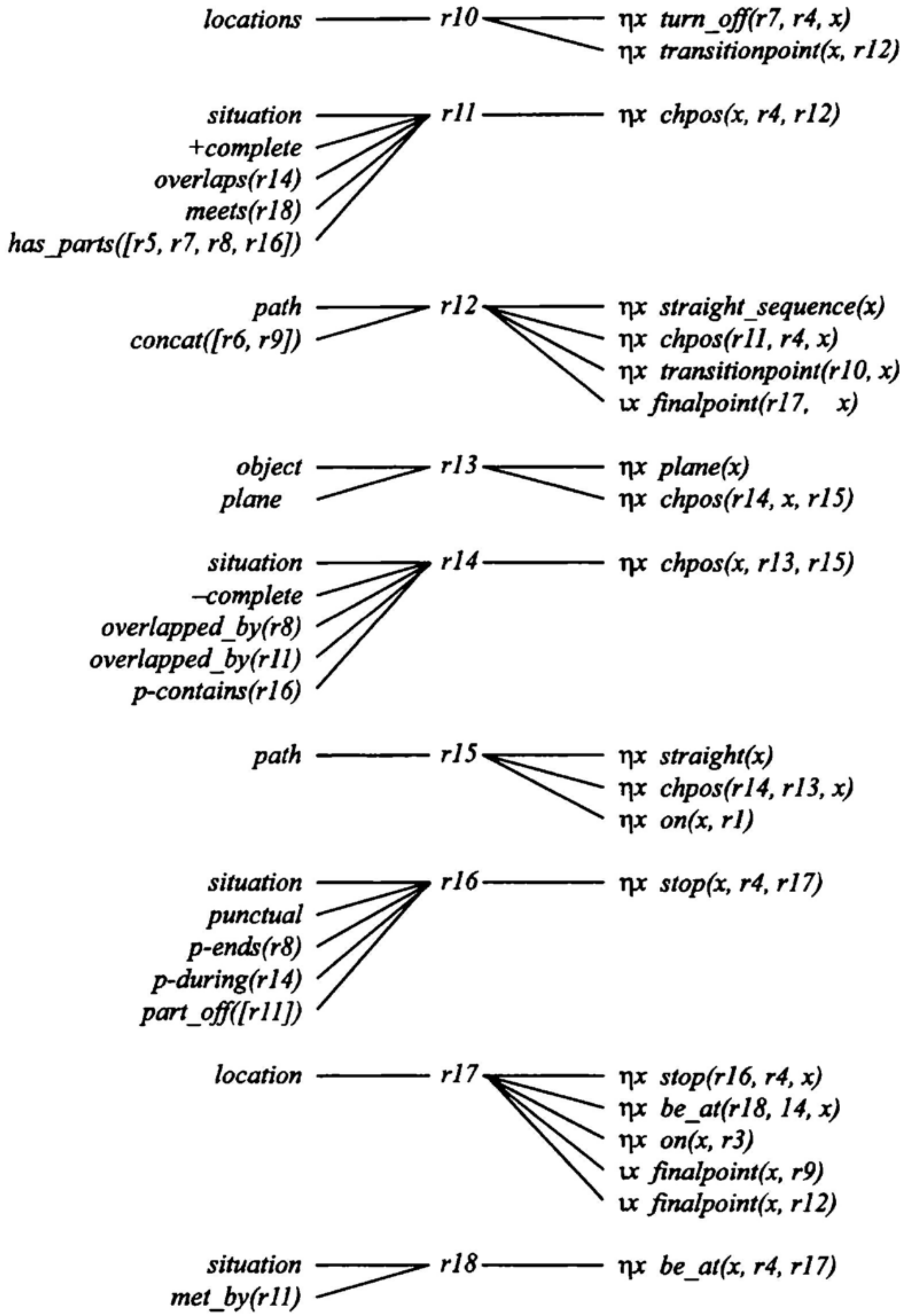


Figure 5. Conceptual representation of the observed scene (some information is left out for better readability; the possible temporal relations between *r14* and *r18* are left out, because they play no role for the discussion of this paper)

Attributes represent further properties. In the example there are two different kinds of object refOs, *plane* and *track*. Properties of situation refOs include the temporal relations introduced in section 3.1. The temporal relations that were introduced in the discussion above are transformed into attributes with one argument in the referential net, e.g. the relation *r5 meets r8* is transformed to *meets(r8)* as attribute of *r5* and to *met_by(r5)* as attribute of *r8*. Extended refOs have the information whether they are completed (*+/–complete*), punctual refOs are marked by the attribute *punctual*. Further attributes (*part_of* and *has_parts*[14]) establish the part–whole hierarchy of the represented entities, e.g. the situation representing the overall movement of the first plane (*r11*) consists of simpler movements (*r5*, *r7*, *r8*, *r16*).

The difference between punctual and extended events is considered in INC's conceptual representations by using a separate type of temporal relations for punctual events. For example, the punctual event *r7* (the plane changes its orientation) ends the extended event *r6* (the plane moves straight on runway 2), which is formally represented as *r7 p-ends r5* or *r5 p-ended_by r7* following a proposal of Vilain (1982). In our approach inferences between extended events can be seen as processes regarding extended events only. Thus, from the perspective of extended events, *r7* is not an event lying between *r5* and *r8*, and therefore *r7* does not interfere with *r5 meets r8*. On the other hand, the temporal relations involving punctual events allow to perform temporal reasoning in an Allen-like style (Vilain 1982).

Designations are either names or descriptions. They contain meaning-related knowledge about the entities. *r1*, for example, has the name *'RUNWAY1'* that can be used to refer verbally to the entity represented by the refO. Examples of descriptions are ηx *plane(x)* or ηx *chpos(r5, x, r6)*. The η-operator – a logical counterpart to the indefinite article – is used in the referential nets approach to link propositional representations with refOs (Habel 1986; Guhe, Habel, and Tschander 2003a). ηx *plane(x)* can be read as 'an object that is a plane' and ηx *chpos(r5, x, r6)* as 'an object, that changes position in situation r5 along path r6'.

A *transitionpoint* description represents the location where a moving object changed its direction, e.g. the location where the first plane turns off; a *straight_sequence* description – for lack of a better name – represents a movement of one bearer of motion along a succession of straight paths. The other descriptions should be self-explanatory.

4. The incremental conceptualizer INC

4.1. Overview

The incremental conceptualizer (INC, figure 6) is a model of the first part of the human language production faculty, the conceptualizer. In this section we give an account of the main components and functionality of INC, see also Guhe and Habel (2001), Guhe, Habel, and Tschander (2003b). Guhe (in prep.) provides a full description. Its input consists of a stream of visuo-spatial percepts (simple concepts), called *perceived entities*. This input stream is computed by a perceptual *pre-processing unit* (PPU) from the space–time coordinates of the observed motion events.[15] The PPU operates according to the empirically founded *cut-hypothesis* of Avrahami and Kareev (1994: 239): "A sub-sequence of stimuli is cut out of a sequence to become a cognitive entity if it has been experienced many times in different contexts." Perceived entities are used to build up a hierarchical conceptual structure, from which situation refOs are selected for verbalization. For each selected situation refO a preverbal message is generated. In this way INC forms preverbal messages out of perceived entities.

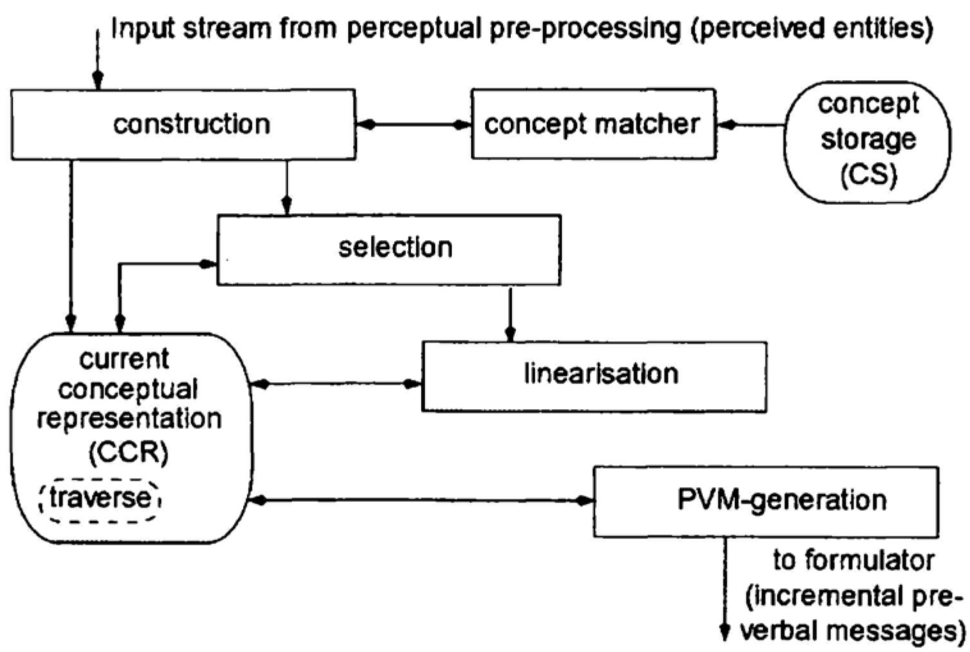


Figure 6. The architecture of the incremental conceptualizer (INC)

As already mentioned in section 1 the processes of incremental models are often depicted as a cascade of processes (Levelt 1989). This symbolizes the simultaneous processing of information on subsequent stages. In INC four processes form this cascade: *construction*, *selection*, *linearization*, and *PVM-generation* (preverbal message generation), depicted in figure 6. All processes operate on the *current conceptual representation* (CCR), which contains a referential net representing the currently observed state of affairs, e.g. the representation shown in figure 5. The *construction* process builds up the CCR from which *selection* chooses the situations to be verbalized. *Linearization* brings the selected situations into an appropriate order, and *PVM-generation* generates preverbal messages for the chosen situations. In this fashion the *traverse* is induced as a path through the CCR.

4.2. Construction

The construction process reads the perceived entities provided by the PPU and builds up the CCR from them. It is supported by the *concept matcher.* Each time a new perceived entity or a modification for an already existing one[16] is read from the PPU, construction calls the concept matcher to determine the best matching concept from the *concept storage.* The concept matcher has three possible results:

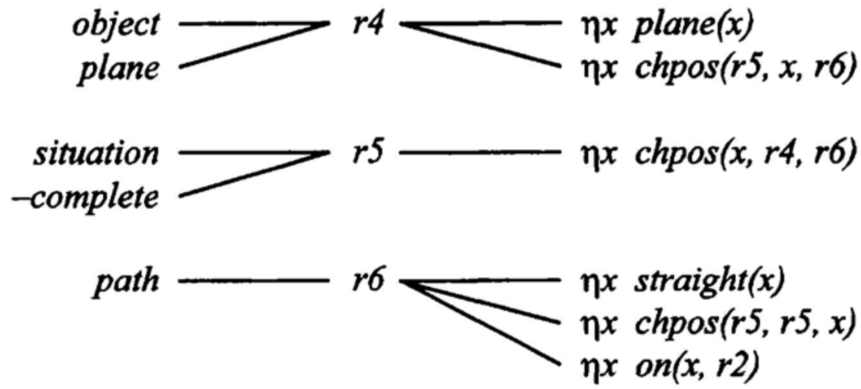
1. it finds an entity that subsumes the new entity and some "surrounding" ones, which are determined by a heuristic,
2. it finds an analysis for the entity, i.e. it finds simpler concepts,
3. it finds nothing.

In the first two cases the concept matcher determines the *best match* by computing the *degree of agreement* (DOA) between the entity or entities to be matched and the concepts stored in the concept storage. The best match is the concept with the highest DOA. A pair consisting of best match and DOA is given back to construction. The best match in case 1 need not be complete; that is, even if parts of this concept that are required for a complete match have not yet been perceived it can nevertheless be introduced into the CCR, with the missing parts marked as *expected.* Following events can then fulfill the expectations, in which case they are marked *regular*, or violate them. Then the expectation is *discarded* and a new one may be generated.

The decision whether the CCR is modified at all depends on the comparison of the DOA with the *degree of agreement threshold* (DOAT). If DOA ≥ DOAT, construction carries out one of three operations: *generate, modify*, or *discard*.

1. *Generate* introduces a new concept into the CCR. This operation is needed when (1) a new perceived entity comes in from the PPU or (2) when DOA ≥ DOAT for the best match given back by the concept matcher, and the best match is not yet contained in the CCR. In the latter case the best match and the missing entities are inserted into the CCR. The most usual case is that simpler concepts are matched onto a more complex one, which is the best match. If DOA < 1 such a complex concept is *expected.*
2. *Modify* is used when (1) the observed events demand an adaptation of an already existing perceived entity or (2) an element of the CCR that is marked *expected* is actually recognized. Then its status is changed to *regular*.
3. *Discard* is executed when expectations in the CCR no longer correspond to the best match. It removes all elements marked *expected.* After *discard* the operation *generate* is usually called to introduce (an expectation for) the new best match into the CCR. This implies that perceived entities cannot be discarded, because they never can be *expected.*

As we already indicated, the referential net given in figure 5 is a representation of the example scene after the scene has been observed completely. However, due to the online character of the task, verbalizations are generated before this is the case. Consider the first phase of the example scene. RefOs *r1–r3* can be considered as given in the representation before the observation starts. As soon as the first plane comes into view refOs *r4–r6* are added, but they do not contain all information given in figure 5.



The first difference is that no temporal relations exist at this stage (there is only one situation). The situation refO *r5* is still *–complete*, because the endpoint of the movement has not been observed yet. Consequently, the path refO *r6* does not contain the final point description *ιx finalpoint(r10, x)*, because *r10* as well as the turning off that ends this movement are not observed yet. Due to the same reason the plane refO *r4* does not have all designations it contains later on and there are no complex refOs (*r11*, *r12*).

The important point about this is the temporal interleaving of the processes construction and PVM-generation. When the first sentence of the example (15)a is generated only the first phase of the scene has been observed. Therefore, the generated verbalization that results in (15)a is generated with the CCR containing only *r1–r6* in the state indicated here. One effect is that *r5* represents a situation equivalent to the situation type *process*. Thus, different verbalizations are possible than when the plane has turned off and *r5* is equivalent to the situation type *event*.

4.3. Selection

Selection, like linearization and PVM-generation, operates on the *traverse buffer*. Its variable length is determined by the parameter *length of traverse buffer* (LOTB). The traverse buffer contains pointers to situation refOs. The refOs these pointers refer to can be modified, because no preverbal messages have been generated for them. Such a buffering is necessary for two main reasons. Firstly, linearization needs some time for its computations, and some selected situations have to be available simultaneously in order for linearization to be able to perform its operations. Secondly, selection needs a possibility for revising its choices, especially with the principle underlying the functioning of all processes: *Extended Wundt's Principle.*

Levelt (1989: 26) formulates *Wundt's Principle*: "an incremental process starts computing as soon as it obtains some characteristic input". Our extension of this principle consists in not only starting to process input as soon as possible but to produce output as quickly as possible as well.

The disadvantage of this method is that quite a lot of choices need to be revised. Yet, the advantage is that linearization has the maximum amount of time to operate on the elements in the traverse buffer, before they are taken out by PVM-generation. Furthermore, this method results in an *anytime* capability of INC: PVM-generation is capable of producing a preverbal message at any point in time. An additional reason for *Extended Wundt's Principle* is that from an empty traverse buffer no verbal output can be produced, which in the end leads to audible gaps in the verbalization.

Selection uses two operations for modifying the traverse buffer: *append* and *replace*. The former appends a pointer to a situation refO to the end of the traverse buffer, the latter replaces one or more pointer with another one. The position of the inserted pointer is the one of the headmost replaced pointer. Although we consider it likely that a deletion operation is needed someday, INC has none, because up to now we never needed it.

Selection applies its algorithms to the refOs that are inserted or modified by construction. If it decides to insert an element into the traverse buffer, i.e. a pointer to a situation refO, the time of insertion (a time stamp) is also saved with the pointer. Since the selection strategies are not in the focus of the present paper, please see Guhe and Habel (2001) and Guhe, Habel, and Tschander (2003b) for a description.

Note that we will often use the formulation "the situation refO in the traverse buffer" in the following instead of the correct "the pointer in the traverse buffer referring to a situation refO in the CCR". This makes the text better readable.[17]

4.4. Linearization

Linearization played only a minor role in our investigations up to now, because we did not investigate concurrent events like the two simultaneously moving planes. For this reason, linearization is not yet implemented. Since we do not need a fully functional linearization process for our present purposes, we only outline the effects of linearization. The effects of linearization on verbalizations is one of the major results we expect from our current study.

The purpose of linearization is to reorder the situation refOs in the traverse buffer so that they are verbalized in an appropriate order. For linearization the notion *event thread* proves useful: an event thread is a sequence of events that have the same bearer of motion, e.g. the first plane. Linearization may be done according to different, conflicting linearization strategies:

- Situations belonging to the same event thread will be grouped together. Then, a sequence in the traverse buffer of $A_1 - B_1 - A_2$ (of event threads A and B) is changed to $A_1 - A_2 - B_1$.
- Situations with temporal overlap will be grouped together; then, then sequence $A_1 - A_2 - B_1$ may be reordered to $A_1 - B_1 - A_2$.
- A situation that is connected to the one uttered last is put in head position.

4.5. PVM-generation

PVM-generation monitors the first element of the traverse buffer, the *head of traverse buffer*. When the specified *latency* (LT), computed as the difference between the current (system) time and the time stamp that was attached at insertion time expired, PVM-generation commences generating a new preverbal message. Since the preverbal messages are generated increment by increment we also call them *incremental preverbal messages* (Guhe, Habel, and Tappe 2000; Guhe in print).

Starting from the situation refO PVM-generation complements this refO with object and spatial refOs. Each time PVM-generation verbalizes a refO from the CCR it generates a new refO that contains the information of the CCR refO used in this particular verbalization. These refOs are called *verbalization refOs*. Each time PVM-generation produces such a refO it sends it to the formulator as increment of the current preverbal message and appends it to the traverse. The traverse is, therefore, the list of verbalization refOs (ordered by the sequence in which the verbalization refOs were created).

In the versions up to now PVM-generation had only access to the head of traverse buffer. However, in order to generate utterances like *While A B* this must be extended, because A and B must both be known to PVM-generation in order to compute the corresponding temporal connective that must be generated even before A. However, *B while A* could be generated without the extension. Thus, if there is a temporal relation between the head of traverse buffer and another refO in the traverse buffer, then the second refO can also be taken out and two preverbal messages are generated in one sweep. In section 5 we will describe this in detail. Additionally, PVM-generation now also has access to the traverse. This is the prerequisite for generating coreferences as it was described in section 2. Only if PVM-generation has the information that an object was already verbalized before it can decide to use a reduced referring expression.

4.6. Parameters

INC contains parameters (currently four) with which its behavior can be determined. In a simulation the parameters are assigned a value and cannot be changed by INC. Note that INC's behavior is not *determined* by the parameter values, because the processes run in (simulated) parallelism, and its behavior is much more dependent on the input from the PPU.

4.6.1. Degree of Agreement Threshold (DOAT)

The *Degree of Agreement Threshold* (DOAT) determines whether the *Degree of Agreement* (DOA) of a *best match* determined by the concept matcher is high enough, i.e. whether the match is good enough, to cause construction to work on the CCR. It is instantiated with a value between 0 and 1.

DOAT can be regarded as a technical necessity as well as being cognitively adequate. The crucial question is, whether humans do not generate expectations if the DOA is not high enough (in our terms: introduce expected refOs into the CCR) or whether they generate any expectation they are capable of creating but do not select these expectations from the CCR for verbalization. Since generating expectations is costly (in the sense that it requires time and other resources), the solution we have used in INC seems to be the plausible one, cognitively as well as technically.

By varying the value for DOAT the amount of expectations generated by the construction process is increased or decreased. The effect of DOAT = 0 is that construction uses any best match to work on the CCR. DOAT = 1 means that no expectations are generated at all.

4.6.2. Length of Traverse Buffer (LOTB)

The *Length of Traverse Buffer* (LOTB) specifies the number of elements that can be stored in the traverse buffer. It models the capacity of the working memory of the speaker that is available for the conceptualization task. Thus, the maximum value of LOTB is determined by an individual's current storage capacities.

If the value for LOTB is too small, elements get lost (are forgotten), because PVM-generation does not use them in time to produce a preverbal message. Thus, by increasing its value the chance is decreased that elements get lost and vice versa.

LOTB = 0 is not possible: INC would not work, because elements must be put in the traverse buffer in order to be accessible for linearization and PVM-generation; this would be like a working memory with no storage capacity, thus LOTB ≥ 1. There is no technical reason to limit the length of the traverse buffer because of processing limitations of INC. Thus, this value is purely motivated by cognitive considerations.

4.6.3. Latency (LT)

The *Latency* (LT) is the span of time that an element is kept in the traverse buffer, before it is taken by PVM-generation in order to produce a preverbal message for it. This time span is necessary for selection to be able to change its choices and for linearization to be able to perform its operations at all.

A lower value for LT has the effect that utterances are temporally closer to the event they describe. A higher value means that the produced output is better and more reliable, because selection and linearization have had more time for improvements, and more input from the pre-processing unit is available, which means that the CCR contains more reliable information about what is happening.

LT models the time pressure on the verbalization task. If the traverse buffer is permanently filled or if elements are even lost, the latency time can be reduced to compensate this – the speaker talks faster. This can be especially important when the value for LOTB is comparatively small.[18]

4.6.4. Attention Threshold (AT)

The *Attention Threshold* (AT) is a parameter that is used by one of the constraints with which incremental preverbal messages are generated. Items (designations in particular) are not used in a verbalization if their activation is below the value of AT.[19]

A lower value of AT causes more designations to be chosen by PVM-generation. This has two effects. First, the observed scene is described in greater detail; second, the number of generated pro-forms is reduced.

5. Coreferences and discourse relations in the generation of sequences of incremental preverbal messages

5.1. Commonalities of both generation tasks

In this section we spell out in detail how elements of incremental preverbal messages (increments) can be related in order to generate an interconnected (ideally coherent) discourse. More precisely, we consider the two cases discussed above, the generation of coreferences on the basis of refOs and the generation of temporal connectives on the basis of temporal relations. Since in Guhe, Habel, and Tschander (2003a, b) we already demonstrated how INC incrementally generates preverbal messages, we will not discuss these problems here.

We already described the notion of the traverse as the list of verbalization refOs in the temporal order in which they were created. It is, therefore, part of the discourse memory. Each time a refO is verbalized, i.e. each time a verbalization refO is generated for a refO in the CCR, the traverse is searched backwards for verbalization refOs to which the currently generated verbalization refO can be linked. For the two cases we consider here this means that if an object refO is verbalized it is checked whether there already exists a verbalization refO for it. In the referential nets approach this is a very cheap operation due to the fact that each refO can be uniquely identified by the term that serves to refer to it. In the case of a situation refO the traverse is searched backwards for verbalizations of situation refOs to which it can be linked. There are two prime candidates for this: firstly, the last situation refO of the same event thread, secondly, the last verbalization of a situation refO (of all situation refOs).[20] In the first case a coreference is generated, in the second case a temporal connective. Both computations have in common that they depend on last elements in the traverse. However, if an object refO is to be verbalized this operation can become quite complex, because a coreference cannot only be established between the last object and the currently generated one. Many other objects may be mentioned in between. Thus, the set of all possible referents has to be determined, and it must be computed how the object can be distinguished from the others elements in this context set.

Note that what follows does not describe an algorithm used by PVM-generation but explores different possibilities that can be used by it. The preferred expressions and factors that favor one verbalization over another one are determined in an empirical study currently carried out.

5.2. Generating coreferences

The task of finding verbal descriptions for objects is an instance of the task called *generation of referring expressions*. Generating coreferences makes this task more intelligent in that it not simply finds a referring expression but exploits cases in which an object is already introduced in the discourse. The reason is that in this case the object to be referred to needs not be distinguished form all objects present in the scene but only from those that were already introduced. Thus, the set of possible referents is reduced.

Our incremental algorithm for the generation of coreferences in sequences of preverbal messages follows the overall idea of the algorithm for the generation of referring expressions by Dale and Reiter (1995). For the generation of referring expressions this algorithm is by now regarded as baseline by most NLG researchers against which alternative algorithms are tested. The strengths of this algorithm are that it

- observes the Gricean maxims of conversation,
- is fast (has linear run-time),
- is psychologically plausible.

It generates referring expressions that single out one object from a set of objects, called *distractor set*. Each object is characterized by a set of *attribute–value pairs*, also called *properties* for short.[21] The algorithm finds the subset of properties that uniquely distinguishes the object from the other objects in the distractor set. This is done in a monotonic incremental fashion, i.e. for each property of the target object it is checked whether using it in the referring expression removes other objects from the distractor set until no other object shares the set of properties with the target object.

The kind of incrementality used here is called monotonic, because if a later addition of a property renders an earlier added one superfluous the earlier property is not removed (Guhe in prep.). Without monotonicity the algorithm loses it linear run-time.

By now a number of extensions and variants of this algorithm exist, see, for example, van Deemter (2002). For the purposes at hand two are especially relevant. Krahmer, van Erk, and Verleg (2002) propose to reformulate the algorithm in graph-theoretical terms so that results from theoretical informatics can be used to enhance the algorithm's capabilities. Due to the network structure of referential nets this is a natural way of capturing the Dale and Reiter algorithm. Yet, the question of how the graph-based variant is applied to referential nets is far more than we need for the purpose at hand, and a detailed discussion would only lead us astray.[22]

A variant of the algorithm addressing problems similar to ours is the one by Krahmer and Theune (2002). They describe a method for generating *reduced anaphoric* referring expressions. These referring expressions are not distinguishing in the sense that they uniquely single out an object from the distractor set, but they suffice to single it out from a distractor set smaller than the set of all possible distractors. This smaller distractor set consists of the most salient objects in the current context, in particular the objects that were previously mentioned in the discourse. This builds on the assumption that speaker and hearer build up a similar set of possible referents so that the reduced expression suffices. Finding the set of possible referents is done by using *salience weights*. In INC the salience weights correspond to activation values (Guhe, Habel, and Tschander 2003b). However, since in the scene at hand we just have to distinguish two planes, we can leave the matter of salience aside and assume that the distractor set consists of maximally two planes – one before and two after the second plane comes into view. Yet, a full solution will exploit the different refO activations.

In the domain of moving planes we have a setting that differs in an important respect from the ones usually used: the objects referred to (the planes) look identical. That means, they cannot be distinguished by shape, form, size, color, or similar attributes. Thus, they must be distinguished by localizing them on the maneuvering area, by the situations they are (previously were) involved in, or by the sequence in which they were previously mentioned. We will mainly consider the last possibility. Note that our approach is capable of explaining all three kinds of discourse anaphora proposed by Webber et al. (to appear, see section 2.2), while the algorithms put forward so far can only explain the first two.

A consequence of this is that the *preferred attributes list* must be extended. In the Dale and Reiter algorithm this list gives the order in which the available attributes are evaluated in order to find out whether using the property (attribute–value pair) in the referring expression for the target object removes elements from the distractor set. The list originally used by Dale and Reiter is ⟨type, color, size⟩. That means first the type attribute is checked, e.g. if using the property *type(plane)* removes objects with property *type(helicopter)* from the distractor set it is chosen. After this, the color, then the size of the object are tried. Krahmer and Theune extend this list by spatial (generally: relational) attributes like *on runway 2*. Their list is ⟨type, color, size, spatial⟩.

After what we said above this extension alone does not suffice for our purposes, because it does not enable us to generate referring expressions that take the order in which the objects were mentioned into account. Hence, using spatial properties allows to generate utterances like the ones in (17). However, if it is to be verbalized that the first plane stops on the connection only a somewhat strange verbalization like the one in (18) can be generated. Furthermore, in the Krahmer and Theune algorithm it is not possible to generate utterances like the ones in (19), because this requires using discourse related properties. So, we propose the following attribute list: ⟨type, color, size, discourse, spatial⟩. We regard discourse related properties as more preferred than spatial ones, i.e. we regard the referring expression *das erste Flugzeug* 'the first plane' in (19)a as preferred over the expression *das Flugzeug auf Runway 2* 'the plane on runway 2' in (17)a. This still must stand the test of empirical evaluation as well as finding the factors (contexts) that lead to favoring one over the other.

(17) a. *Das Flugzeug auf Runway 2 biegt ab.*
'The plane on runway 2 turns off.'
b. *Das Flugzeug, das abgebogen ist, hält an.*
'The plane that turned off stops.'

(18) *? Das Flugzeug auf der Verbindung stoppt auf der Verbindung.*
'? The plane on the connection stops on the connection.'

(19) a. *Das erste Flugzeug stoppt auf der Verbindung.*
'The first plane stops on the connection.'
b. *Das zweite Flugzeug fährt auf Runway 1.*
'The second plane moves on runway 1.'

c. *Ein anderes Flugzeug fährt auf Runway 1.*
'Another plane moves on runway 1.'

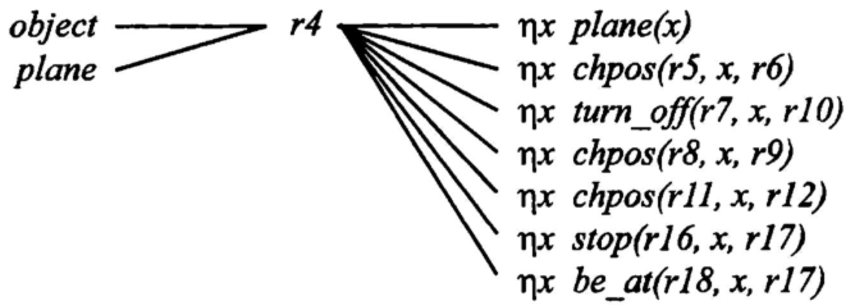
Discourse related properties differ from the other properties in two respects. Firstly, they are relational descriptions not with respect to other objects in the scene but with respect to the order in which they were mentioned. Secondly, the properties are computed and added to the verbalization refOs during the generation of a referring expression. Thus, they are not present in the CCR.

All in all there are four ways to generate coreferences:

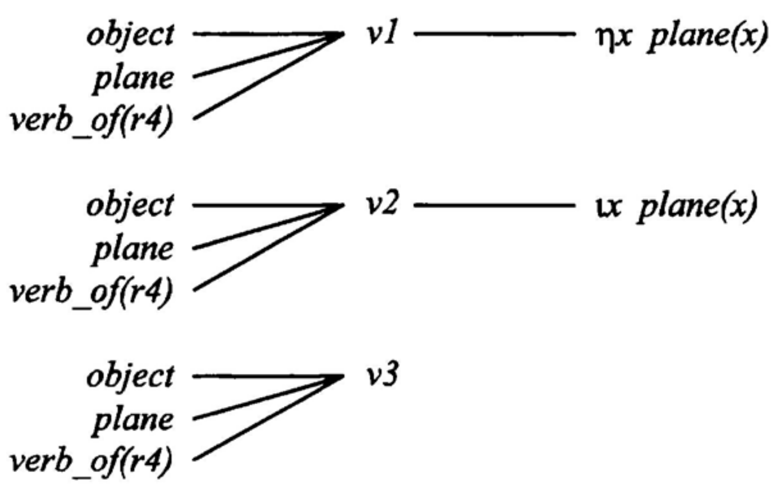
- using the ι-operator instead of the η-operator, i.e. using a definite instead of an indefinite description (*das Flugzeug* 'the plane' instead of *ein Flugzeug* 'a plane'),
- choosing no designation to describe the refO and relying on the coreference, which the formulator may realize anaphorically, e.g. as pronoun (*es* 'it'),
- leaving the referent implicit, e.g. by means of an NP ellipsis,
- using a semantically related concept, e.g. *Maschine* 'vehicle' instead of *Flugzeug* 'plane'.

These possibilities differ with respect to how much of the overall computation is done in which component (conceptualizer or formulator), depending on how much of the conceptual and discourse knowledge is required for the computation. In the first case the computation is done entirely in the conceptualizer, in the second case it is shared between both. In the third case it is done entirely in the formulator, more precisely, the decision to use an anaphor or an ellipsis is made in the formulator on the basis of the same preverbal message, see also Guhe and Schilder (2002b). In the fourth case it most likely is a combination of conceptualizer and formulator. This will strongly depend on the assumptions made in the overall model. We will only take into account the first two cases here.

Consider *r4* in its final state. There are two descriptions that can be used for generating a referring expression: ηx *plane(x)* and ηx *be_at(r18, x, r17)*. Since the former is *directly grounded* it is preferred over the second one. Directly grounded means that it is referring to no other refO (Guhe, Habel, Tschander 2003a). All other designations relate *r4* to situations or spatial entities.[23]

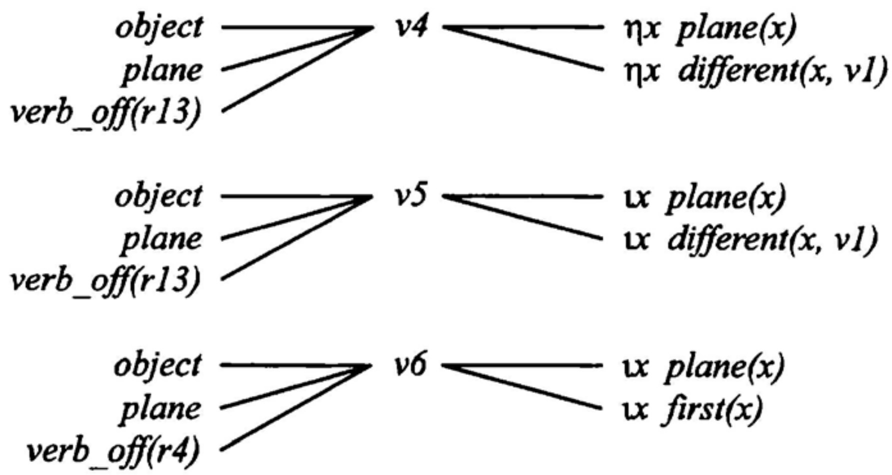


Three different verbalization refOs can be generated form the *ηx plane(x)* description, given as *v1* to *v3*. Note that the verbalization refOs given in the following will only contain the information that is important for the discussed examples. The verbalization refOs generated by INC contain more information, and the output of INC can also be tailored to the requirements of a formulator using its output as input.

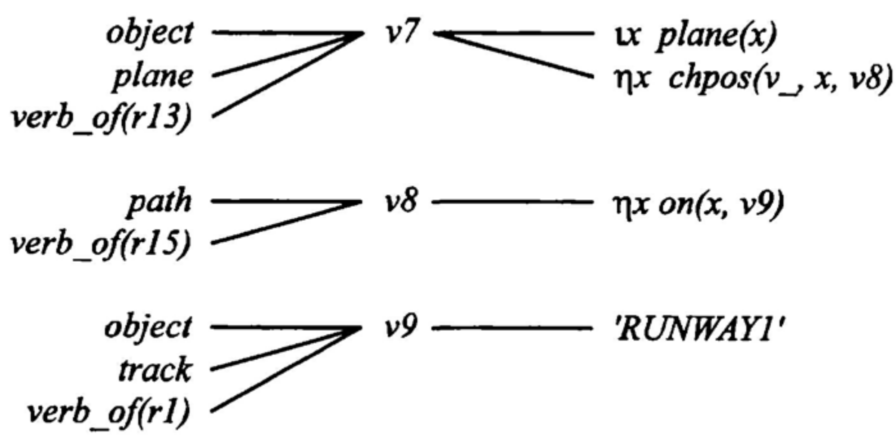


Since verbalization refOs are a different kind of refO than the other refOs in the CCR, the terms naming them contain a *v* instead of an *r*. *v1* corresponds to a verbalization of the referring expression *ein Flugzeug* 'a plane'; *v2* to *das Flugzeug* 'the plane'. The coreference of both verbalization refOs is expressed by *verb_of(r4)*, i.e. the CCR refO from which both verbalization refOs were generated is *r4*. Thus, both verbalizations refer to the same entity. *v3* shows the case that no designation was chosen. The underlying rule currently used by INC is that if a designation was already used it is not used again – except if it is required by the discourse, e.g. to distinguish an object from another one, see below. Thus, in (15)b a verbalization refO like *v3* is used, because the *ηx plane(x)* description, from which *ιx plane(x)* would be computed, was already used in (15)a.

If the distractor set consists of both planes like in the second part of (15)b and in (15)c, it is sometimes necessary to suppress coreferences. In section 2 we argued that this can be done by way of two linguistic means, corresponding to the expressions *ein anderes Flugzeug* 'another plane' (*v4*), *das andere Flugzeug* 'the other plane' (*v5*), and *das erste/zweite Flugzeug* 'the first/second plane' (*v6*). The indefinite variant is chosen in the first mentioning, the definite ones afterwards.



Referring expressions like *the plane on runway 1*, which is used in (15)c, are generated out of verbalization refOs like *v7* with additional ones like *v8* and *v9* (the situation refO *v_* is not given in this example).



5.3. Generating temporal connectives

The temporal relations introduced in section 3.1 are represented as attributes at the situation refOs. Therefore, they are not considered by the grounding algorithm of PVM-generation with which an incremental preverbal message is produced (Guhe, Habel, and Tschander 2003a, b). For our present purpose we make the assumption that there is always exactly one temporal relation between two preverbal messages. That is, each time the generation of an incremental preverbal message commences it must be linked to a previous one. Since in INC the first increment of an incremental preverbal message always is the verbalization of a situation refO, the temporal relations between preverbal messages can be established on the basis of the temporal relations between the situations. As default case we regard that the event for which currently a preverbal message is generated occurred after the one that was described in the previous verbalization (see also section 2.3).

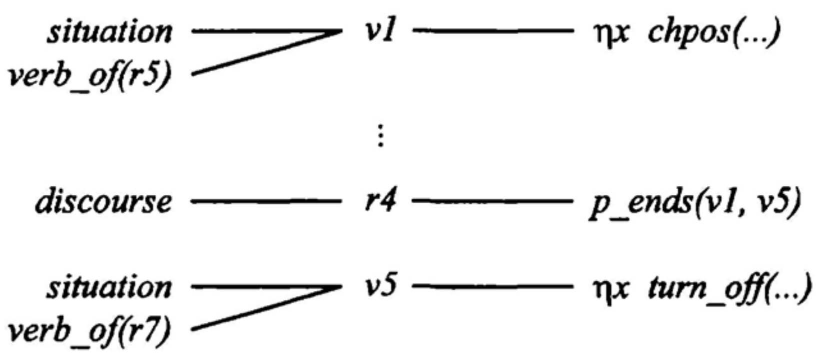
Consider verbalization (15), which is repeated here as (20). The CCR refOs representing the corresponding situations are given in brackets.

(20) a. *Ein Flugzeug fährt auf Runway 2 (r5, e_1).*
'A plane is moving on runway 2.'

b. *Nachdem es auf die Verbindung abgebogen ist (r7, e_2), fährt ein anderes Flugzeug auf Runway 1 (r14, e_3).*
'After it has turned off onto the connection, another plane moves on runway 1.'

c. *Dann stoppt das erste Flugzeug (r16, e_4), während das Flugzeug auf Runway 1 weiterfährt (r14, e_5).*
'Then the first plane stops, while the plane on runway 1 is moving on.'

Utterance (20)a describes the situation represented by *r5*. It is generated when only refOs *r1* to *r6* are present in the CCR. After the plane turned off onto the connection, refOs *r7–r10* together with *r11* and *r12* are added to the CCR. Assume that the next refO to be verbalized is *r7*. In INC terminology: *r7* is the head of traverse buffer. The temporal relation between the two preverbal messages for *r7* and *r5* can now simply be determined, it is *r7 p-ends r5*. For representing the temporal relation between the two preverbal messages an additional verbalization refO is generated and sent to the formulator. We term this kind of refO *discourse refO*, because these refOs can connect incremental preverbal messages also by other than temporal relations, e.g. causal ones.[24] Assume that the incremental preverbal messages resulting in (20)a and the first part of (20)b have as their first increment the verbalization refOs *v1* and *v5*, *v1* being a verbalization of *r5*, *v5* being a verbalization of *r7*. Then, there is a discourse refO *v4* connecting the two.

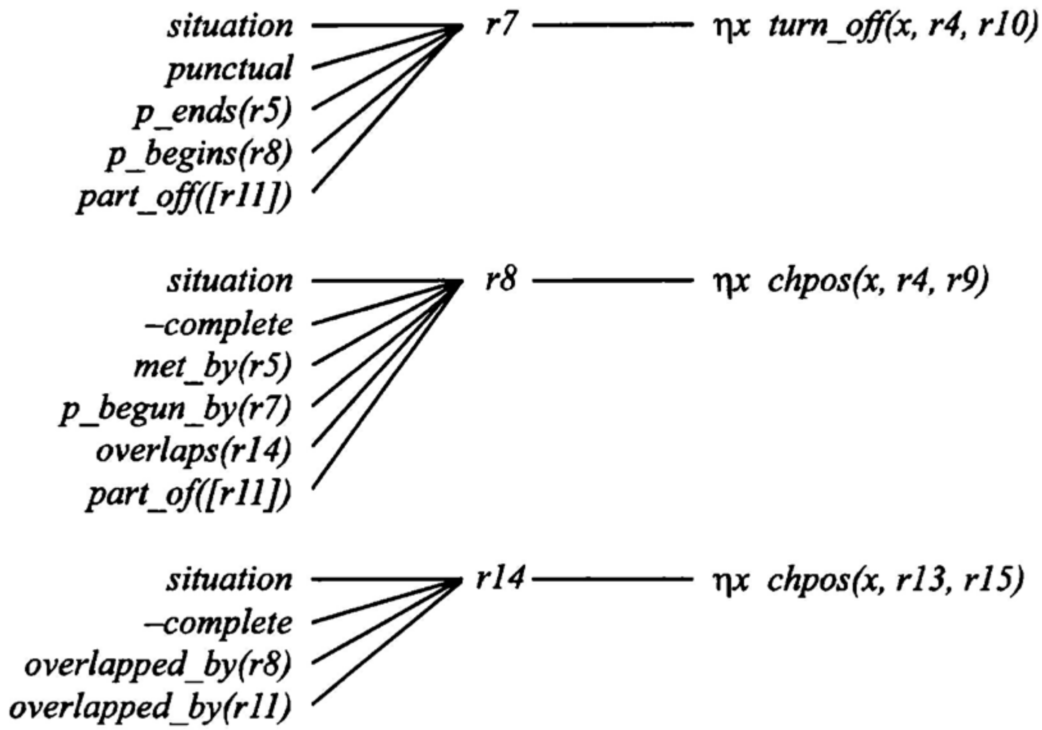


As you will have observed the one-place temporal relations of *r5* and *r7* became one two-place functional expression at *v4*. In this example doing so may seem superfluous: why not leave the temporal information at the situation refOs? As we will see below there is one major argument: it must be possible to establish the temporal structure of multiple preverbal messages in advance. This is necessary, because the grammatical tense of the first part of (20)c, for example, depends on its temporal relation to the second part.

Before *v4* and *v5* are actually generated a second computation takes place, viz. that PVM-generation establishes the temporal connection between the refOs the pointers in the traverse buffer refer to. We assume that a temporal relation can be established between all situation pairs. While this may be not generally true, for the scene at hand it certainly is.

The reason for this is that once PVM-generation starts generating incremental preverbal messages it tries to verbalize as many elements of the traverse buffer as possible.[25] Since, typically, no more than three to four elements are stored in the traverse buffer, the resulting utterances do not become too long. However, PVM-generation is concerned with the generation of preverbal messages, not sentences. Thus, it is required by a formulator using INC's preverbal messages that the utterances are generated in an appropriate form.

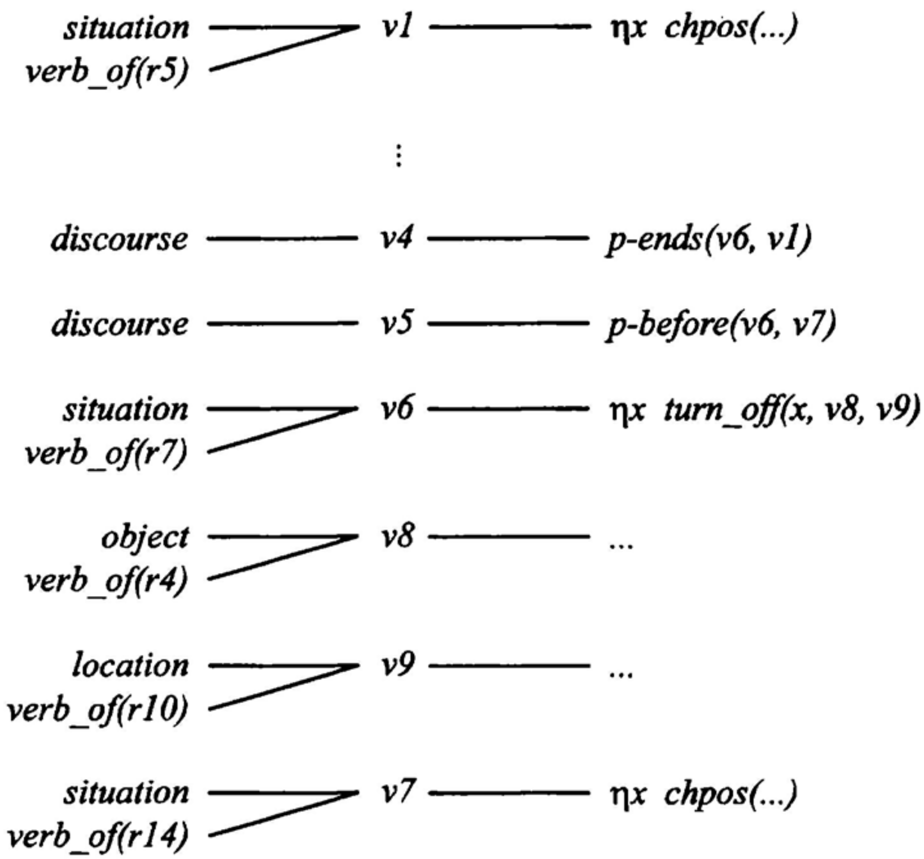
When utterance (20)b is generated the traverse buffer contains two elements that point to *r7* and *r14*. As no direct temporal relation exists between the two refOs, PVM-generation has to determine them first. At this point of time the CCR refOs look like follows:



Both, *r7* and *r14* have a temporal relation to *r8*. The relation *r8 overlaps r14* states that *r8* started before *r14*, i.e. the first plane moves on the connection before the second plane moves on runway 1. Furthermore, *r7 p-begins r8*, i.e. the first plane turned off before it moved on the connection. From this one can conclude that *r7* must be finished before *r14* starts, because it is punctual. Therefore, the relation *r7 p-before r14* and its inverse, *r14 p-after r7*, hold. The first point of *r8* cannot coincide with the first point of *r14*, because of the *overlap* relation. It would only be the case if *r8 p-begins r14* held. Written as a general rule:

$$\{s_1 \textit{ p-begins } s_2,\ s_2 \textit{ overlaps } s_3,\ s_1 \textit{ punctual}\} \models \{s_1 \textit{ p-before } s_3,\ s_3 \textit{ p-after } s_1\}$$

The inference can be made by substituting s_1 with *r7*, s_2 with *r8*, and s_3 with *r14*. Thus, the verbalization of (20)b starts not only with one discourse refO connecting *v5* to *v1* but with two, the second one announcing that another incremental preverbal message (verbalizing *r14*) will follow immediately. The formulator can then generate a sentence frame of the form *Nachdem A ... B ...* 'After A ... B ...' on the basis of the temporal relations and the situation types (*punctual, +/–complete*). The outline of the incremental preverbal messages, therefore, is as follows:



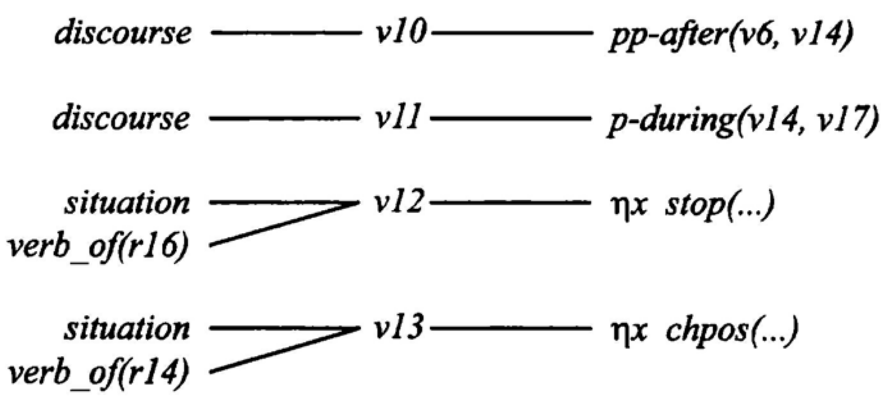
Note that *v7* is announced rather early in comparison to the point of time at which it is actually generated: before that, the verbalization refOs *v8*, *v9* that are needed to complement *v6* – and perhaps some more – are sent to the formulator. These verbalization refOs were announced by the description *ηx turn_off(x, v8, v9)*. Since they are required for a complete preverbal message, they must be generated first. But see note 23 on the generation of preverbal messages within preverbal messages.

When the next utterance (20)c is generated the traverse buffer contains pointers to refOs *r16* and *r14* (in this order). *r16* can be related to two different discourse segments, e_2 (verbalized as *v6*, CCR refO *r7*) and e_3 (*v7*, *r14*). Since *r16* belongs to another event thread than the last verbalized situation refO *v7* (*r14*), the traverse is searched backwards further to the verbalization of *r7* (*v6*), belonging to the same event thread as *r16*.[26] Attaching *r16* to *r7* is rather straightforward. The temporal relation is *p-after* because of the following rule:

$$\{s_1 \textit{ p-begins } s_2,\ s_3 \textit{ p-ends } s_2,\ s_1 \textit{ punctual},\ s_3 \textit{ punctual}\} \models \{s_1 \textit{ pp-before } s_3,\ s_3 \textit{ pp-after } s_1\}$$

s_2 in this case is instantiated with *r8*, s_1 with *r7*, and s_3 with *r16*. *pp-before* and *pp-after* are relations between two time points (punctual events). Remember that *p-before* and *p-after* relate a time interval and a time point. The two rules for making inferences on temporal relations given in this section are, of course, only examples of an encompassing set of rules.

Taking this together the following verbalization refOs establish the interconnections of the incremental preverbal messages for an utterance like (20)c:



6. Conclusions

We explored two means for interconnecting preverbal messages that are generated in an incremental fashion: coreference and temporal relations. Such interconnections are part of the overall problem of establishing coherence in the generated discourse.

We used two different techniques for the two tasks. Coreferences are established by using refOs (referential objects) from the CCR (current conceptual representation) that were already used earlier in the discourse. In INC this means that two or more verbalization refOs are generated for one refO in the CCR. On the linguistic surface such refOs can be realized by reduced referring expressions, e.g. pronominal anaphora. Temporal connectives are generated from the temporal relations between situation refOs, which appear, for example, as *nachdem* 'after' on the linguistic surface, but they can also stay implicit with regard to the *order of mention contract*.

For the INC conception it is important to observe that it is an incremental model. For the problem we discussed in this paper this means that not only the generation of preverbal messages but also of the construction of the CCR takes place incrementally. Consequently, observation of the scene, construction of the CCR, and generation of preverbal messages (and all other tasks INC performs) are temporally interleaved. This way of processing enables the system to keep track of what is going on, and to generate utterances describing the observed events shortly after they took place. Additionally, it facilitates the selection task (deciding what to say), because only the most recent events must be considered, and the coherence inducing interrelations are established with what was said last.

Notes

1. The research reported in this paper was conducted in the project ConcEv (Conceptualizing Events), which has been supported by the DFG in the priority program "Language Production" under grant Ha-1237/10 to Christopher Habel.
2. These verbalization experiments, which will be finished at the end of 2003, are the basis for comparing human descriptions of the presented events with the verbalizations generated by INC.
3. Concepts stored in long-term memory are another source for constructing an internal conceptual representation. However, we focus on the data-driven aspects of conceptualization so as to better control the content of conceptual representations.
4. Since the verbalizations we record in our studies are German, we provide German examples. The English glosses are only rough translations.
5. Another linguistic phenomenon depending on coreference are NP ellipses as in (i), where the bearer of motion in the second statement must be identified with the one in the first statement.

   (i) The plane is moving on runway 2 and turns off to the left at the branch.

   We do not focus on the production of elliptic constructions in this paper but see Guhe and Schilder (2002a, b) and Schilder and Guhe (2002), who demonstrate how the INC approach can be combined with underspecification semantics in order to incrementally generate coordinative structures for self-corrections, VP ellipses, and gapping constructions.

6. Since these discourse anaphora establish conceptual coreference, Webber et al. (to appear) call them *coreferential discourse anaphora.*
7. Discourse anaphora of this kind are indirectly related to conceptual coreference. Therefore, they are named *indirect discourse anaphora* by Webber et al. (to appear).
8. Other examples of discourse anaphora of this kind are *other, such, elsewhere* or *larger*. They determine lexically the function that is applied to the relation between the denotation and the referent. Therefore, Webber et al. call them *lexically-specified discourse anaphora.*
9. We consider only temporal factors (relations). For the example it is quite likely that one would use utterances mentioning a (conjectured) causal relation like the following:
   (ii) The first plane stops, *because* the second plane moves on.
10. *Während* 'while' has two meanings, a temporal and a contrastive one. However, we exclusively consider the temporal meaning here.
11. This orientation to cognition distinguishes referential nets from kindred approaches of linguistic semantics, e.g. *discourse representation theory* (Kamp and Reyle 1993). The referential net representations we use in the *current conceptual representation* (CCR) of INC correspond to the conception of *situation model* used by Lewis (1993) in an NL-Soar approach.
12. The identifier *s1*, which corresponds to kindred naming conventions in DRT (Kamp and Reyle 1993), is used in this section, to remind the reader, that *s1* represents a situation. In section 3.2, which describes the referential net approach in more detail, all refOs are named uniformly with identifiers beginning with an *r*. The information about the ontological type (*situation*) is attached to a refO as an attribute.
13. However, a situation in which a movement takes place that is *–complete* is just another representation of a process. Similarly, for the other situation types.
14. In earlier publications the *has_parts* attribute was called *sum*. However, the more complex concepts are not sums in the sense proper, e.g. *r11* is not a sum of situations like a cube is the sum of its six sides, because a situation can consist of many different combinations of sub-situations.
15. Technically, the position of all movable entities is recorded each 50 ms. The PPU then computes simple movements from the sequence of these snapshots. Based on these simple movements triples of perceived entities are computed, consisting of the moving object, the situation, and the path of motion (or the location).
16. A perceived entity needs to be modified if, for example, the movement of an object is made known to INC by the PPU before the movement is finished. Due to the incremental character of processing this is the normal case. Consider a moving plane. As soon as the PPU detects this movement it generates the corresponding perceived entities for INC. If the plane then stops moving this in-

formation must be sent to INC as well as if it continues moving, especially if the path is segmented due to a change of orientation.

17. The traverse buffer contains pointers to situation refOs in the CCR, not the situation refOs themselves, because the refOs in the CCR can change while they are "in" the traverse buffer. Such a change to the CCR refO would then also have to be applied to the situation refOs in the traverse buffer, which is a needless doubling of operations.
18. LT is the prime candidate for a parameter that is only set to an initial value and changeable during a simulation by the system. That would mean that the mentioned decision to talk faster is made by INC and already not before the simulation is started.
19. This is not completely true; if an element is required for a well-formed preverbal message then it is selected even if its activation is below AT.
20. It is always a situation refO that establishes the link between two incremental preverbal messages, because the first refO of each preverbal message verbalizes a situation refO. Only situation refOs are stored in the traverse buffer, which is, of course, no general solution, but since we are concerned with the verbalization of *events*, it suffices for our purposes.
21. Attributes in the sense of Dale and Reiter comprise attributes as well as designations in the referential net formalism. Attributes and designations in referential nets, however, encompass much more possible expressions than the attributes of Dale and Reiter.
22. Roughly, a referential net can be translated into the graph theoretical approach as follows. RefOs correspond to the nodes of the graph. Directly grounded descriptions – descriptions that refer to no other refO (Guhe, Habel, and Tschander 2003a) – as well as names are edges that refer back to the refO. Two-place descriptions are edges between two nodes (refOs). Three-place (n-place) relations must be transformed into two (n – 1) two-place relations, introducing auxiliary nodes (refOs). The same must perhaps be done for some special attributes that are used for verbalization. This allows to use the algorithm described in Krahmer, van Erk, and Verleg (2003) more or less unmodified.
23. Selecting one of these descriptions might, however, be used for generating relative clauses, e.g. *The plane that moved on runway 2 moves on the connection*. This would be a case of recursive generation of preverbal messages: the generation of one preverbal message is temporarily suspended in order to generate another one specifying an increment of the embedding preverbal message. Note that our approach does incorporate the possibility to use this means for singling out an object from a distractor set, because finding a referring expression for an object is just one particular case of finding a description for some entity. The Krahmer and Theune algorithm does not encompass this possibility.

24. Other discourse refOs are used in Guhe and Schilder (2002a), where they represent correction terms. These in turn are similar to discourse refOs establishing coordination relations in the generation of VP ellipses, simply called *list* in Guhe and Schilder (2002b).
25. In this respect the version of INC proposed here differs from earlier ones, which only verbalized one preverbal message at a time. In those version it was not necessary to compute relations between elements of the traverse buffer.
26. An additional reason for not using the last situation refO in the traverse could be that one of the events to be verbalized is the one described by the previous preverbal message (e_3 = e_5, *r14*). This possibility must be explored when evaluating the empirical verbalization data.

References

Allen, James F.
1984 Towards a General Theory of Action and Time. *Artificial Intelligence* 23: 123–154.
Asher, Nicholas, and Alex Lascarides
2003 *Logics of Conversation.* Cambridge: Cambridge University Press.
Avrahami, Judith, and Yaakov Kareev
1994 The Emergence of Events. *Cognition* 53: 239–261.
Bach, Emmon
1986 The Algebra of Events. *Linguistics and Philosophy* 9: 5–16.
Clark, Herbert H., and Eve V. Clark
1977 *Psychology and language.* New York: Harcourt Brace Jovanovich.
Dale, Robert, and Ehud Reiter
1995 Computational Interpretations of the Gricean maxims in the Generation of Referring Expressions. *Cognitive Science* 18: 233–263.
Eschenbach, Carola, Ladina Tschander, Christopher Habel, and Lars Kulik
2000 Lexical Specification of Path. In *Spatial Cognition II*, Christian Freksa, Wilfried Brauer, Christopher Habel, and Karl F. Wender (eds.), 127–144. Berlin: Springer.
Grice, Paul
1975 Logic and Conversation. In *Syntax and Semantics* (vol 3), Peter Cole and Jerry L. Morgan (eds.), 41–58. New York: Academic Press.

Guhe, Markus
2003 Incremental preverbal messages. In *Mediating between Concepts and Grammar*, Holden Härtl and Heike Tappe (eds.). Berlin, New York: Mouton de Gruyter.
in prep. Incremental conceptualisation for language production. PhD thesis, University of Hamburg.
Guhe, Markus and Christopher Habel
2001 The Influence of Resource Parameters on Incremental Conceptualization. In *Proceedings of the Fourth International Conference on Cognitive Modeling*, Erik M. Altmann, Axel Cleeremans, Christian D. Schunn, and Wayne D. Gray (eds.), 103–108. Mahwah, NJ: Lawrence Erlbaum.
Guhe, Markus, Christopher Habel, and Heike Tappe
2000 Incremental Event Conceptualization and Natural Language Generation in Monitoring Environments. *INLG-2000*, 85–92, Mitzpe Ramon, Israel. (Published by ACL).
Guhe, Markus, Christopher Habel, and Ladina Tschander
2003a Describing Motion Events: Incremental Representations for Incremental Processing. In *Proceedings of the 5th International Workshop on Computational Semantics (IWCS-5)*, Tilburg, The Netherlands, 410–424.
2003b Incremental Production of Preverbal Messages with INC. In *Proceedings of the 5th International Conference on Cognitive Modeling (ICCM-5)*, Bamberg, Germany.
Guhe, Markus, and Frank Schilder
2002a Incremental Generation of Self-corrections Using Underspecification. In *Computational Linguistics in the Netherlands 2001. Selected papers from the twelfth CLIN meeting*, Mariët Theune and Anton Nijholt, and Hendri Hondorp (eds.), 118–132. (Language and Computers: Studies in Practical Linguistics 45.) Amsterdam,/New York: Rodopi.
2002b Underspecification for Incremental Generation. In *Proceedings of KONVENS 2002, 6. Konferenz zur Verarbeitung Natürlicher Sprache*, Saarbrücken, Germany, 37–43.
Habel, Christopher
1986 *Prinzipien der Referentialität*. Berlin: Springer.
1987 Cognitive linguistics: The Processing of Spatial Concepts. *T. A. Informations (Bulletin semestriel de l'ATALA, Association pour le traitement automatique du langage)* 28, 21–56.
Kamp, Hans, and Uwe Reyle
1993 *From Discourse to Logic*. Dordrecht: Kluwer.

Kehler, Andrew
2002 *Coherence, Reference, and the Theory of Grammar*. Stanford, CA: CSLI Publications.
Kempen, Gerard, and Edward Hoenkamp
1987 An incremental procedural grammar for sentence formulation. *Cognitive Science* 11: 201–258.
Krahmer, Emiel, and Mariët Theune
2002 Efficient Context-sensitive Generation of Referring Expressions. In *Information Sharing: Reference and Presupposition in Language Generation and Interpretation*, Kees van Deemter and Rodger Kibble (eds.), 223–264. CSLI Publications.
Krahmer, Emiel, Sebastiaan van Erk, and André Verleg
2003 Graph-based Generation of Referring Expressions, *Computational Linguistics* 29(1): 53–72.
Lascarides, Alex, and Nicholas Asher
1993 Temporal Interpretation, Discourse Relations and Common Sense Entailment. *Linguistics and Philosophy* 16(5): 437–483.
Levelt, Willem J.M.
1989 *Speaking: From Intention to Articulation*. Cambridge, MA: Cambridge University Press.
1999 Producing spoken language: a blueprint of the speaker. In *The neurocognition of language*, Colin M. Brown and Peter Hagoort (eds.), 83–122. Oxford: Oxford University Press.
Lewis, Richard L.
1993 An Architecturally-based theory of Human Sentence Comprehension. PhD thesis, Carnegie Mellon University.
van Deemter, Kees
2002 Generating Referring Expressions: Boolean Extensions of the Incremental Algorithm. *Computational Linguistics* 28(1): 37–52.
Vilain, Marc B.
1982 A system for reasoning about time. In *Proceedings of AAAI-82*. 197–201.
Webber, Bonnie, Aravind Joshi, Matthew Stone, and Alistair Knott
to appear Anaphora and Discourse Structure. In *Computational Linguistics*.

Generating definite descriptions: Non-incrementality, inference, and data

Claire Gardent, Hélène Manuélian, Kristina Striegnitz, and Marilisa Amoia

1. Introduction

The generation of referring expressions is a central task for systems that automatically generate natural language texts. Indeed, the correct use of natural language referential devices is crucial for generating successful utterances, i.e., utterances that are easily and correctly understood by the hearer, because referring expressions play an important role in linking an utterance to the previous discourse, the non-linguistic situation the utterance is produced in, and the knowledge of speaker and hearer.

One algorithm that has been particularly influential in this field is the incremental algorithm for generating referring expressions presented in (Dale and Reiter 1995). In this paper, we both extend this basic algorithm to deal with more complex, inference based definite descriptions and propose an alternative, non-incremental algorithm which circumvents the shortcomings arising from incrementality. Moreover, we present the results of a corpus study on definite descriptions in French which suggest some research directions that could help both widening the range of expressions that can be generated and improving the form and content of generated definite descriptions.

We start (Section 2) with a brief presentation of Dale and Reiter's (1995) incremental algorithm. In Section 3, we turn to the generation of bridging descriptions. Bridging descriptions are definite descriptions that refer to entities which are new to the discourse, but can be linked through world knowledge to an entity that has been mentioned before. We look at the contextual reasoning necessary for generating such definite descriptions and then integrate it into a variant of Dale and Reiter's basic algorithm. In Section 4, we present results of a corpus study which examines the relations that link bridging descriptions to the context. The results offer valuable insights on what kind of information has to be provided for the generation of bridging descriptions. Section 5 looks at another aspect of Dale and Reiter's algorithm and algorithms derived from it: incrementality.

We identify a number of problems that arise from incrementality and present an alternative, non-incremental approach. Section 6, finally, summarises our conclusions and points out open questions that further research on the generation of referring expressions needs to address.

2. Dale and Reiter's incremental algorithm

The algorithm described in (Dale and Reiter 1995) provides the basis for many of the later approaches to the generation of referring expressions (Horacek 1997; Krahmer and Theune 2001; van Deemter 2002). As we will also build on it, we will now sketch the way it works. The input to this algorithm are

- the *context*: a set C of positive literals associating entities with relations of arbitrary arity such as shown in Figure 1,
- the *target entity*: the entity t which needs to be referred to.

$$\{rabbit(r_1),\ rabbit(r_2),\ rabbit(r_3),\ hat(h_1),\ hat(h_2),\ bathtub(b_1),\ white(r_1),\ black(r_2),\ white(r_3),\ in(r_1, h_1),\ in(r_2, h_2),\ in(r_3,b_1)\}$$

Figure 1. Representation of a context for Dale and Reiter's incremental algorithm.

The task of the algorithm is to find a *distinguishing description* for the target entity, i.e., a subset L of C which uniquely identifies the target. In other words, given the context, L should give so much information about the target entity that it cannot be confused with any other entity mentioned in C. For example, if, given the context of Figure 1, r_2 is the target entity, $\{rabbit(r_2), black(r_2)\}$ would be a solution, as there are no other entities which are rabbits and black in the context. However, if r_1 is the target entity, $L=\{rabbit(r_1), white(r_1)\}$ would not be sufficient, because there is one *distractor*, i.e., one other entity which also fits the description given by L, namely entity r_3. $L=\{rabbit(r_1), white(r_1), in(r_1,h_1), hat(h_1)\}$ would be a solution in this case, because r_1 is the only entity in C, which is a white rabbit and is in a hat. Formally, this can be captured as follows: L is uniquely identifying the target entity t, if $D(t,L,C) = \{t\}$ with $D(t,L,C)$ the set of objects satisfying the description or more formally:

$$D(t,L,C) = \{o \mid \exists \text{ substitution } s \text{ such that } s(t)=o \text{ and } s(L) \subseteq C\}.$$