

Grounded Discourse Representation Theory

Towards a Semantics-Pragmatics Interface for Human-Machine
Collaboration

Von der historisch-philosophischen Fakultät der Universität Stuttgart zur
Erlangung des Doktors der Philosophie (Dr. phil.) genehmigte
Abhandlung

Vorgelegt von

Tillmann Pross

aus Karlsruhe

Hauptberichter: Prof. Dr. Hans Kamp PhD

Mitberichter: Prof. Dr.-Ing. Alois Knoll

Tag der mündlichen Prüfung: 12.10.2009

Institut für maschinelle Sprachverarbeitung

2010

Contents

I	Introduction	5
1	Introduction	6
1.1	What is this dissertation about?	6
1.2	Method and Scope	7
1.2.1	Preliminaries	7
1.2.2	The technical foundations of GDRT	8
1.2.3	Restrictions	8
1.2.4	An example for discursive interaction	10
1.2.5	Architecture	11
1.3	Structure of this thesis	13
II	Theoretical Foundations	14
2	Language, Reality, Mind	15
2.1	Reference	15
2.1.1	Theories of Reference	15
2.1.2	Reference and Explanation	18
2.1.3	Outlook	24
2.2	Models	24
2.2.1	Theories of Models	24
2.2.2	Outlook	29
2.3	Meaning	29
2.3.1	Semantic meaning	29
2.3.2	Pragmatic concepts of meaning	32
2.3.3	Outlook	33
3	Presentation and Representation	35
3.1	Preliminaries	35
3.1.1	Data sources	35
3.1.2	Perspective representations	35
3.2	Objects, Things, Individuals	36
3.2.1	Basic architecture	36
3.2.2	Inward construction of reference	37
3.2.3	Outward identification of reference	42
3.3	Temporal variation	42
3.3.1	Preliminaries to the treatment of time	42
3.3.2	Perception of time	43
3.3.3	Construction of temporal reference markers	44
3.3.4	Identification of temporal reference markers	50
3.4	Outlook	51

4	The Agent Layer	52
4.1	Mind and Action	52
4.1.1	From actions to strategies	53
4.1.2	Mental attitudes	56
4.1.3	Summary and Outlook	62
4.2	The skeleton of active control	62
4.2.1	The main control cycle	63
4.2.2	Plans	65
4.2.3	System states	66
4.2.4	The agent layer as explanatory mirror structure	70
4.2.5	Know-How	71
4.2.6	Summary and Outlook	71
5	On the way to discourse processing	73
5.1	The notion of discourse	73
5.2	Discourse elements	75
5.2.1	Internal perspectives on action	75
5.2.2	Bodily actions	76
5.2.3	Cognitive actions	76
5.2.4	Speech actions	76
5.2.5	External perspectives on action	76
5.2.6	The structure of discourse	77
5.3	Representing time-individuals	77
5.3.1	The traditional approach	78
5.3.2	Temporal entities in GDRT	81
5.3.3	Interactions of explanations and tense	84
5.3.4	Relating utterances, thoughts, representations and actions	86
5.4	The construction of IRS-individuals	88
5.4.1	Representing the progress of time	89
5.4.2	Thing-individuals	90
5.4.3	Time-individuals	91
5.4.4	The <code>update(IRS)</code> ;-function	101
5.5	A foretaste of discourse processing with GDRT	104
5.6	Summary	105
6	Putting Things Together	106
6.1	The SMS: Sensorimotor Structures	106
6.1.1	SMSs and SMS structures	106
6.2	The EPS: External Presentation Structures	110
6.2.1	The multi-purpose function of the EPS	110
6.2.2	Syntax of EPSs	111
6.2.3	The EPS structure	112
6.2.4	Information extraction from the EPS	115
6.2.5	SMS, EPS and the BDI-interpreter	116
6.2.6	The set of atomic actions <i>Actions</i>	120
6.3	The IRS: Internal Representation Structures	121
6.3.1	Syntax of IRSs	121
6.3.2	Semantics of IRSs	123

III	Practical Application	140
7	Application to examples	141
7.1	Preliminaries	141
7.1.1	The limitations of the 'paper-and-pencil' approach	141
7.1.2	Example discourses	141
7.1.3	Examples of semantic-pragmatic concepts	142
7.1.4	The <code>initialize-state</code> procedure	147
7.2	Building a cube bolting	148
7.3	Building a corner cube bolting	153
7.4	Plans involved in the processing of examples	154
7.4.1	The main plan for building a corner cube bolting	154
7.4.2	Initialization and finishing of a discourse	158
7.4.3	Plans involved in the construction of a corner cube bolting	160
7.4.4	Resolution of IRSs, anchors and errors	164
7.4.5	Resolution of errors	169
7.4.6	Feedback	171
7.5	Assistance mode	173
7.6	Collaboration mode	183
7.7	Teaching Mode	197
7.8	Free interaction	218
IV	Conclusion	231
8	Conclusion	232
8.1	Summary	232
8.1.1	Grounded Discourse Representation Theory	232
8.1.2	Overview of the argumentation	232
8.2	Some notes on open questions	233
8.2.1	The relation between DRT and GDRT	233
8.2.2	GDRT and Axiomatization	234
8.2.3	GDRT and other approaches to human-machine interaction	235
8.3	Outlook	235
8.3.1	Future research	235
8.3.2	Closing words	236
V	References	237
VI	Zusammenfassung	245

Part I

Introduction

Chapter 1

Introduction

This section gives a general overview of the topics discussed in this thesis and an outline of the course I will pursue.

1.1 What is this dissertation about?

This thesis aims at developing a formalism for the semantics-pragmatics interface of a robot equipped with speech and vision processing units, two gripper arms and a head with an animated face (figure 1.1).



Figure 1.1: The robot, named Clara.

The acceptance of robotic assistance crucially depends on the possibility of natural interaction between man and machine. This includes in particular the potential to communicate by the means of spoken language, gestures and facial expressions. This study spells out an approach to goal-directed joint interactions between humans and robots. The goal of this thesis is to ground *semantic* theory (of thought and speech) in the *pragmatics* of a robot's sensorimotor capabilities, or, putting it the other way round, to equip a robot's sensorimotor pragmatics with the ability to construct and make use of semantic representations. The framework is a sample setup geared to the work packages of the large-scale european research project JAST (Joint-Action Science and Technology) that are located at the Technical University of Munich.

1.2 Method and Scope

The theory of Grounded Discourse Representation Theory (GDRT) as developed in this thesis aims at a uniform treatment of speech, thought and action, of object recognition, motor control and natural language meaning, of semantics and pragmatics. Consequently, both with respect to method and scope this thesis is caught in the middle - between semantics and pragmatics, between linguistics and computer science, between philosophy and psychology. The following introductory remarks should provide the reader with a first glimpse of the set of problems to which this thesis is devoted and the methods employed to tackle it.

1.2.1 Preliminaries

GDRT: an interdisciplinary approach of the semantics-pragmatics interface

Developing a formalism that enables a robot to naturally engage in joint interaction is a demanding enterprise which combines major problems from areas such as computer science, linguistics, robotics, logics, psychology and philosophy. The theory of GDRT developed in this thesis seeks to unify insights from research in these different areas. While there exist reasonable models of discourse processing in each of these areas, only few attempts exist to integrate the results of the particular lines of investigation into a unified theory. But the crucial problems pertaining to discursive interaction are located at the intersection of different sciences, e.g. a proper computational account of reference involves (at least) linguistics, psychology, philosophy and computer science. I hope that it will become clear in the course of discussion that the treatment of such comprehensive problems can only be solved by an interdisciplinary approach. Not only between different branches of science, but also within certain disciplines this thesis needs to bridge gaps. E.g. for the case of linguistics, this thesis seeks to integrate semantic and pragmatic perspectives on notions such as meaning or reference.

External vs. internal perspective

I tackle the enterprise outlined above from the external third-person *perspective of a designer* in that I seek to develop a theory that satisfies the demands imposed by the processing of real-time interaction with humans from the internal first-person *perspective of a robot*. I announce a switch between these two perspectives to the reader when necessary. The difference between the external designer's perspective and the internal perspective of the robot is in particular important with respect to the different access that we as designers and the target system (the robot) have to the information structures and processes defined with GDRT. From the designer's point of view, the architecture and function of the robot's information structures and processes are completely transparent. This is not necessarily the case from the robot's point of view. In fact, I will argue that there exists good reason that we (as designers) intentionally limit the robot's access to a certain subset of her architecture and the associated information structures and processes.

The claims (not) raised by GDRT

This thesis will probably raise the question whether GDRT should be considered not only a theory of the design of a robot's semantics-pragmatics interface (in the sense of "cognitive engineering") but

also a theory of the *human* semantics-pragmatics interface (in the sense of “cognitive modelling”). It is not my concern with GDRT to raise philosophical, psychological or neurological claims about the nature of the human semantics-pragmatics interface. Instead, I take core ideas from these fields as an inspiration to engineer a reasonable semantics-pragmatics interface for a robot. Thus I claim no more than the following. It is necessary to consider research results - from linguistics, philosophy, psychology, computer science - concerning the human semantics-pragmatics interface when designing a theory that is supposed to enable a robot to execute natural (viz. human-like) interactions with humans for which, so I shall argue with GDRT, the semantics-pragmatics interface plays a central role. GDRT is not to be considered an attempt of modelling the nature of human cognition but an attempt to develop the conditions of possibility that enable a robot to participate in goal-directed interactions with humans. The answer to the question to what extent GDRT is also to be considered a theory of human capacities is not a matter of GDRT itself but depends on whether the research I draw upon in this thesis is considered an accurate picture of human cognition or not.

1.2.2 The technical foundations of GDRT

This work does not start from nowhere in that it would develop a completely new theory of the semantics-pragmatics interface. Quite the contrary, the argumentation is based on existing frameworks that have the status of de-facto standards. The prominent aspect highlighted in this thesis is that the combination of these respective theories leads to a revised and novel picture of the standard conceptions of meaning, reference and model theory. From a technical point of view, this boils down to an embedding of Discourse Representation Theory (DRT, [Kamp and Reyle, 1993, Kamp et al., 2007]) into a system of Computational Tree Logic [Emerson, 1990] and the Procedural Reasoning System [Georgeff and Lansky, 1987]. Combining action theory with natural language semantics allows for the elegant treatment of phenomena which constitute the core concepts of language and action: propositional attitudes, planning and practical reason. There exists a vast number of approaches to those topics in the literature, such as Segmented Discourse Representation Theory (SDRT, [Asher, 1993, Asher and Lascarides, 2003]), the theory of Dialogue Acts [Grosz and Sidner, 1986] or plan-based theories of speech acts [Cohen and Perrault, 1986] but all these approaches are limited in that they consider only one of the mentioned concepts as crucial to the analysis of discursive interaction and in that they try to derive a theory of discourse without an underlying theory of action. My concern with this thesis is to show that an integrative approach can do better. This pertains in particular to Discourse Representation Theory. While there exist attempts to employ DRT in a more abstract way as a ‘language of thought’ (an idea that goes back to [Asher, 1986]), the enterprise of redesigning DRT to apply to non-linguistic cases has never been undertaken seriously. Thus this thesis can also be understood as an attempt to break the ground for other areas of application to DRT without loss of the ability to process the wide range of natural language phenomena DRT has been designed for.

1.2.3 Restrictions

It is necessary to place some reasonable restrictions on the scope and structure of this thesis.

Structure over content

First, the main interest of this thesis is of an architectural nature: how must the flow of information between hardware and software be organized so that complex phenomena can be captured with simple means? One of the major goals is to overcome the unstructured information flow of common architectures caused by the obvious need to have different sources of information in place during all stages of the processing of discursive interaction¹.

Restriction to the semantics-pragmatics interface: Ideal in- and output

Second, my interest is directed *exclusively* towards the semantics-pragmatics interface. Thus the scope of investigation pursued in this thesis is necessarily restricted to the analysis of what happens between logical forms and the robot's sensorimotor instruments. I assume that sensorimotor in- and output engines work in a reasonable way and that interfaces can be defined to access them. This includes modules such as the natural language parser or object recognition device (see definition 1 below). This does not mean, of course, that I am assuming that parsing or object recognition are necessarily perfect. Incomplete parses and failures in object recognition are not excluded.

Practicability over richness of theoretical detail

The primary intent of what follows is to *design* a *practicable* formalism that enables a robot to conceptualize an unknown and dynamic environment in a way that enables her to solve tasks in joint cooperation with humans. This involves natural language communication, manipulations of the robot's real world environment, the construction and anchoring of complex representations as well as reasoning processes. That is, I seek to develop a cognitive architecture for a robot that spells out the *conditions of the possibility* that underly a robot's ability to participate in an interaction with humans. With regard to this paper's enterprise, the term 'practicable' is to be understood as pertaining at least to the following aspects:

Note 1 *Demands on the formalism*

- *Transparency: The formalism should allow to express explicit statements concerning the thoughts and actions of an agent involved in discursive interaction via a symbolic representation and the application of ideas from formal logic.*
- *Realism: The formalism should avoid counterintuitive analysis and capture the limited nature of human-like abilities and behavioral structures (i.e. it is not concerned with omniscient and omnipotent creatures with unlimited storage and processing power). This applies in particular to the use of knowledge and models: these notions need to be derived from scratch, as we cannot assume the robot to possess any intuitive form of 'world knowledge' besides the knowledge which is explicitly defined.*
- *Dynamics: The formalism should contain reactive as well as active components which can be thought of as algorithms or programs triggered by contextual demands, i.e. the formalism should allow for autonomous and context-dependent interactions.*

¹For some of such 'spaghetti' architectures, see e.g. the papers collected in [Huhns and Singh, 1998, ch. 3]

1.2.4 An example for discursive interaction

This thesis is tailored to the needs of multimodal goal-directed interaction between humans and robots of which an example is given in the sample dialog in example 1. The interaction is about building a *baufix*² 'corner cube bolting', consisting of a cube bolted to the corner hole of a slat.

Example 1 *Example dialog for free interaction.*

1		<i>Initial Setting: On table are a cube, a blue and a red slat and a screw. The user (A) approaches the table and stops in front of the robot (B). The robot opens her eyes and looks at the user. The user does not know the robot's name.</i>
2	A:	Hi, my name is Tillmann. What is your name?
3	B:	My name is Clara.
4	A:	OK.
5	A:	I am going to build a corner cube bolting.
6	B:	OK.
7	A:	Do you know how to do that?
8	B:	Yes.
9	A:	Please take the slat and I screw it to the cube.
10	B:	Which slat?
11	A:	This one.
12		<i>A points to the red slat.</i>
13	B:	OK.
14		<i>B grasps the red slat.</i>
15	B:	OK.
16		<i>A grasps the cube and the screw.</i>
17	A:	Please hold the slat between the cube and the screw.
18	B:	OK.
19		<i>B holds the slat between the cube and the screw.</i>
20		<i>A screws the slat into the cube.</i>
21	B:	OK.
22	A:	That's it. Thank you.

²Baufix is a construction kit for children, consisting of wooden cubes, screws, nuts, slats with holes and other parts.

Even though example 1 is comparatively simple, it exhibits the set of problems which this thesis is intended to develop a treatment for. The next few paragraphs provide the reader a first taste of what problems we must face in the further course.

Begin with the initial setting described in (1). The transformation of a visual perception of the scenery described by (1) into a *semantic representation* (in the sense of a 'mental picture') of (1) does not pose any problems to humans - but it does so for robots. Robots are not per se equipped with the ability to construct semantic representations from *visual* input. Instead, mechanisms must be developed that specify in detail the conditions that render possible to obtain semantic representations from visual input. The other way round, it must be precisely stated *how* semantic representations refer to perceived states of affairs. This task is complicated by the fact that we ('designers') can not foresee the situations a robot will perceive. Consequently, we can not rely on predefined and static approaches to the construction, maintenance and use of semantic representations but must spell out how semantic representations can be constructed, used and maintained by a robot *dynamically* in *realtime*.

The second set of problems comes into play once we consider the contribution of *utterances* to interactions between humans and robots. As with visual information, robots do not come with the pre-installed ability to transform utterances into semantic representations. Instead, mechanisms for speech recognition, parsing and semantic construction as well as utterance generation must be explicitly provided to the robot (but this is not the focus of my thesis, see section 1.2.3). Given (5), we have to take into account that A's utterance not only contributes to the semantic representations of B that B has constructed alongside the progress of discourse but also that A's utterance expresses a certain *intention* of A that substantially contributes to the interaction between A and B. Because 'intention' is a term of *action theory*, besides a semantic theory of representation, we must develop action-theoretic (i.e. pragmatic) means to capture the specific contribution of intentional agency to interactions as in example 1.

Finally, when we move to (10)-(13), we see that both sets of problems introduced above are inextricably tied together. Visual and linguistic information may influence semantic representations and pragmatic intentions of the participants and the other way round - and all this happens in a parallel manner to the other actions (such as grasping or holding a cube) that are executed to realize the goal of the discourse. A precise modelling of how semantics and pragmatics influence each other in the framework of goal-directed interaction is the main goal of this thesis. In addition, the demand of this thesis is not only to describe what is going on during an interaction between a human and robot as in example 1 but to spell out a *control architecture* that enables a robot to execute reasonable continuations of a given interaction or to initiate the realization of her own goals. The next part II of this thesis tackles the introduced problem sets with the development of a formalism that enables me to give a detailed analysis of example 1 in section 7.8 in part III of this work.

1.2.5 Architecture

The formal structure of GDRT developed in this thesis involves a combination of three layers each of which has distinct duties and functions. The proposed division into layers has several reasons motivated by practical and theoretical assumptions. The simpler and I think intuitively evident reason is of a purely practical nature. With regard to such a complex topic, it is useful - for us as designers - to work with a limited number of clearly defined larger functional units. This avoids the fragmentation of the analysis into a confusing structure that unnecessarily complicates argumentation. Second, each of the proposed

layers has a distinct functional domain. The robotic hardware layer should not have to be concerned with the details of natural language semantics, while the structure of speech actions is independent from topics of object recognition. The various interdependencies that nevertheless exist will be treated by the definition of interfaces between the different layers.

It should be kept in mind, that the architecture as a whole is only accessible from a designer's point of view, but the robot herself can only access her architecture through the instruments of representation that are developed with GDRT at the level of the mental layer. This limited access of the robot to her different layers of internal structure provides an intuitive modelling of common approaches to human nature from philosophy and psychology. The architecture of three interacting layers proposed below can be considered as representing body (robot), mind (representation) and control (the intermediate agent layer).

Definition 1 *The layer architecture*

- **a mental layer** covering mental functions for the construction, grounding and manipulation of representations including
 - acquaintance with individuals and temporal variation,
 - representation of thoughts, percepts and utterances
 - handling of high-level action specifications
- **an intermediate agent layer**, managing the flow of information between information stored at the hardware level and its representation(s). The 'runtime environment' for a robot's conscious interaction with reality is concerned with the following tasks:
 - coordination and planning of interactions
 - timing information flow between representation and reality (timing of perception, thinking and acting)
 - presentation of perceptual data
 - derivation of means-end structures and future possibilities of acting
 - evaluation of representations against the sensorimotor layer
 - error and feedback handling
 - parsing and generation of natural language
- **a sensorimotor hardware layer** dealing with the agent's external reality. The discussion of this layer in the framework of this thesis is restricted to the definition of an interface that allows for data exchange with the sensorimotor engine (see section 1.2.3). For most of the tasks listed above, reasonable solutions exist that do not directly affect the topics of this thesis. If such cross-dependencies occur, the respective problems (e.g. parsing and generation) will be briefly discussed. It is supposed that the layer includes modules that handle
 - sensorimotor data
 - speech and object recognition

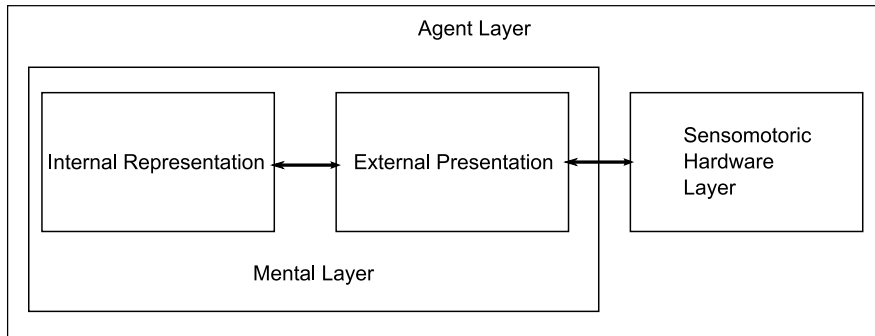


Figure 1.2: Schematic diagram of the architectural backbone of this thesis.

- *pointing and gesture detection,*
- *hand, gaze and (positioning) face tracking*
- *handling of action primitives, motion planning, manipulations and collision detection*
- *attention control*

Figure 1.2.5 schematically pictures the architecture underlying this thesis.

1.3 Structure of this thesis

The thesis is divided into three main parts. Subsequent to this introduction the next part is of a theoretical nature and starts with the discussion of issues related to the architecture proposed with definition 1 and figure 1.2.5. This includes the delineation of concepts of reference, models and meaning (chapters 2 and 3) as well as the design of the control architecture (chapter 4) of the agent layer. Chapter 5 develops a notion of discourse suitable for the goals of this thesis. Then, the formalism of Grounded Discourse Representation Theory is spelled out in one go in chapter 6. I should mention that because the second part is dedicated to theoretical issues it contains only a minimum of examples. The reader interested in applications of GDRT to examples is referred to the third part (chapter 7) which applies the findings of the second part to the analysis of example interactions.

Part II

Theoretical Foundations

Chapter 2

Language, Reality, Mind

The purpose of this chapter is to outline the general motivation for the use of the basic terms *reference*, *model* and *meaning* in the context of this thesis, as these notions constitute the theoretical basis for the subsequent argumentation. After sketching common attempts to the analysis of those terms I will discuss the issues crucial to the development of an artificial agent who is able to make meaningful and successful use of language. It should be noted that this section primarily deals with theoretical issues. Hence for the sake of readability I won't discuss examples. I do not think that jumping back and forth between examples and fundamental theoretical issues leads to a clear argumentation. The reader interested in tangible analysis of examples and familiar with the theoretical issues related to reference, models and meaning can skip this section and return to it when necessary.

2.1 Reference

For the intended development of a formalism that allows an agent to dynamically interact with reality the notion of reference plays a central role. Reference provides the means to connect experiences (of reality or oneself) to representations in a meaningful way. Starting from the analysis of linguistic reference in formal semantics and pragmatics, I argue for a general notion of reference in terms of explanation, i.e. the concept of 'reference as explanation' which is introduced in the following is not limited to the case of linguistic reference but also and primarily pertains to the personal acquaintance with reference. Put another way, I argue that the referential use of mental representations is a necessary condition for the referential use of language.

2.1.1 Theories of Reference

Reference can be understood as the correlation between symbolic representations (however they may be realized, mentally or linguistically) and objects of subjective experience (either of the self or of external reality). Thus reference is concerned with the question how the relation between representations, reference and referents is to be determined and what representations and experiences are supposed to be.

Frege's account to reference

Gottlob Frege was the first to give a systematic and formal account [Frege, 1960] of the relation between representations and referents in terms of formal logic. In Frege's theory, the reference of an expression - the referent - is the entity the expression denotes, e.g. names refer to objects, predicates refer to functions and sentences refer to truth or falsity. The reference of sentences can be systematically derived from the referential contributions of the atomic expressions the sentence in question contains (such as nouns or predicates). This is what is known as the principle of compositionality. Unfortunately, things are not that simple, as Frege already noticed, as the relation between sign and referent can be ambiguous. That is, several signs can refer to the same object or the same sign can refer to several objects¹. Frege was most concerned with the first case, i.e. equivalent expressions that refer ("bedeuten") to the same entity but nevertheless mean ("meinen") something different. The difference in meaning can then only be explained if the difference between "the signs corresponds to a difference in the mode of presentation of that which is designated" [Frege, 1993, p.24]. The additional information about the 'mode of presentation' is what Frege calls the sense of an expression. This leads Frege to the following relation between symbols, reference and sense:

"The regular connexion between a sign, its sense, and its reference is of such a kind that to the sign there corresponds a definite sense and to that in turn a definite reference, while to a given reference (an object) there does not belong only a single sign." [Frege, 1993, p. 25]

Offspring theories

As Frege did not exactly specify what sense is supposed to be, several refinements and interpretations evolved from Frege's conception of reference and sense. Three of them are relevant for the argumentation pursued here: Rudolf Carnap's formalization of Fregean sense [Carnap, 1947] within a modal framework, the theory of reference as identification by Peter Strawson and John Searle [Strawson, 1950, 1959, Searle, 1969] and the conception of sense as algorithm by Iannis Moschovakis [Moschovakis, 1994]. Each of these theories sheds light on different parts of Frege's conception of sense and reference:

- Carnap's modal intensionality has, at the latest with Richard Montague's project of Universal Grammar [Montague, 1979], become a core feature of formal semantics
- the Strawson-Searle theory of reference as identification is located at the heart of what is known today as speech act theory
- Moschovakis' algorithmic interpretation of reference has a strong affinity to computer science and programming

In preparation of what is to come later (a combination of ideas from Carnap, Searle and Moschovakis) each of these theories will be succinctly introduced. Before I do so, a note on the terminology used in the following seems to be appropriate. Frege himself did not use sense, reference and referents in a unique way, nor do the theories that develop his ideas further. Thus from now on I will use the term 'reference' for the relation between expressions and entities, 'referent' for the entity an expression refers to, 'reference marker' for the expression that refers to an entity and 'sense' as a specification of how

¹In addition, the relation between sign and referent may be unknown. This point is picked up in section 6.3.2

reference determines the referent for a given reference marker². The term 'meaning' is reserved for later discussion.

Semantic Interpretation of Reference: Intension

The treatment of reference that is spelled out in Carnap's "Meaning and Necessity" [Carnap, 1947] makes use of advances in formal logic that were not available at Frege's time. These concern in particular Tarski's formal conception of truth and the use of set-theoretic models, as well as Carnap's own formal conception of possible world semantics. Basically, Carnap models Fregean reference as a function that assigns expressions their objects of reference, the so-called denotation function. The Fregean sense of a sentence is formalized as the set of worlds in which the corresponding proposition is true. Carnap calls this set of worlds the intension of that sentence. That is, "the intension of a sentence ϕ is the function which assigns to each possible world W the truth value of ϕ in W ." [Moschovakis, 1994, p. 7]. Carnap's concept of intensionality is a formalization of reference that seeks to capture Frege's 'double vision' of reference by means of the denotation function which is assumed to relate expressions and their reference. With respect to cases of underspecification or missing knowledge, the assumption of such a denotation function is problematic insofar as many "problems in the philosophy of logic turn on the question how we *identify* individuals; how they are individuated." [Hayes, 1985, p. 72]. The intrinsic problems connected with the identification of referents are not clarified by the denotation function but assumed to work somehow. These problems are the central point of investigation for the pragmatic interpretation of reference³.

Pragmatic Interpretation of Reference: Identification

"What conditions are necessary for the utterance of an expression to be sufficient to identify for the hearer an object intended by the speaker?" [Searle, 1969, p. 82]. This is the question that leads John Searle through his examination of reference as speech act. In order that language serves its purpose as means of communication, the information contained in an utterance must be extracted from it, i.e. the reference of it must be retrieved. These 'referential objects' intended⁴ by the speaker must be presented in a way that Searle describes as follows:

- (1) "There must exist one and only one object to which the speaker's utterance of the expression applies [...] *and*
- (2) The hearer must be given sufficient means to identify the object from the speaker's utterance of the expression" [Searle, 1969, p. 82]

Searle's notion of identification employs an idea of Strawson, who replaces Frege's conception of reference by the notion of identification, which both yields the object of reference of an expression and specifies the expression's sense. But the concept of identification as proposed above gives rise to new questions. What exactly is a 'sufficient means'? Pointing may be sufficient for existing and accessible objects, but

²The introduction of this terminology is mainly motivated by the further course of this work, where I introduce a three-level architecture, where referential relations exist between each of the levels.

³I exclude other aspects of reference commonly assumed to be of a pragmatic nature, such as presupposition for later discussion. Nevertheless, presuppositions already shine through in the analysis of reference as identification I will discuss below.

⁴The concept of intention is discussed in detail in section 2.1.1

does not help for predicates and for expressions that refer to objects out of perceptual reach. In addition, compound objects (such as an airplane) pose problems with regard to exact pointing. Equally important, the use of descriptions presupposes the ability to utilize predication, thus a theory of identification will also have to address the phenomenon of predication. Furthermore, there are problems with the use of properties to identify things, as “certain properties of a thing serve to distinguish it from other things and to identify it as being the thing that it is, while other properties of the thing are merely properties that it happens to have.” [Hayes, 1985, p. 72]. Finally, Searle’s theory of linguistic reference presupposes the acquaintance with reference in a non-linguistic sense, as speaker and hearer must possess the ability to deal with reference non-linguistically in order to make referential use of language. Consequently reference must not only include the identification of referents (the ‘outward’ direction of reference) but also the construction of reference markers (the ‘inward’ direction of reference), i.e. the personal acquaintance with reference. These points indicate some of the problems related to reference that this study has to deal with. In a first approximation, this means that I am after detailed *specifications of procedures for the identification and construction of reference*. This specification should, if intended to be implemented on a robot’s hardware, satisfy another demand: algorithmic feasibility.

Formal Interpretation of Reference: Algorithm

Iannis Moschovakis’ account to the Fregean notion of sense is to “reduce the notion of sense to that of algorithm ” [Moschovakis, 1994, p. 2]. He understands reference as the algorithmic identification of reference markers, where the triple of reference marker, reference and referent is considered as input, algorithm and value. Algorithmic determination of reference is relevant for the goals of this thesis inasmuch as this conception of reference allows for the formulation of procedures that are easily implementable and naturally connect with accounts of reference in dynamic semantics [Bos and Blackburn, 2010]. In addition, the algorithmic interpretation of reference renders possible a detailed structure of identification and construction procedures without loss of formal precision.

Reference as algorithmic theory of identification and construction

Summing up, the concept of reference that will be developed in the following combines semantic, pragmatic and algorithmic conceptions of reference. Central to this account will be the notion of reference as explanation which will be introduced next.

2.1.2 Reference and Explanation

This section introduces the concept of reference as explanation, one of the major tools of analysis used throughout the following argumentation.

Reference as Explanation

With respect to reference, I will use the term ‘explanation’ in several contexts. First, there are explanations which capture how referents are identified and reference markers are constructed, respectively. Here, a theory of explanation replaces the concept of a denotation function with a more detailed specification how an expression relates to an object and vice versa. Second, such an account of reference as explanation fits within a larger picture of how a robot can relate her own internal states to her perception

of reality and the self. Here, a theory of explanation can specify the referential dimension of thoughts in a way that enables a robot to autonomously acquire and maintain meaningful representations. Third, explanations are employed to discuss reference in the following assigns the concept of explanations an ontological function. Explanations are 'ontology-makers' in that they render possible to derive basic ontological categories and to group entities according to those categories.

Identification and Explanation Reference has a natural connection to explanation which shines through Frege's notion of sense but is most obvious in Searle's conception of reference as identification. Identifying a referent is stating "certain essential and established facts" [Searle, 1969, p. 169] about the entity in question. But what exactly are these essential and established facts supposed to consist of? In the following, I will argue that it is certain types of explanations which generate the facts (and beliefs) that are necessary for the identification of referents.

Construction and Explanation Explanations do not only apply to the identification of referents (connecting a representation to an experience) but also to the construction of referents and reference markers from given experiences. This in fact is the intuitively more fundamental type of reference: in order to make referential use of natural language, there must first be a capacity for the acquisition of referents and reference markers that is not directly related to natural language but to acquaintance with reality and the self in general, i.e. the referential aspect of thoughts. The construction of reference markers as constituents of private thoughts can be seen as the primary goal of explanations, consequently the use of reference in communication relies on the ability to deal with reference on a personal level.

Representation and Presentation

This section delineates the theoretical foundations for the treatment of reference pertaining to objects and their temporal variation. In doing so, I introduce the basic methods involved in the identification of thing referents and temporal referents.

Mind and Experience Explanations were said to apply to the reciprocal referential relation between mind and experience. But what do mind and experience stand for in the context of this thesis? First of all, experiences can be related to internal or external states of affairs. For now, I restrict the notion of experience to the latter⁵, as even the relation between mind and external reality is beset with very tricky philosophical problems. It should be noted that in the following I do not intend to argue for or against the existence of either mental or real entities. Instead, I adopt a pragmatic position, in which the concepts of mind and reality provide a helpful means for the analysis of the processes related to the conscious commerce of a robot with her 'external reality'⁶.

The status of reality With respect to the related philosophical disputes about whether there is a reality or not and if yes, what reality 'really' looks like, I will only note an essential limit to philosophical

⁵This will ease the discussion as experience of the self requires a notion of mind which is not yet present but will only be developed later on in section 2.2.1

⁶The distinction that is made in this thesis between what is inside and what is outside of an agent is no arbitrary dualism but a fundamental feature of the organization of organisms: "The distinction between everything on the inside of a closed boundary and everything in the external world [...] is at the heart of all biological processes" [Dennett, 1991b, p. 174]

insight that we have to face up too. Reality is not circumventable. Human agents cannot perceive reality as it might be beyond the boundaries of their senses, i.e. they cannot step outside the intrinsic limitation of their sensory instruments to access reality in a more 'direct' way. In particular, the neural stream of perception cannot be accessed directly but only in a preprocessed form of conscious presentation. Consequently it is assumed that speaking of a perception only makes sense if it is available to an agent in the form of a presentation - if sensory instruments yield a *presentation* of reality. Such a presentation can then be processed to take the form of a *re-presentation* of reality, capturing the content of consciousness. Re-presentations are means to reproduce what an agent sees herself to be aware of in a certain situation.

Explanations and reality I return to the topic of explanations. Explanations are related to presentation and representation resp. reality and mind as they allow for the extraction of individual entities from a presentation that constitutes the basic elements of the corresponding representation. The next two sections spell out how the explanatory derivation of such representations of reality can be modeled. This requires me to draw a basic distinction concerning the source of complexity of representational entities. With regard to the perceptual configuration of a robot, perceptions of reality are distributed along two dimensions, as the presentation of reality involves *sensory and temporal distribution*.

Explaining sensory distributed presentation

Objects and Things The term 'sensory distributed presentation' refers to a single snapshot that contains information from different sources of perception. As there is no duration or dynamics, time plays no role in the content of a single snapshot. Thus the only type of entity that can be extracted from a single snapshot are states of things⁷. I will try to make this clear with a simple example.

Example 2 *Imagine a photo taken during a penalty kick in a soccer game. Of course it is possible to identify the ball, the players, the goal and other things by means of the visual relations in which they stand and the visual properties they have. But it is impossible to say something about the action that is going on without the use of background information, e.g. what soccer and a penalty kick is about.*

In a first approximation the construction of individuals from a given percept could be modeled in a simple-minded way by assuming that a thing is identified by a set of properties it has. Such a notion of identification of things may be sufficient for the construction of a simple object recognition but unfortunately it is too simplified for the purposes of this thesis. Explaining the identification of things in such a way is too crude if one takes into account that we also want to deal with compound objects and with objects that exhibit temporal variation. In such cases the identification of an object requires the use of additional, more complex kinds of information than mere bundles of perceptual properties.

Explanations in the proper sense This is where explanations in the proper sense come into play: explanations allow for the systematic assumption of information by means of the introduction of higher-order entities - individuals - for which complex and unique identification specifications can be given

⁷Even if this notion of states of things seems to be intuitively evident, it comes along with a problem that can be formulated as follows: what is a thing if not a set of states it is in? And what is a state if not a thing's having some property? And if change comes into play, is it the thing that changes, its properties or its states? I do not have a conclusive answer to this question but will address the problem in section 3.3.3 where I introduce a concept called thing-individual which allows for an abstraction from the specific constitution of states, properties and things and their temporal variation. That is, the type of formalization to come leaves it open whether temporal variation should be interpreted as a change of things, states or properties.

that go beyond collections of characteristic visual properties. In the case of objects exhibiting temporal variation, the use of relational properties does not suffice but functional information has to be employed to identify the object. The notion of functional explanation leads to the second type of perceptual distribution mentioned above, viz. temporal variation⁸. That is, I am after the individuals in the sense of spatiotemporal continuants as they manifest themselves in series of successive snapshots.

Explaining temporally distributed presentation

Temporally distributed experience is a quite complex topic. The basic material that is at hand is a set of snapshots. As we have no sensory organs for time, the experience of events involves “second seeing” [Dretske, 1969], cognitive vision assisted by explanations of temporal variation. Put another way, processes can’t be perceived in a direct way because a “process isn’t a sequence of events *which* stand in certain causal relations to one another. It is their *standing* in these relations to one another.” [Dretske, 1988, p. 35]. This section argues that it is the various types of explanations that make up this ‘standing-in relation’ of temporal variation. This holds in particular for the prediction of future outcomes of temporal processes which will be discussed in section 3.3.3.

The individuation of temporal entities The explanatory analysis of the individuation of events proposed in the following is derived in an unusual way, as it doesn’t start from a definition of events but with an examination of agency. I take agency as the fundamental concept involved in the cognitive processing of temporal variation because agency captures both the objective and subjective dimension of temporal variation. The objective dimension of agency is about *how* properties of things change, its subjective sense about *why* those properties changed. Both these dimensions can be related to the notion of cause. Agency is the ability to exert power, to be an actor or a cause that is able to bring about effects, and thus to induce temporal variation. The question of interest to my enterprise is why and how agency is ascribed to an object’s temporal variation with respect to the interdependent temporal variations of several objects. In other words: how can one find an explanation for the (more or less complex) temporal variation of a certain object’s properties by linking this variation to another object’s temporal variation (‘external cause’) or forces inherent to the object itself (‘internal cause’)⁹? Consequently, I have to examine which types of explanations of temporal variation are available.

Deductive-nomological explanation A person taking the perspective of the natural sciences will maybe try to utilize the Hempel-Oppenheim scheme (resp. the deductive-nomological model) to explain the link between cause and effect. An explanation in the form of this scheme is pictured in definition 2.

Definition 2 *The Hempel-Oppenheim scheme [von Wright, 1971, see p. 11]*

<i>Premises:</i>	<i>singular facts</i>	A_1, A_2, \dots, A_n
	<i>universal laws</i>	G_1, G_2, \dots, G_n
<i>Conclusion:</i>	<i>The event to be explained</i> E	

⁸This claim can also be formulated the other way round: most objects can only be identified if they do exhibit temporal variation [Marr, 1982]

⁹[Dretske, 1988] makes a more fine-grained distinction that will be adopted next. He differentiates two types of internal causation and one of external causation.

This kind of explanation may do its job in the domain of e.g. physics, where

- universal laws can be explicitly stated (and are available)
- the singular facts capturing the antecedent conditions can be isolated and controlled in an experimental setting.

E.g. a ball's rolling off the table can be explained in terms of the ball's shape, the table's slope and the laws of gravitation. But the type of agency I am concerned with is not limited to the domain of physics, it also involves mental activities and their realizations. Explaining such kinds of agency in terms of the Hempel-Oppenheim scheme poses at least two severe problems:

- the involvement of non-deterministic and non-causal temporal variation for which it is hard to formulate universal laws and
- the detection of the significant singular facts to which the laws can be applied. The reason for this second problem is inherent to mentality: mental activities are not perceivable directly, but they are hidden 'inside' the physical manifestation of the agent (the so-called 'problem of the third person').

Representation and Behavioral Reductionism Nevertheless there have been numerous attempts to develop theories that circumvent the 'problem of the third person' by reducing mental states to physical causes which should render them amenable to the application of the Hempel-Oppenheim scheme. Unfortunately, such a way of explaining an agent's behavior does not fit our needs for a transparent and symbolic representation of the agent's mental states, as we need to make the agent's internal states visible and accessible to mental functions. Assuming a transparent representation of an agent's mental processes does not entail a commitment to the real existence of mental representations, but I need a 'toolbox' that permits the explicit symbolic formulation of processes of interaction between an entity and its environment¹⁰.

Intentional explanation If the Hempel-Oppenheim scheme does not fit my need for the explanation of mental agency, what other ways do exist? There is a long tradition dating back to Aristotle for the use of the so-called practical syllogism, an explanation scheme for intentional, i.e. conscious and deliberate human acting [Anscombe, 1957]. A formulation of the practical syllogism is given in definition 3.

Definition 3 *The practical syllogism [von Wright, 1971, see p. 96]*

Let A be an agent, p a state of affairs and a an action. Then As acting can be explained as follows.

A intends to bring about p.

A considers that he cannot bring about p unless he does a.

Therefore A sets himself to do a

While the practical syllogism is intuitively appealing and seems to state the obvious, it involves a bunch of philosophical subtleties that need to be mentioned:

¹⁰The assumption of mental representations is not incompatible with the neurophysiological or subsymbolic basis of cognition. In fact the representations I develop in the subsequent sections will be grounded in the sensorimotor mechanisms underlying cognition.

- The practical syllogism is an abductive inference scheme, i.e. it is not a deductive reasoning pattern, but merely a form of 'educated guessing'.
- The practical syllogism involves the concept of intentionality which is roughly to be seen as the directedness of mental states toward reality. This topic is elaborated in extenso in section 2.2.1.
- The practical syllogism requires practical knowledge about means-end relations that derive from practical experiences - a topic that is hard to capture in terms of a formalism.

However, the practical syllogism is a powerful scheme that not only allows for the explanation and prediction of the behavior of other human agents, but also of self-awareness and other mental functions. For now, I will be content with just adopting the practical syllogism as an explanatory schema, reserving the discussion of its detailed structure and formalization to a later time.

Behaviorism Causal and intentional explanations still fail to cover all kinds of agency we should be able to explain, as there is one kind of temporal variation that is neither causally determined nor guided by intentions. This temporal variation is behavior in the literal sense and can be explained as behavior triggered by physical needs, where there is some kind of rudimentary planning of behavior which is too complex to reduce it to purely physical causes but at the same time too simple for the assumption of intentions. Unfortunately, I can give no precise definition of behavior yet, as we don't have any formal apparatus at hand. So I must leave the reader with the promise the formal definition of behavior to be given later on. A definition in terms of desires will be proposed in section 4.1.2.

Ascription of agency The actual application of explanations is tricky. Spelling out conditions for the use of the three different types of explaining a thing's temporal variation is a task I can't accomplish by following formal principles. In principle any temporal variation can be explained with any of the three types of explanation. Whether such explanatory ascriptions of agency are really true or adequate, depends merely on cultural or religious conventions rather than on purely scientific matters. E.g. a relic bone will be explained as agentic, if the effect of a person's healing is traced back to the impact of the bone. Determining which objects possess which kind of agency is part of a person's fundamental categorization and conceptualization of the world. For our enterprise this means that we should rather focus on the development of precise definitions and conditions for the use of explanations (in the sense of 'conditions of the possibility') than on predefined thing-agency-explanation mappings.

Explanations in context The trichotomy of explanation spelled out above is no arbitrary classification, but finds strong backing by philosophical theories surrounding the notion of 'folk psychology' [Dennett, 1991a]. It corresponds most notably to Dennett's theory of intentional stance [Dennett, 1989]¹¹. Following Dennett, explanation and prediction of an object's behavior start off at varying levels of abstraction: physical, design and intentional stance. Each type of explanation comes with its own terminology:

- the physical stance is concerned with physical or chemical terms such as mass, energy, velocity,

¹¹Which is in turn inspired by David Marr's theory of three levels of analysis (implementation, algorithmic and computational level, cf. [Marr, 1982])

- the design stance is about biology and engineering: purpose, function and design.
- the intentional stance is located at the level of minds: belief, thinking and intentionality.

Similar three-fold categorizations of basic distinctions in explaining the world's temporal variation can be found in various approaches such as Dretske's theory of three system types [Dretske, 1988], or the three basic types of empirically learned distinctions in the elaboration of the Marburg school of Erlangen constructivism [Hartmann and Janich, 1991]. Table 2.1 sketches the introduced types of explanation of temporal variation. These types of explanation will be employed to subsume the temporal variation of a thing's properties by the notion of a temporal referent, an abbreviation for the complex functional property underlying the special temporal individuality of a thing.

Explanation	Causal	Behavioral	Intentional
Stance	Physical	Design	Intentional
Scheme	Ded.-Nom.	Mixed	Practical Syllogism
Telos	-	Desire	Intention
Type	Law	Mixed	Convention
Prototypical Agent	Thing	Animal/Plant	Human

Table 2.1: The types of agency which are used in this thesis.

2.1.3 Outlook

The discussion of the relation between reference and explanations has brought us one step further in the direction of establishing mechanisms for the identification of referents and the construction of reference markers. It is the type of explanation employed in the individuation of entities that provides the information for the proper identification of the referent.

The course of further work suggested by this analysis of reference requires that explanations are used in a more detailed and formal way. This will be done in chapter 3, employing the formal methods of Carnap to model the Strawson-Searle idea of reference as identification, and in a way that will enable me to arrive at an algorithmic formulation of referent identification and construction in chapter 6.

2.2 Models

The notion of model is of central importance to the tasks of this thesis. This section elaborates on the different aims and domains of application of models as well as the basic theoretical motivations underlying the use of different types of models.

2.2.1 Theories of Models

With respect to the challenge of this study, three concepts of models will be introduced in the following:

- **Formal set-theoretic models** Logic-based theories of meaning define models as abstract set-theoretic structures that allow for the definition of truth and reference for the sentences and terms of a given language.

- **Mental models** In cognitive science models are understood as the manifestations of the fundamental human ability of constructing, maintaining and reasoning with representations.
- **Models of the mental** It is common in the philosophy of mind to take intentions as the central feature in modeling the conscious human commerce with reality.

In this section, I want to contrast these three notions of models by showing that they differ in their application domain as well as in their aims. That is, each type of model serves a different purpose. Formal models are concerned with structural issues, cognitive models focus on utility while models of the mental deal with the content and structure of consciousness. Close examination of these different notions of models is important for me, inasmuch as I won't be able to rely on just one of them to the exclusion of others - I can't rely on an agent possessing the ability to have models of one kind or another but must explicitly state what models she should acquire in what way under which circumstances for which purpose. By doing that I will prepare the ground for a notion of meaning that incorporates the key features of set theory, cognitive science and philosophy of mind. Spelling out this in a way that matches the desiderata of this thesis as they are spelled out in definition 1 is one of the major tasks of the present work.

Formal Models

Tarski Models The use of the term 'model' in formal logic is mainly due to Alfred Tarski, who introduced models as part of a formal solution to the problem of giving a definition of truth [Tarski, 1956]. Tarski, disenchanted with the ambiguity of natural language, proposed that truth is consistently definable only for disambiguated languages - formal languages such as predicate logic. For such languages, it is possible to establish a recursive definition of truth that evaluates given formulas against a model by checking whether the respective formula is satisfiable - whether there is an assignment that makes the formula true in the given model¹². Models that are adequate to this task basically mirror the symbolic structure of the language in terms of set-theoretic constructs built from individuals, functions and relations. Consequently, a formal model for a formal language L is a structure M (in this example first-order predicate logic) that consists of the following items:

Definition 4 *Structure of a formal model*

A formal model M consists of the following components

- (1) *A set called the domain of M (the universe) and written $dom(M)$; it is usually assumed to be nonempty;*
- (2) *for each individual constant c in M , an element c_M of $dom(M)$;*
- (3) *for each predicate symbol P of arity n , an n -ary relation P_M on $dom(M)$;*
- (4) *for each function symbol F of arity n , an n -ary function F_M from $dom(M)$ to $dom(M)$.*

Models of the above type are purely structural, i.e. they are concerned with formal properties of abstract structures. This applies most directly to the human commerce with reality - Tarski models are not

¹²In this case, the formula is the object language, for which truth is defined in the metalanguage of satisfaction in a model.

designed to capture the cognitive nature of reality but serve formal purposes such as the definition of satisfiability, whereas the connection to natural language semantics is not immediate. Nevertheless the above conception of model theory is the basis of formal natural language semantics. This was probably not Tarski's intention, as he did not aim at a theory of natural language meaning but looked for a "materially adequate and formally correct definition of the term 'true sentence' " [Tarski, 1956, p. 152]. Only in connection with the philosophical concept of truth-conditional semantics can this produce a theory of formal natural language semantics¹³.

The dynamics of reality and Tarski models Since the days of Tarski, formal model theory has not undergone major revisions; the ontology has been upgraded to match some of the needs of natural language analysis, for instance that the universe can include not only individuals but also eventualities [Davidson, 2001a] and the interpretation functions have been indexed to capture modal configurations (Carnap [1947], Kripke [1980]). That the cognitive importance of models has not been in the focus of researchers in formal semantics and artificial intelligence may be due to the fact that at least in the early days of formal natural language semantics, any reference to the cognitive or psychological dimension of natural language semantics was suppressed. But the main problem with today's model theory from the perspective of this thesis is located at another point: while the definitions of object languages and truth-conditions of formal semantics have advanced to fine-grained dynamic formalisms, the corresponding model theory is still static. In the light of linguistic and other actions, the models of formal semantics are constructed offline, i.e. they are defined before the use of natural language as one form of action has had a chance to affect reality. A formal model captures *structural properties of formal languages*, i.e. it captures how expressions of a certain language systematically match structures that have no necessary connection to the mental states of the agent who is processing the respective expression. But the connection between models, mental states and perceptions is a major constraint on a formalism that can serve the goals I have set myself in this dissertation as they were stated in section 1.2.5. Thus the formal structure of the models that will be developed in the subsequent chapters must contain a notion of *model dynamics* - the model's universe and its relations must be allowed to vary over time. In anticipation of what is to be spelled out later, the formal model that I seek to deploy is to be constructed from the information delivered by the sensory instruments of the robot and the agent layer; it *presents* the experience of external and internal reality.

Mental Models

Craik Models The use of models in cognitive science serves other interests than the clarification of issues related to formal languages and truth - it is about how reality is represented in human thought: "thought models, or parallels reality" [Craik, 1967, p. 57]. Even if the roots of this idea date earlier, this use of models was first explicitly spelled out by Kenneth Craik [Craik, 1967]. His ideas on the 'nature of explanation' heavily influenced the contemporary psychological theory of mental models [Johnson-Laird, 1983, 2004]. The theory of mental models is not intended to capture formal structures of mathematics and logic but is designed to capture how *representation of and reasoning about reality* works, that is, its focus is on the usefulness of models in accounts of human commerce with reality. One of Craik's major

¹³As I want to keep the topic of meaning apart from the notion of model, I will delay the discussion of truth-conditional semantics until chapter 2.3.1. For now, I am only concerned with clarifying what the term 'model' can mean.

insights was that explanations are intrinsically dynamic, they are subjective valuations of a mutable situation that have to cope with the non-determinism of temporal variation and with the coincidence of perceptual data. Nevertheless, humans are able to deal with the dynamics of reality in an efficient way. In order to explain this fact, it can be assumed that there is a layer of mind that organizes the mental commerce with reality in the form of representations of presentations of reality. Craik called these 'models', but I will use throughout this thesis the term representation for such 'models' in order to avoid confusion with formal models. With the help of such representations, the referential interplay between mind and reality can be processed in a way which permits abstractions from the vast majority of actual variations in time. The mental representation of presentations of "external reality" [Craik, 1967, p. 61] is part of what I introduced under the term mental layer in figure 1.2.5¹⁴. Internal mental representations are supposed to have "a similar relation-structure to that of the process it imitates." [Craik, 1967, p. 51]; they *represent* commerce with reality.¹⁵

The utility of representations Craik proposes the following procedure concerning the use of models: (cf. [Craik, 1967, p. 50])

- (1) "Translation' of external processes into words, numbers and other symbols.
- (2) Arrival at other symbols by a process of 'reasoning', deduction, inference, etc. and
- (3) 'Retranslation' of these symbols into external processes (as in building a bridge to a design) or at least recognition of the correspondence between these symbols and external events (as in realizing that a prediction is fulfilled)."

This procedure exhibits a quite interesting feature of Craik's model theory - his models are not only supposed to represent reality but also allow for reasoning about reality, which comes close to what I introduced under the term 'reference as explanation' above:

- (1) the construction of a model from perceptions - "Perceptions yield models of the world that lies outside us." [Johnson-Laird, 2005, p. 185],
- (2) the derivation of representations with the help of explanations
- (3) the anchoring of representations in reality

A mental representation takes the presentations of sensory instruments (in our case the object recognition and other sensory devices of the robot) and transforms them into "a convenient small-scale model" [Craik, 1967, p. 59], as the limited capacity of the human mind does not allow for infinite representations. Thus in building and anchoring of a representation only the most salient features of the current situation can be included.

¹⁴Remember that we are still concerned with the experience of reality and not the self

¹⁵Discourse Representation Theory has a quite similar objective, even if the focus is directed toward the analysis of natural language phenomena. If applied to a natural language discourse, the discourse representation structures proposed by DRT yield a partial model of the discourse that expresses the meaning of the discourse. One of the main goals of this thesis is to formulate a Craikian reinterpretation of DRT that widens the scope of DRT to capture not only the commerce with language but also with reality.

Models of the mental

I should briefly mention that besides formal and mental models there is a third way of understanding the notion 'model' that I must discuss; that of an agent's model of her own internal states. Such models of the conscious states of the self play a central role in the design of the agent layer. For the purposes of this thesis, two concepts are of central importance that should not be confused: that of intentions and that of intentionality.

Intentions Intentions play a central role in action theory where intentions are a distinguished way of rationally purposing the consequences of one's actions. Thus intentions are closely connected to reasonably structured sequences of actions, a feature that I will make extensive use of in section 4.1.1. Another feature of intentions that should be mentioned here is that they exhibit the experience of the difference between mind and reality. To illustrate this, let us have a look at a more detailed version of the practical syllogism (definition 5):

Definition 5 *The syllogism of intentional acts (cf. [Hubig, 2002, p. 18])*

Let

- x be the actor,
- P be a subjective, (imagined as being possible to realize) means
- Q' be a subjective, (imagined as being possible to realize) goal
- Q'' be the actually realized goal
- M be an outer, real existing means

Then the general form of an act is:

$$\begin{array}{c} x \text{ intends that } Q' \text{ via } P \\ \quad \quad \quad \underline{P \text{ via } M}^{16} \\ x \text{ doing } M \text{ brings about } Q'' \end{array}$$

I mentioned the uncertainty related to intentional explanations in section 2.1.2. Considering the above definition 5, this uncertainty can now be traced back to the difference between what the agent supposes reality to be like in her imagination and what reality actually looks like. In definition 5, this gap is marked by the difference between the intended goal Q' and the real state of affairs Q'' ¹⁷.

Intentionality Intentionality refers to the aboutness or directedness of an agent's mental representations, i.e. "[i]ntentionality is by definition that feature of certain mental states by which they are directed at or about objects and states of affairs in the world" [Searle, 1980]. The aboutness of mental representations will be captured by means of anchoring in section 3.2.2.

¹⁶This line constitutes the crucial point of the explanation scheme for intentional acts. The transition from P (thought) to M (reality), displayed as *via*, involves the choice of the right real means M to realize the subjective means P and plays the role of a link between intention and reality.

¹⁷In fact, the realized goal Q'' and the intended goal Q' will always differ, as the intended goal has only finitely many properties, while the real goal has infinitely many properties.

Summary The concepts of intentions and intentionality apply to two different things: intentions are related to actions while intentionality is concerned with mental representations. Intentions will be discussed in more detail in section 4.1.2, while intentionality will be handled by means of anchoring of representations (section 3.2.2).

2.2.2 Outlook

The model theory that will be developed in chapter 6 incorporates the core features of all three types of models discussed above. The central idea that will be employed to that effect is based on an integration of the three types of models discussed here into the overall architecture as pictured in figure 1.2.5 on page 13.

2.3 Meaning

Basically, the theories of meaning that are relevant to this thesis are rooted in two philosophical traditions that are very different both in their focus and methods: philosophy of 'ideal language' and philosophy of 'ordinary language'. The concepts of meaning pursued in these two approaches can roughly be equated with semantic meaning (focusing on truth) and pragmatic meaning (focusing on success). I will examine these two concepts in what I take to be the standard presentations before I develop my own concept on the basis of the notions of reference and models I have introduced above. As a preliminary, I have to say something about meaning in general. Meaning is not only associated with languages as objective abstract structures but also has a personal and private dimension: the fundamental ability of directing internal states toward external situations. I will deal with this dimension of meaning via the concept of anchoring (see the coming section 3.2.2).

2.3.1 Semantic meaning

I have already introduced Frege's principle of compositionality, Tarski's ideas on models and truth and Carnap's theory of intension. What is still missing for a theory of natural language semantics is how to relate meaning to these formal concepts.

Static semantics

Basically, semantic theories of meaning can be grouped according to the way and order in which they developed.

Truth-Conditional Semantics While the cognitive dimension of human commerce with internal and external reality is ignored in Tarskian model theory, the use of such models in order to explicate natural language meaning essentially involves a conception of reality. The connection rests on an idea that still constitutes the philosophical backbone of today's semantic analysis, an idea that is put by Ludwig Wittgenstein as follows: "To understand a proposition means to know what is the case if it is true." [Wittgenstein, 1922, 4.024]. This claim states the philosophical essence of what is today known as 'truth-conditional semantics', whose central idea is that the meaning of a sentence consists in the conditions under which the sentence is true. In other words, truth-conditional meaning is about checking whether

the proposition in question depicts reality in a congruent way; “In order to tell whether a picture is true or false we must compare it with reality” [Wittgenstein, 1922, 2.223]. Thus meaning and reality are inseparably tied together. Formal Semantics has been heavily influenced by that philosophical conception of truth-conditional meaning. One prominent example of this idea is e.g. the view spelled out in Donald Davidson’s “Truth and Meaning” [Davidson, 2001b], where he argued for the definition of a truth-predicate that provides a “clear and testable criterion of an adequate semantics for natural language.” [Davidson, 2001b, p. 35]. Davidson’s idea of truth-conditional semantics is fundamentally connected with Tarski’s formal definition of truth¹⁸; it is a reformulation of Tarski’s convention T for translations of natural language expressions into logical formulas. Tarski proposed to define the truth-predicate for an object language in a meta-language, e.g. set-theory. Applied to natural language semantics, this requires to map a given natural language expression E to its logical form L (its truth conditions) which can then be evaluated with respect to satisfiability in a formal model M as given in definition 4.

Montague Grammar The probably most influential manifestation of truth-conditional semantics is due to Richard Montague, who was the first to undertake the enterprise of systematically deriving logical forms of utterances for non-trivial fragments of natural language. In a series of papers, he argued that there is “no important theoretical difference between natural language and the artificial languages of logicians” [Montague, 1979, p. 1] and supported this by spelling out in detail how natural language sentences can be mapped to higher-order intensional logic - a milestone for formal natural language semantics.

Dynamic Semantics

The central role of context While Montague’s agenda is still the principal foundation of research in formal semantics it became clear in the early 1980s that Montague’s approach to formal semantics has drawbacks when it comes to sequences of utterances with cross-sentential dependencies such as anaphora. Montague treated pronouns as variables and pronominal anaphora as involving quantificational binding. This entails that if a pronoun in a sentence S_i is anaphoric to an antecedent in an earlier sentence S_j , then S_j and S_i must be analyzed as involving the same variable occurring both in the position of the pronoun and in that of its antecedent; and this variable cannot be bound until the later sentence S_i has been revealed, with the effect that no interpretation of the earlier sentence S_j is possible until that later point in the discourse. Such an analysis of anaphora seems quite unrealistic if one assumes that human agents process utterances online and incrementally, i.e. if the construction of discourse representations proceeds in parallel to the binding of referents. This insight was one of the main motivations for the development of Discourse Representation Theory (DRT, one of the possible formalizations of dynamic semantics [Kamp et al., 2007]), where anaphoric binding to an antecedent is compatible with a complete interpretation of the earlier sentence before the later sentence with the pronoun is realized. In this way, discourse interpretation becomes genuinely incremental, in a way which resembles much more closely the way discourse is processed by human speakers. This incrementality of interpretation takes the following form: the interpretation assigned to the part of the discourse that has been interpreted so

¹⁸Consequently this kind of semantics has to face all the problems that have been mentioned in the last section on models - classical truth-conditional semantics relies on ‘offline’ models that do not capture the cognitive dimension of commerce with reality but provide a virtual formal structure for formal languages.

far serves as 'discourse context' for the interpretation of the next sentence. The context-dependency of the interpretation of utterances is one part of the dynamics of meaning handled in dynamic semantics, the other part has to do with the way context itself is influenced by the interpretation of an utterance. The processing of each new utterance updates the context with the information it contributes. This updated discourse context is then to be used as background for the interpretation of further utterances, i.e. discourse context itself has a dynamic nature.

Context and information It should have become clear in the last paragraph that one of the key notions of dynamic semantics is that of a discourse context. One possible way to capture the notion of discourse context is to identify it with an information state. The concept of an information state extends the 'propositional' approach to formal semantics in that it does not only capture the set of possible worlds W in which a Discourse Representation Structure (DRS) K is true but also records the verifying embeddings f that make K true in a world $w \in W$. This idea can then be employed for a definition of the meaning of an utterance as its potential to change representations of discourse contexts resp. information states.

Truth and dynamics The revised conception of meaning in dynamic semantics as interpretation in context and context update requires an improvement of the interpretation process of possibly partial representations. As discourse representations may change after each processing step, the semantics of such dynamic discourse representations can not rely on static assignments of reference markers to entities of the universe. Instead, after each update of the representation the current stage of representation construction must be evaluated by an updated verifying embedding of the representation in the given model structure. This incremental procedure of interpretation leads to a dynamic notion of truth, where the dynamics of truth results from the requirement that fixing verifying embeddings for the incremental evaluation of DRSs is not only about truth-conditions but requires a successful embedding of the representation of a discourse into the given model structure.

The statics of representational approaches The way DRT handles the dynamics of the interpretation process through the stepwise construction of DRSs has been criticized as capturing only part of what the dynamics of interpretation is about: the dynamics of DRT-like interpretation processes "resides solely in the incremental build-up of the representations, and not in the interpretation of the representations themselves." [Groenendijk and Stokhof, 1999, p. 10]. This observation has led to the development of Dynamic Predicate Logic (DPL, [Groenendijk and Stokhof, 1991]), an attempt to capture the nature of interpretation dynamics by modeling information states with the help of sets of "model theoretic objects, represented in the metalanguage, not with expressions of the object language" [Groenendijk and Stokhof, 1991, p. 13]. This makes an intermediate representation formalism superfluous. While I agree with Groenendijk and Stokhof's observation that DRT handles the dynamics of interpretation by means of representations of discourse context rather than the dynamics of the context itself, I propose another solution than they do. First, it will become clear during the next chapters that we cannot abandon the use of intermediate representations, as we need a transparent tool to treat the various processes involved in the interplay between language, reality and mind¹⁹. Second, I will argue

¹⁹Chapter 4 gives a detailed discussion on this point

that the dynamic interaction between representations and the world is in the first instance a matter of action-theoretic pragmatics and not of truth-conditional semantics - a claim central to this thesis and one that motivates the strategy of investigation I will pursue in the following chapters. This can also be understood as a plea for a formal concept of dynamic 'utterance context' [Kamp, 2008] as opposed to 'discourse context'.

2.3.2 Pragmatic concepts of meaning

Meaning and Use

While meaning is, by definition, the central concept of semantics, it has also been a prime target within pragmatics. But the pragmatic concept of meaning is not given in a well-defined formal theory, such as Tarski's theory of truth and Tarskian model theory. Rather, the pragmatic conception arose out of opposition to the semantic conception. Central to the early days of pragmatics was the claim that the "true-false fetish" [Austin, 1962, p. 151] of formal semantics is not able to capture what constitutes the meaning of many utterances in the daily use of ordinary language. Pragmatic accounts of meaning focus on what can be *done* with the help of utterances. Utterances are to be analyzed as actions and the primary distinction is that between failure and success. Sometimes, when the utterance is an assertion, its success may consist, wholly or partly, in its being true. But truth is only one among several properties of utterances that are needed to characterize success, and so loses the unique relation in which it stands to meaning in the view of formal semantics. Put another way, the structure of utterances is in some way similar to that of 'normal' non-verbal actions. In the development of my own conception below, I will, mindful of Austin's recommendations, not even make a fundamental distinction between utterances and other kinds of action. This implies that the problems related to the pragmatic conception of meaning are similar to those which arise for the explanation of action in general. Utterances resemble other actions in that in either case the intentions behind them are hidden from direct observation. Thus it is part of the pragmatic conception of meaning that interpretation cannot stop at the surface of overt symbolic form, but must use reasoning mechanisms to get at the intention behind the utterance. This entails that interpretation is, like the explanation of non-verbal actions, subject to an element of 'guessing' that is inherent in all forms of abduction. Consequently, pragmatic meaning is not so much a formal but first and foremost a social phenomenon. This general character of the pragmatic conception of meaning is mirrored by the concepts usually involved in its analysis such as conventions [Lewis, 1983], the observance of ceremony-like guidelines [Austin, 1962], institutional conventions [Searle, 1969], the maxime of cooperation [Grice, 1975], the relevance of utterances [Sperber and Wilson, 1986] or language games [Wittgenstein, 1953].

Formal pragmatics

In view of the 'social' nature of pragmatic concepts it is not surprising that there are not many formalizations of such concepts. Nevertheless, the history of formal pragmatics started around the same time as formal semantics of natural language with the publication of David Lewis' book on conventions [Lewis, 1969]²⁰. Lewis showed how pragmatic phenomena can be analyzed by the methods of game theory. More

²⁰I do not mention the school of Erlangen Constructivism here, as the approach proposed there is not primarily intended as pragmatic theory of utterance meaning.

recently, the application of game theory within pragmatics has experienced a strong revival (see e.g. the collection of papers in [Benz et al., (2005)]). Another line of formal development, inspired by Searle’s work on speech acts is found mainly within artificial intelligence. Here speech acts are analyzed within the setting of multi-agent interactions, using the tools of modal logic (cf. [Cohen and Perrault, 1986, Singh, 1998]).

2.3.3 Outlook

Summing up the discussion of meaning in the previous paragraphs, in the following I outline how we will proceed with the topic of meaning in subsequent chapters. It is a philosophical commonplace that language is no end in itself but serves a purpose; that one can “do things with words” [Austin, 1962] and that words are there in order that things can be done with them. This applies to pragmatic meaning as well as semantic meaning, as utterances code both types of meaning. That is, the “toolbox” [Wittgenstein, 1953] of language allows for the joint coding of information (the propositional content) *and* intentions (the illocutionary force in the Austin-Searle ([Austin, 1962, Searle, 1969]) jargon, or Speaker-Meaning in terms of Gricean analysis [Grice, 1957]). An analysis of natural language meaning must address both these aspects. And it must also show how they can go hand in hand. For the two aspects are interdependent, decoding the information contained in an utterance requires recovering the speaker’s intention and conversely. In order to do justice to this two-sided nature of utterance meaning, it is necessary to explain how utterances are directed at the hearer’s capacity for mental representations of information, the reality that speaker and hearer share and how utterances serve the realization of a speaker’s intentions.

In order to encounter this set of problems, at first it is required to explicate the general role of semantic representations, thus not only taking into account semantic representations that stem from utterances but also semantic representations that pertain to (e.g. visual) perceptions of internal and external reality. Subsequently, I will define two different *stages* of the interpretation of semantic representations, *semantic binding* and *pragmatic profiling*. The term ‘semantic binding’ comprises the assignment (or ‘binding’) of thing reference markers occurring in a semantic representation to model-theoretic entities. The term ‘pragmatic profiling’ is intended to capture the pragmatic dimension of meaning that manifests itself in the temporal profile of a semantic representation defined by the occurring temporal reference markers. The distinction between semantic binding and pragmatic profiling is motivated by the following observation. From a technical point of view, the primary concern of semantic analysis is to determine the treatment of thing reference markers concerning conditions such as argument relations and quantification in close dependency on the syntactic surface of the utterance²¹. With respect to argument relations, this amounts to filling their open slots with the appropriate number and with the right kinds of reference markers. In turn, filling in the argument slots of conditions determines the semantic binding of the representation, a logical skeleton from which pragmatic profiling can take effect in that it determines options of action associated with the then semantically bound temporal reference markers.

Then, in order to approach the intentional dimension of goal-directed discursive interactions, I will consider two basic *modes* of the two-staged interpretation of semantic representations, roughly adopting

²¹Montague grammar is the most obvious example of the close connection between semantic processing and syntactic parsing. In Montague’s formalism, the construction of logical forms goes hand in hand with syntactic parsing.

the distinction that [Searle, 1983] draws between the word-to-world direction of fit between words and the world and the world-to-word direction of fit between the world and words.

An explication of the concepts of semantic binding, pragmatic profiling, plain interpretation and reactive interpretation and their role in the modelling of the semantics-pragmatics *interface* is given in chapter 5, the formalization of these concepts is developed in chapter 6.

Chapter 3

Presentation and Representation

This chapter prepares the ground for the formalization of the concepts related to reference and models introduced in the last chapter. That is, the issues on reference and models I have discussed here are linked with the architecture proposed in chapter 1.

3.1 Preliminaries

3.1.1 Data sources

To avoid confusion it has to be mentioned in advance that this section is part of the development of a theory about how to enable a robot to interact with humans in the general sense outlined in the introduction (section 1.2.2). Consequently, I develop the analysis of spoken interaction starting from a general theory of action without singling out speech actions as a distinct class of actions. For the argumentation in this chapter that means that I won't make use of examples involving natural language. Relying on the distinction between speech actions and other actions from the start would have been an impediment to developing a general theory of interaction, because the surface of natural language does not directly correspond to its underlying action structure, i.e. the structure of speech is different from the structure of actions¹, where the structure of action in the general sense is fundamental to the structure of speech actions. Consequently, in the following I develop a theory of speech actions starting from a general theory of action which, on a very fundamental level, does not distinguish speech actions from other actions. The fact that I do not start from linguistic data does not imply that the theory delineated in the following is not grounded in empirical data at all. Quite the contrary, the main motivation for the following analysis is to show how representations (no matter how they are derived) can be grounded in a robot's sensorimotor functions.

3.1.2 Perspective representations

A robot able to engage in cooperative action must be capable of representing information from a first person perspective, so the formalism in which it represents that information must permit this. The

¹While speech has a systematic structure, actions have a methodic structure. E.g. in order to tie one's shoes, a particular methodic order has to be observed. A description of this task can reverse the order. (cf. [Hartmann and Janich, 1998] [Hartmann and Janich, 1991]).

capacity to represent information from a first person perspective is essential because it is crucial to the suggested notion of realism (see definition 1). The robot must be able to recognize and deal with the same problems of knowledge and ontology as humans and, like them, cannot take an omniscient and omnipotent third-person perspective. This is especially important for the natural interaction with humans. Humans cannot access other humans' mental states directly, nor can they observe themselves from the outside. Humans are bound to their private first-person perspective and so should a robot be ².

3.2 Objects, Things, Individuals

The purpose of this section is to introduce the notions of acquaintance with individuals as outlined in chapter 2.1.2.

3.2.1 Basic architecture

The following introduces the architecture proposed in figure 1.2.5, page 13 in more detail and discusses the construction of representations from single snapshots.

Architectural structures

To develop a representation formalism that does justice to the demands of realism and transparency (definition 1) from scratch would get us into a huge amount of work concerned with the form of representations. Instead I will adopt a highly prestructured format imposing constraints that would have had to be explicitly stated otherwise. The basic ingredient is quite similar to Standard-DRT's notion of a Discourse Representation Structure (DRS). DRSs can "be regarded as the mental representations which speakers form in response to the verbal input they receive" [Kamp, 1984, p. 5]. As stated above, I am not concerned with natural language for now but with developing the cognitive functions underlying the processing of representations. Hence I seek to extend the DRS-formalism to the treatment of non-linguistic data via a general notion of mental representation. The basic representational structure of an agent is pictured in figure 3.1. It consists of a box with two parts, where the upper part of the box holds representations of entities, and in the lower part of the box conditions which specify which or what kinds of entities these are.



Figure 3.1: The basic representational structure

Presentation and Representation

The architecture proposed in figure 1.2.5 was said to consist of

- (1) a sensorimotor frame

²Entering the third person perspective would also be impracticable, as the designer of a theory cannot ensure in advance complete knowledge about the mental states of all participants of a discourse.

(2) a presentation of reality

(3) a representation of reality

This coarse structure can be refined as follows:

Definition 6 *The refined architecture consists of*

- (1) *A sensorimotor structure (SMS) capturing the stream of data delivered by the object recognition engine and motoric control - the agent's notion of uncircumventable perception and motor control.*
- (2) *An external presentation structure (EPS) which presents SMS data - the agent's model of reality (pertaining to perceivable properties for now, future possibilities are considered in the next chapter).*
- (3) *An internal representation structure (IRS) displaying the mental representations constructed from presentations of SMS-data and future options.*

These three structures form an agent's architectural skeleton, which is pictured in figure 3.2 (cf. the general architecture given in figure 1.2.5).

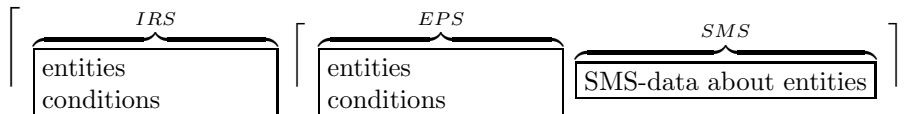


Figure 3.2: An agent's architectural skeleton

In order to display the different status of the entities in the three parts of the architecture, they are represented by a distinguished set of symbols:

- SMS-entities are represented by lower case greek letters α, \dots, ω
- EPS-entities are represented by latin letters from the beginning of the alphabet a, b, \dots
- IRS-entities are represented by latin letters from the end of the alphabet u, \dots, z

3.2.2 Inward construction of reference

The substructures of the agent's architecture are interconnected in two directions: information can flow from right to left or from left to right corresponding to the inward and outward direction of reference, i.e. the construction of reference markers and the identification of referents, respectively. Roughly speaking, information flow in the inward direction takes the form of constructing reference markers and conditions on them based on the output of the SMS layer. The outward direction of information flow is about the identification of IRS-conditions and IRS-reference markers in the EPS resp. SMS. It should be noted that all this informational processes take place *inside* the tripartite architecture of IRS, EPS and SMS.

Objects and things

In view of the concrete aims of this thesis I have to assume an intermediate step in the derivation of presentations of reality. The robot's object recognition engine presents its information in the discrete form of *single snapshots*³ containing information about the current distribution of color, shape and position in the robot's field of vision. To keep the analysis independent from particular object recognition systems with different output formats, I assume a *translation function* that transforms the raw data delivered by the object recognition system into a presentation format that allows for further processing. This translation function transforms the information about a recognized object in the SMS into a presentation of a thing in the EPS that takes the form of a number of conditions on a given EPS reference marker for a thing. EPS-Conditions are n-place relations, capturing properties such as color, position and shape of a thing (see section 6.1 for a detailed specification). The connection between a SMS-object and the corresponding EPS-thing is recorded by an 'external anchor'⁴ between the respective two entities (in this case between a SMS object and an EPS thing). The following example 3 gives a first idea of the translation of objects to things.

Example 3 Translation rule for things

The information "object α has properties C_1, \dots, C_n ", assumed to be represented in the SMS as $\alpha : \{C_1, \dots, C_n\}$ is translated into an EPS-term via

- the introduction of a variable a for a thing in the upper part of the EPS (if it does not already exist),
- the introduction of conditions $c_1(a), \dots, c_n(a)$ in the lower part of the EPS; the translation of C_1, \dots, C_n to c_1, \dots, c_n is discussed in more detail later.
- the introduction of an external anchor $\langle a, \alpha \rangle$ in the lower part of the EPS.

If the above rule is applied to an example setup, the presentation in figure 3.3 is obtained, illustrating the translation of an object on the table identified as red screw⁵ to a thing reference marker specified by the conditions 'screw', 'on-table' and 'red'.

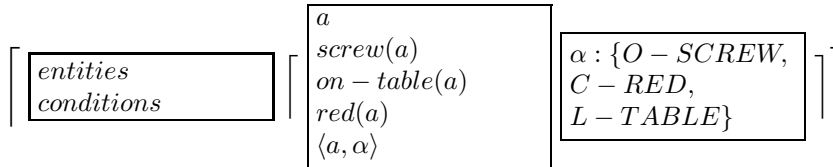


Figure 3.3: Example of an SMS to EPS translation

Things and individuals

A terminological note It should be noted that I distinguish between the terms 'objects', 'things' and 'individuals' to keep apart the different types of reference markers used in the SMS (for objects),

³I use the terms snapshot and moment interchangeable.

⁴The concept of anchoring is defined in definition 7 below

⁵The use of the SMS-terms 'screw' and 'red' here is chosen for simplicity. In actual implementations, the object-recognition data will probably come in some format not readable by humans.

EPS (for things) and IRS (for individuals). My use of these three notions is motivated by the idea that objects, things and individuals stand for concepts with an increasing complexity. Objects are of purely physical nature while things involve sensory processing and individuals require mental activities.

The purpose of Thing-individuals The translation of objects to things is only one part of the acquaintance with reference, the second part consists in the construction of distinguished reference markers in the IRS: individuals. In the approach proposed here, 'thing-individuals' (i.e. IRS reference markers that represent individuals of type 'thing') receive an analysis that is motivated by technical rather than philosophical considerations (see also section 2.1.2). This is for several reasons. First, individuals are supposed to include compounds, i.e. an individual may consist of several parts connected by some relational property. Second, individuals allow us to deal with incomplete SMS-data, as individuals need not necessarily be externally anchored but can also have the status of purely mental reference markers for which external identification has to be established. This applies in particular to occurrences of individuals in future scenarios where reference markers must be employed which do not have to be related to any existing objects but are intended to come into existence later. Third, together with the notion of things, individuals allow us to deal with temporal variation of properties. A thing's properties can change (the thing may even change completely) while the corresponding individual is still the same. In other words, individuals are abstract entities related to reality through an abstraction layer - the concept of things. Lastly, individuals allow to capture the ontological difference between things and individuals - while things are thought to capture perceptual categorizations, individuals involve the explicit use of mental functions such as explanations.

Construction of Thing-individuals Under which circumstances can a thing be represented as an individual? Technically, the current concept of individuals is introduced in the IRS, internally anchored in EPS-conditions concerning a thing that represent the unique quality of the respective individuals. IRS-individuals represent a 'handle' with which we can grasp the complex identification conditions of the respective individual. Such a handle for an individual is what I call a *name* for the individual (note that this has nothing to do with the linguistic distinction between proper names and common names). Example 2 gives a first impression of the individuation procedure.

Note 2 *Constructing thing-individuals*

A thing-individual x is individuated from an EPS containing conditions

$C_1(a_1, \dots, a_n), \dots, C_n(b_1, \dots, b_n)$ that make up the specific quality of x by undertaking the following steps:

- *Collect the respective individuating conditions in a new sub-EPS K*
- *Introduce a new reference marker d for K in the universe of the EPS*
- *Introduce a new reference marker x in the universe of the IRS*
- *Introduce a condition $name(x)$ in the body of the IRS, where 'name' specifies a handle for x*
- *Introduce an internal anchor $\langle x, d \rangle$ in the body of the IRS*

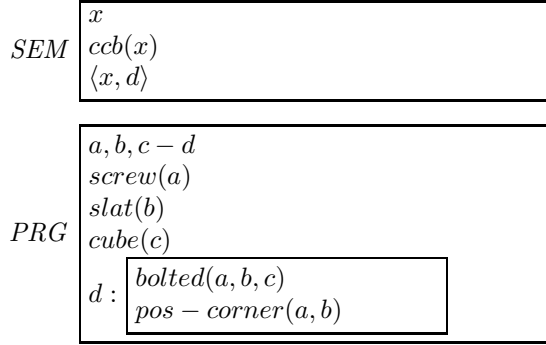
Sem-Prag-Concept 1 *Corner cube bolting (thing-individual)*

Figure 3.4: Example specification of a corner cube bolting.

The following example illustrates an application of the thing-individuation rule to a compound object, a corner cube bolting (abbreviated 'ccb') resulting in figure 3.6. For a specification of the meaning of the SMS-terms see section 6.1. The specification of a thing-individual consists of a semantic IRS-part *SEM* which specifies its representation and a pragmatic EPS-part *PRG* that specifies its identification.

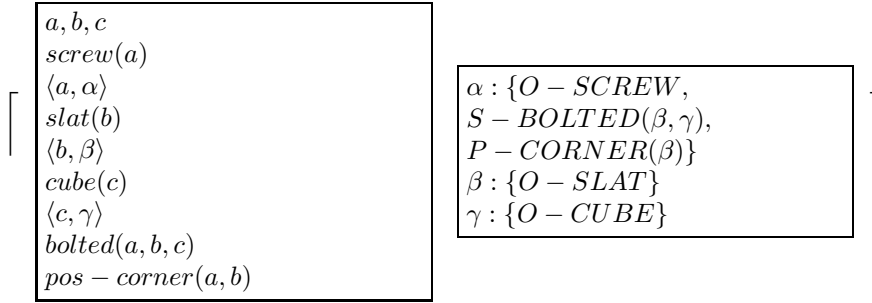


Figure 3.5: EPS(left)-SMS(right) configuration triggering the corner cube bolting individuation

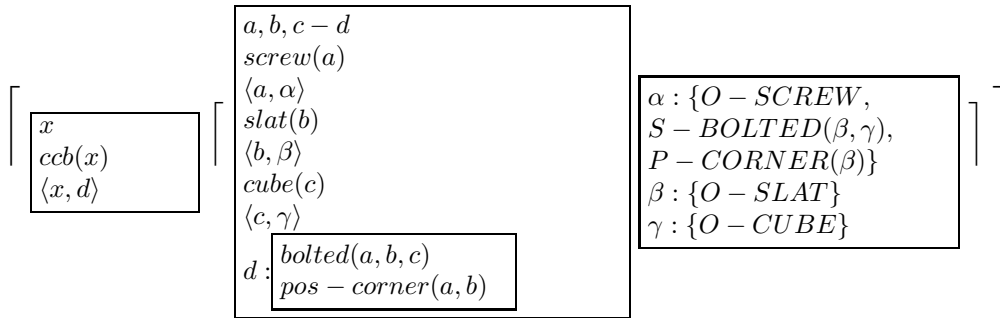


Figure 3.6: Configuration after the individuation of the corner cube bolting: IRS(left), EPS(center), SMS(right).

Figure 3.7 sums up the described construction of individuals from SMS-data.



Figure 3.7: Inward information flow from SMS to IRS.

anchors

An anchor is a relation between two reference markers or a reference marker and an SMS object. An anchoring relation states that the reference of the reference marker occupying the first slot of the anchor (which I call floater) is fixed to the reference marker resp. SMS object in the second slot (the source of the anchor).

Definition 7 Anchors

An anchoring relation is called

- *internal iff its floater is an IRS-reference marker and its source an EPS-reference marker*
- *external iff its floater is an EPS-reference marker and its source an SMS object term*
- *anaphoric iff its floater is an IRS-reference marker and its source an IRS-reference marker*

An important part of the interpretation of IRSs is the resolution of anchor sources of IRS reference markers. E.g. for the interpretation of an IRS representation derived from an utterance, anchor sources for the reference markers of the IRS representation must be determined as a preliminary to successful interpretation of the respective IRS representation as a whole. I call anchor sources that are not yet resolved 'unresolved anchor sources'. In addition, I introduce representations for unresolved anchor sources and definiteness constraints in note 3. The meaning and use of these notations will be discussed in detail in section 6.3.

Note 3 Representation of unresolved anchor sources

I distinguish four cases of unresolved anchor sources:

- *The specification of an anchor by means of a definite description (as occurring e.g. in an utterance) requires the floater to be anchored in a distinguished singular entity (e.g. 'this red cube'). This is represented by an arrow over the respective source:*

$$- \langle \text{floater}, \overrightarrow{\text{source}} \rangle$$

- *The floater of the anchor is not identified with any specific entity. This case requires the introduction of placeholders for the source to be resolved.*

– $\langle \text{floater}, ! \rangle$ *While the floater is not yet identified, it must be externally anchored i.e. the source must be retraceable to an SMS-object.*

– $\langle \text{floater}, ? \rangle$ *While the floater is not yet identified, it must not be externally anchored (it may be anchored in some future object).*

– $\langle \text{floater}, ?a \rangle$ *While the floater is not yet identified, it must be anchored in an anaphoric occurrence of the floater.*

3.2.3 Outward identification of reference

Now that the inward direction of thing-individual construction (deriving internal representations of thing-individuals from reality) has been outlined, it is time to say something about the reverse direction of outward reference marker identification; the identification of thing-individuals with real referents. As I introduced anchors to capture the inward construction of thing-individuals, I can use these anchors to trace back a reference marker for an individual to its object of reference. In the current setup, which is limited to externally anchored reference markers, the identification of thing-individuals can be executed by following the chain of anchors via the corresponding EPS-thing to a SMS-object as illustrated in figure 3.8).



Figure 3.8: Reference tracking from IRS to SMS, the outward flow of information.

3.3 Temporal variation

3.3.1 Preliminaries to the treatment of time

Following the discussion of temporally distributed perception in section 2.1.2 we will have to extend the procedures for reference computation introduced above to the analysis of temporal variation.

Individuation of temporal variation

Central to the treatment of temporal variation is the development of a notion of time that fits the needs of this thesis. This will be done in several steps, leading to a concept of time that matches the agent's architecture as pictured in figure 3.2. Concretely, besides thing-individuals another type of individuals, called 'time-individuals', is introduced. Time-individuals are grounded in EPS translations of successions of snapshots provided by the SMS-engine and consequent derivation of future possibilities by the agent layer. The relation between a time-individual and its corresponding EPS structure is established by the application of an explanation scheme of temporal variation as introduced in section 2.1.2 to the EPS. The basic idea is that a thing's temporal variation, i.e. the succession of two incompatible properties being true of the same thing can be individuated as an IRS time-individual. This task of individuating IRS time-individuals relies on the derivation of IRS thing-individuals from properties as proposed in the previous section 3.2.

Demands on the conception of time

The conception of time devised in the following should be related as closely as possible to a human being's conception of time while leaving open philosophical issues on time such as the opposition between intrinsic and extrinsic time and the relation between psychological and physical time. In addition, time has to be represented in an explicit way that provides contextual information for the processing of IRS representations (the demand of transparency). Similar to reference to objects, temporal reference works in two directions; inward and outward. The latter outward direction identifies IRS representations of

temporal variation in the SMS and EPS, the former, to be examined now, constructs IRS representations of temporal variation provided by the SMS and EPS. Hence, the central question of the following paragraphs is about how time-individuals are actually individuated with the help of temporal explanations. I will tackle this enterprise in two steps. First, I develop a notion of time based on perceptions (i.e. concerning the agent’s past up to the present as captured by a robot’s sensory instruments) which is then extended in a second step to a conception of future which is not grounded in perceivable reality but in the plan-based derivation of possible evolutions of reality.

3.3.2 Perception of time

Basically all that is available about temporal variation in terms of reality-grounded information is a perpetual sequence of snapshots delivered by the SMS. In the following, I discuss how a conception of time can be derived from the SMS data that fits the needs of this thesis.

The atomic structure of time

As the basic material at hand is a point-like structure of snapshots⁶ a set of time points will be induced in the following way: each snapshot that is delivered by the object recognition triggers the addition of a new EPS enhancing the agent’s existing set of EPSs with information about the current state of affairs. Each such new EPS is annotated with a time t_{n+1} introduced in the universe of the EPS, where n is the index of the previous EPS annotated with t_n , where $n \in \mathbb{N}$. The set of all time-indices is the set of time points we are after. Figure 3.9 pictures the construction of time points. The use of natural numbers to annotate time points neither indicates an ordering, an intrinsic meaning nor an internal structure of time points but is used only to clarify the design of the EPS structure to the reader.

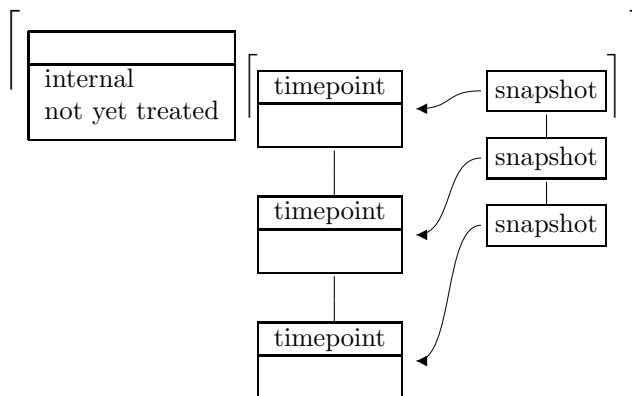


Figure 3.9: Construction of time points. Each SMS-data snapshot triggers a new EPS annotated with a new time point. Transitions between EPSs are called ‘atomic actions’, defined by the differences of successive EPSs. Transitions between SMSs are called ‘time slices’.

⁶The assumption of the discrete structure of time can be motivated by research on psychological, physiological and linguistic thresholds of perception: “People perceive ongoing activity in terms of discrete temporal events.” [Zacks et al., 2006, p. 1], see also [Fingelkurts and Fingelkurts, 2006] for neurophysiological arguments or [Fernando, 2006] for a linguistic argumentation

Order and direction of time

Intuitively, the most important property of time is its order and direction. In the following, I propose a separate introduction of order and direction at the SMS and EPS level, respectively. First, I assume the SMS structure to come with a direction-neutral order on the set of SMS-snapshots which is formally modeled as a three-place relation *between* in section 6.1. Second, at the level of the EPS, the set of EPS time points stemming from the indexing of SMS-snapshots as introduced above is assumed to be ordered and directed: the EPS is an acyclic directed graph (formally modeled with a binary relation $<$ on the set of EPS time points later on). The reason to do so is that we should not rely on an object recognition system having implemented basics of set theory such as the notion of a strong partial order. That is, the SMS output could also fit into a theory of reversible time, i.e. I do not require that the plain output of the SMS renders possible to calculate whether one snapshot is earlier than another snapshot. For the EPS, the additional assumption of a direction is mainly motivated by the fact that the EPS presents an agent's temporal view on the world in terms of earlier, later and coincident states of affairs. As the SMS is not directly accessible by an agent (but only transparent to us 'designers'), the transitions between EPS time points present the atomic intervals of time to an agent. In addition, I do not require the granularity of atomic actions to correspond to the granularity of SMS time-slices. Figure 3.10 pictures the notions of order and atomic actions.

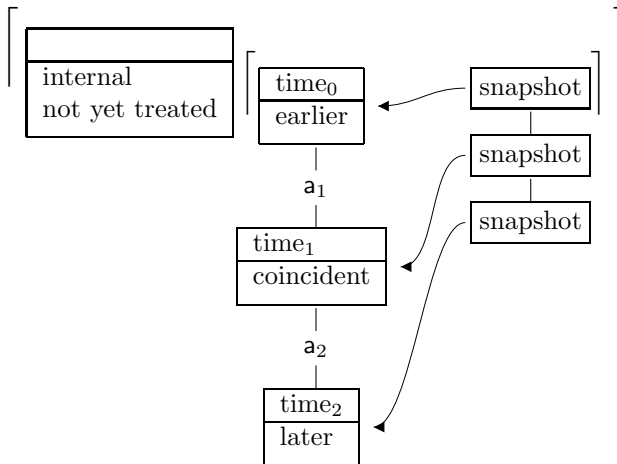


Figure 3.10: Introduction of a temporal order at the EPS level, where a_1 and a_2 are 'atomic actions', i.e. transitions between two EPS time points, a concept that is discussed in more detail in section 4.1.1.

Figure 3.10 pictures time as far as it is determined by the translation of perceptions to presentations of experiences. This is of course only half the story: there is also a future toward which the agent is directed. Consequently, the picture of figure 3.10 has to be extended to a presentation of the future and mechanisms have to be introduced to construct those presentations.

3.3.3 Construction of temporal reference markers

Incorporating the future

The modal structure of the future An essential ingredient to the design of an agent who is able to perform meaningful actions is the concept of future. Without a conscious conceptualization of future

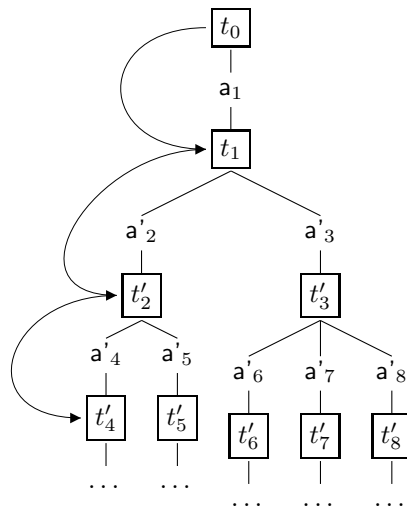


Figure 3.11: EPS branching-time structure. The arrows indicate a scenario, a path through the tree. Each time presents a state of affairs. In this drawing, t_1 is the current position of the agent in time, as from this point on, the future possibilities branch, while the past is determined by the output of the SMS.

possibilities there can be no planning or intentional performance of actions⁷. The future is not captured by SMS-data and thus not presented in the current setup of the EPS. Nevertheless, future possibilities must be integrated in the picture of time in figure 3.10. To do that we must take into account that reality may evolve from the agent's current now in different ways⁸, depending on what actions are performed by the agent herself and other processes going on at the same time. In other words: while the past of an agent is determined by the information provided by the SMS, the future consists of numerous possibilities for which the outcome is not determined at the current 'now'. The vocabulary used in this description indicates the tools I will use to formalize the nature of the future - I will deploy a variant of modal logic, branching-time logic. In this formalism, time is modeled as a tree branching in the direction of the future, where each of the branches represents a possible evolution of the world depending on the agent's own choice, her abilities and the actions performed by other agents. The cognitive ability to build and maintain such models of the future plays a central role in the planning of verbal and nonverbal actions and we will devote chapter 4 to spelling out how this is done in detail. For now it is assumed that the possible evolutions of reality can be captured by the introduction of possible EPS-times; those are annotated with a prime to indicate their modal status. Figure 3.11 gives an impression of what the intended structure of time looks like.

⁷In fact, the ability to 'predict' the future has been said to be one of the most distinctive features of humans: "the fundamental purpose of brains is to produce future." [Dennett, 1991b, p. 177]. Only the use of the concept of future allows for meaningful commerce with reality, as e.g. Craik states: "A man observes some external event or process and arrives at some 'conclusion' or 'prediction' expressed in words or numbers that 'mean' or refer to or describe some external event or process which comes to pass if the man's reasoning was correct." [Craik, 1967, p. 50]

⁸The concept of an agent's current 'now' will be introduced in detail a few paragraphs below.

Populating subjective time The next step in the development of the agent's concept of time is to introduce the notions of past, present and future, which can be considered as mental functions that define a space for temporal reasoning⁹. A first observation is that agents are not bound to deal with time in terms of atomic actions, but they can group sets of atomic actions together so as to form new entities that can serve as time-individuals. That is, the agent picks out certain temporally distributed properties of things from EPSs that are part of the EPS tree structure. She adopts a set of atomic actions which chain the respective EPSs together as a 'package' of temporal variation. This package can then potentially be transformed into a time-individual under the assumption of an explanation in terms of causality, desires or intentions. Such an explanation must relate a chain of atomic actions with the chained EPSs stating properties of things. That is, the explanation should explain the change in a thing's property as being induced by some underlying causal force, desire or intention. By presuming that a thing's temporal variation is driven by an underlying force, this force can be consulted to explain the thing's temporal variation. Consequently, a time-individual is a 'package' of atomic actions, properties and things that is extracted from the EPS time tree with respect to a certain type of explanation of agency. This is what the discussion in section 2.1.2 proposed - the explanation of temporal variation constructs representations of time-individuals. Consequently, the next task is to specify the types of reference markers for time-individuals as outlined in section 2.1.2 in more detail.

Properties and States

Individuation of States For the argumentation I pursue here it is important to keep in mind that states have, like events, the status of individuals and are thus represented in the IRS. In the EPS, states have two manifestations. First, the properties of an EPS-thing at a certain time can be considered as an 'instantaneous' state. The linguistically more relevant case is that of an 'extended' state, where a thing's temporal variation is thresholded over a certain period of time. Figure 3.12 pictures the relation of stative IRS time-individuals and the EPS.

⁹The conception of time as space is a quite powerful metaphor that is used throughout the literature on time, see e.g. [Smart, 1949], [Evans, 2003]

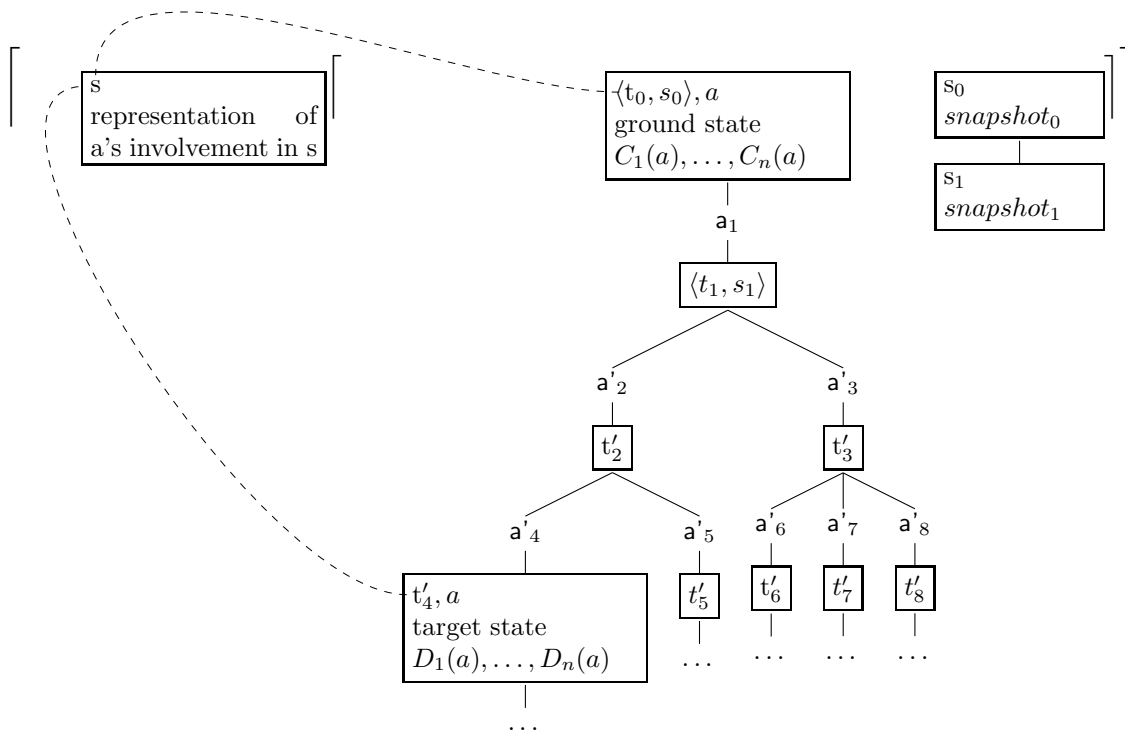


Figure 3.12: Individuation of an IRS time-individual of type state. Assuming that the EPSs at t_0 and t'_4 do not differ with respect to a certain threshold concerning the conditions specifying the thing reference marker a , the sequence can be individuated and represented in the IRS as a 's being in a certain state. Note that the EPS is concerned with the identification of states and not with the predication of states. That is, while predication of a state expresses stable conditions on a linguistic level, this stability is not mirrored on the physical level where there is always some kind of change going on.

Events

The connection between states and events States play a fundamental role for the derivation of events: events are defined by their being a cause between incompatible properties of a thing. If there is temporal variation in the properties of a given thing a given by a succession of two slices s_1 and s_2 such that for some property P , a has P in s_1 but not in s_2 , then it is assumed that some event e is responsible for this variation and that e is temporally located between s_1 and s_2 ¹⁰. As is the case for states, events can be either instantaneous or extended in time. Of special interest is the case where events have duration. This arises when the states s_1 and s_2 are separated by several times in between s_1 and s_2 . Note that this notion of duration does not entail any metric (i.e. quantity of duration).

Individuation of events A tree segment of EPS times and transitions is individuated as an event through the assumption that the structure serves the realization of an intention, brings about a goal or is driven by physical causality. This notion of event conceptualization is strongly supported by recent research in the behavioral and neurological sciences, where experiments suggest that ongoing activities are automatically and spontaneously segmented into hierarchically organized parts and subparts with the help of “bottom-up processing of sensory features such as movement and [...] top-down processing of conceptual features such as actors’ goals.” [Zacks and Swallow, 2007, p. 80], where top-down processing seems to be preferred by the subjects of the experiments. This ‘hierarchical bias hypothesis’ states that observers spontaneously encode events in terms of partonomic hierarchies: “observers are biased to perceive ongoing activity in terms of discrete events organized hierarchically by ‘part-of’ relationships.” [Zacks et al., 2001, p. 30]¹¹. Consequently, the boundaries of events coincide with the segmentations imposed by the explanation of behavior - “behavior episodes” [Barker and Wright, 1954]. Among the types of temporal variation that promote a segmentation of an ongoing activity are e.g. [Barker and Wright, 1954, p.236]:

- (1) Change in the “sphere” of the behavior from verbal to physical to social to intellectual, or from any one of these to another.
- (2) Change in the part of the body predominantly involved in a physical action [...]
- (3) Change in the physical direction of the behavior. [...]
- (4) Change in the behavior object “commerced with” [...]
- (5) Change in the present behavior setting. [...]
- (6) Change in the tempo of activity [...]

If we abstract from these concrete indicators for segmentation (which will nevertheless play an important role in the actual construction of events later on) a more technical notion of events can be derived. A time-individual of type event is individuated from a structure of EPS times and transitions ranging

¹⁰Recent research in psychology has brought up evidence for this way of analyzing events: “the organization of event descriptions closely paralleled the behavioral segmentation data.” [Zacks et al., 2001, p. 63]

¹¹This idea is closely related to research in text understanding, where understanding narrations of events involves the extraction of plan-like structures (called scripts [Schank and Abelson, 1977], schemata [Rumelhart, 1975, 1980] or frames [Minsky, 1972]. One can also establish parallels to theories of discourse realization, such as the recipes of [Lochbaum, 1991].

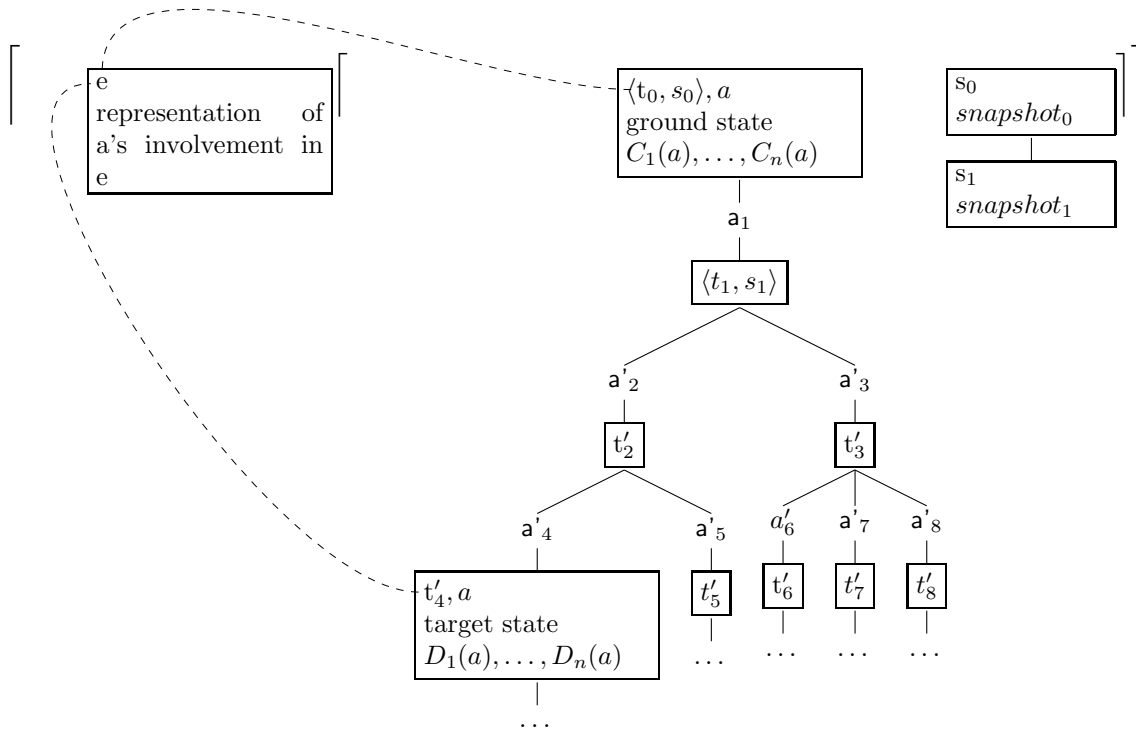


Figure 3.13: Individuation of a time-individual of type event from the EPS branching-time structure. The dotted line indicates the extension of the IRS time-individual, which is determined by the *difference* between the EPS-conditions at t_0 and t'_4 concerning a to which the explanation is applied.

from a starting state to a target state, employing the EPS-information on temporal variation as well as information about involved causes, goals or intentions. It is obvious that capturing time-individuals in a representation requires a lot more work than the representation of thing-individuals. I will devote a separate section to the detailed discussion of this topic and only give a short outlook on what is to be spelled out in more detail in section 5.3. In that section, I introduce anchors for time-individuals. Roughly speaking, the anchor source of a time-individual ev specifies the explanation serves the referential identification or construction of the respective time-individual. I will leave it at that and picture the current status of analysis in figure 3.13.

The agent's now

Indexical now One last point is still open to discussion and missing in figure 3.13: the introduction of an agent's present that allows her to relate IRS time-individuals to EPS-times. This is done by relating IRSs to the currently processed EPS with the introduction of an indexical now-constant at the IRS level. The IRS now-constant always points to the EPS time-point currently constructed from a SMS-snapshot. With each new EPS, the IRS now-constant is anchored in the most recent EPS time point. It can therefore be understood as a pointer constantly moving in the direction of the future. With the help of the now-constant, future and past can be defined by establishing relations between time-individuals and the now placing them either before, inside or after the now¹². It should be noted that an agent's now may outlast a single moment - it can become an 'extended now', if the temporal extension of a time-individual the agent is aware of at the time of the now goes beyond a single time. This arises in connection with plans and intentions of the agent which can extend the now to an interval starting at the current EPS time and reaching into the future defined by currently active plans and intentions.

3.3.4 Identification of temporal reference markers

What has been said so far captures the inward direction of temporal reference in an informal way and a formal treatment is still to come. But the outward direction of temporal reference still awaits discussion, too. Outward identification of temporal reference markers can be seen as the inverse operation to individuation and construction. The reference of a time-individual involves a structure of temporally distributed EPSs (i.e. time points) of which the exact specification is to be given by the IRS-formulas which represent the time-individual. The information that should be provided by an IRS-representation of time-individuals for the purpose of identifying a time-individual in the EPS time tree consists of:

- involved agents and things
- the explanation type which was employed to construct the time-individual (causality, desires, intentions)
- a representation K capturing the target states of the time-individual (causes, goals or intended states of affairs)
- the relation between the respective time-individual and the now-constant

Given this information, the idea which will be employed to trace the reference of time-individuals back to the EPS can be described as an attempt to accommodate the IRS time-individual in the EPS, i.e. the EPS structure from which the time-individual was individuated must be recovered. Speaking in semantic terms, this task attempts to identify EPS structures that satisfy the demands of the IRS formula representing the time-individual. In other words, determining the reference of a time-individual can be understood as trying to satisfy the formula for the respective time-individual by checking whether an EPS-structure exists that allows for the "embedding" of the time-individual in this structure¹³. Chapter 6 specifies the interpretation algorithm for time-individuals in full detail.

¹²This foreshadows how tense will be treated and is in accordance to the Standard-DRT analysis of eventualities and tense.

¹³The second option comes close to what [van Lambalgen and Hamm, 2004] call 'temporal profile calculation'. In fact, the approach suggested here can be used to feed the event calculus developed there.

3.4 Outlook

This chapter has introduced the architectural skeleton of the formalism in an informal way. In order to spell out the notions of IRS, EPS and SMS formally, some more work has to be done. This concerns the design of the agent layer which is needed to set up the modal possibilities of the EPS and provides the means which are necessary to capture explanations of temporal variation, thus making it possible to specify semantic interpretations for representations of time-individuals. The agent layer is developed in the next chapter 4. This will enable me to give the missing formalization of this chapter's findings in chapter 6.

Chapter 4

The Agent Layer

This chapter carries out the shift of perspective announced in section 3.1.2 by moving the focus of analysis from an external descriptive viewpoint to an internal agent's prescriptive point of view. The previous chapters merely considered an agent as a passive receiver of presentations of perceptual information and future options that triggers the formation and manipulation of her representations. This chapter develops the agent's (re)active component. In the course of it it should become clear that it is this component - the agent's active and focused use of her mental functions - that shapes her representations and consequent actions. It may not be surprising that the procedures that the agent employs to control her own activities are the same she uses in her explanation of the external entities she perceives¹. A distinctive feature of intelligent and rational agency is that it connects behavior and mental attitudes in a reasonable and controlled way given the pressure of limited resources such as time and abilities. Capturing this special quality of the relation between mind and action poses new demands on the theory developed up to now. In particular, the following issues have to be discussed:

- (1) the setup of feedback processing mechanisms: *the basic connection between mind and action*
- (2) the design of plans capturing means-end structures: *the complex connection between mind and action*
- (3) the delineation and anchoring of belief and facts in the context of possibility and necessity
- (4) the integration of intentions and desires.

4.1 Mind and Action

This section discusses issues concerning the relation between mind and action, picking up the basic structure of the EPS structure and relating it with mental attitudes such as intentions, desires and beliefs.

¹The parallel nature of explaining reality and the self is supported in particular by the discovery of mirror neurons [Rizzolatti and Fogassi, 1996].

4.1.1 From actions to strategies

Actions can be distinguished according to their complexity, ranging from simple 'atomic' actions to complex plans. The following paragraphs discuss this hierarchical structure.

Atomic actions

Atomic actions The term 'atomic' or 'basic' action has been introduced in a preliminary way in sections 2.1.2 and 3.3.2. Atomic actions were given as relations between two EPS times constituting the smallest unit of temporal granularity at the EPS level. While there are no formal objections against this, I cannot rely on entities which are given in these terms but have to ground them in the robot's functions and abilities. A hint of how atomic actions could be grounded is what we said about their explanation in terms of causality (in the sense of definition 2). Atomic actions are to be considered as the smallest causally interacting units of reality in terms of which all other more complex conceptions of action (such as plans or intentions) can be analyzed. But if atomic actions are to be grounded in the interplay of causes and effects, this requires the introduction of adequate notions of cause and effect. That is, the definition of atomic actions essentially involve causes and their resulting effects. Thus the definition of an atomic EPS action involves at least two states of affairs, one at the start and one at the end of the atomic action. The atomic action is then defined as the transition between these two states of affairs identified as cause and effect. Given these considerations from the designer's point of view, next we should discuss how atomic actions are constituted from the internal perspective of an agent.

From causality to feedback loops From the agent's perspective on current actions (both her own actions and those not performed by her), an atomic action is a relation between her perceived external states of affairs and her internal mental configuration. This is true in particular of her atomic actions. While the preconditions of a current atomic action are grounded in reality, this is not the case for a current atomic action's postconditions. This does not go beyond what we have said about atomic actions already; but the crucial point here is one about rational agency: an intelligent agent will try to anticipate the causal effects of her action by devising an imaginary state of affairs that she believes to hold after the action has been performed. Whether and how the agent's belief about the future outcome of her action matches reality is not determined until the agent actually receives feedback on her performance in the form of an actual EPS presentation - which updates the agent's last presentation of reality (see section 2.2.1). This updated presentation of reality can in turn be used to generate new options for further actions to achieve further effects. The entire process takes the form of a circle, a *feedback loop* that constantly updates the agent's beliefs about the future with new actual facts that her own actions and those of other agents produce. Figure 4.1 pictures such a basic feedback circle.

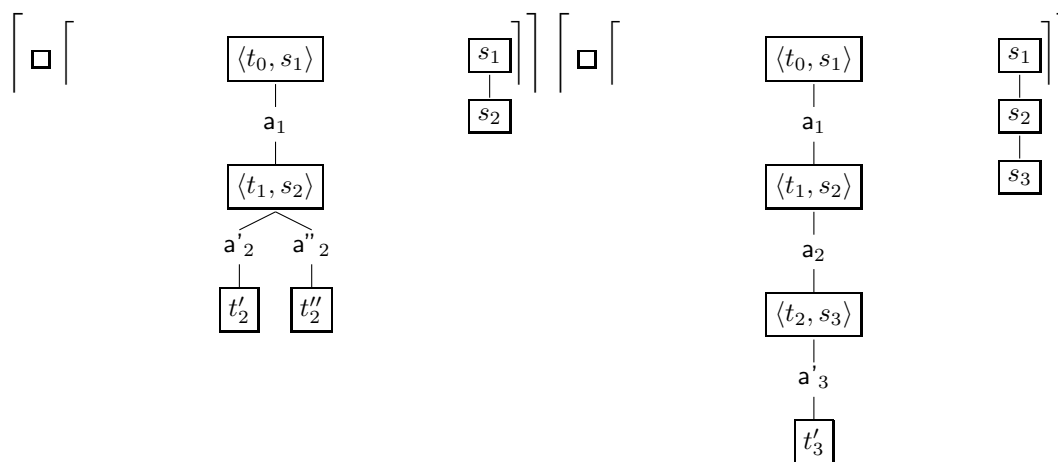


Figure 4.1: A feedback loop of an atomic action. In the left figure, the agent's EPS presentation allows at the present t_1 for the choice between the possible actions a'_2 and a''_2 , leading to the projected states of affairs t'_2 resp. t''_2 . If the agent chooses to perform action a'_2 (right figure), the real state of affairs t_2 derived from snapshot s_3 is not necessarily equal to t'_2 . This causes the agent's previous beliefs about the future represented by t'_2 to be updated by the actual state of affairs s_3 at t_2 .

The utility of feedback It may have been noticed that the analysis of atomic action sketched above already exhibits the close-knit relation between feedback and core questions of epistemology and action-theory. In order to discuss questions on topics such as belief revision or intentions, we need to look at certain sequences of atomic actions - those sequences which constitute plans. This is what the next section does.

Plans

From atomic to complex actions In the non-trivial case, the realization of an agent's goals requires the execution of more than one atomic action. In order to bring about the desired state of affairs a sequence of actions has to be performed. It is distinctive of rational agents that many of their actions do not succeed each other in some arbitrary order but follow a *plan*².

Hierarchical Structure of Plans Plans are characterized by their *hierarchical* structure. On the top level, a plan can be considered as a "global representation of an action" [van Dijk and Kintsch, 1983, p. 65]. But at lower levels we find subplans; each of these contributes to the plan as a whole, while the executions of some subplans create the preconditions for executing some other subplan. In other words, a plan elaborates an 'abstract' action by specifying which choices should be made depending on the currently holding state of affairs. Consequently, planning is intrinsically dynamic and context-dependent. It must take into account the variability of the environment by incorporating a reactive component which allows for the control of the large number of degrees of freedom of one's own actions, the evolution of reality and interactions between the agent's own actions and other processes going on at the same time. At this point, the concept of feedback loops introduced above (section 4.1.1) comes into play again, as it is necessary to detect in the course of executing a plan "upcoming problems so they can be dealt with ahead of time or at an early stage when they are harmless" [Hofstein, 1994, p. 63]. This requires the processing of feedback from atomic actions as well as from higher-order level plans. Thus feedback processing has to be extended to apply to plans too, but that does not pose serious problems. Depending on the granularity of analysis one adopts, whole plans can be treated like atomic actions so long as their internal structure is ignored and only their pre- and postconditions are being considered. So we can substitute complex plans for the atomic actions in the conception of feedback loops (figure 4.1) without running into trouble.

Demands on Plans A further aspect of planning has to do with the *limited resources* that are available to human and other agents (such as processing power and speed, memory storage, physical abilities and knowledge) and with the need for *coordination* of our actions with our own varying needs and our changing environment. Plans reflect these two constraints in that they (1) are only *partially specified* and (2) their specification is *dynamic*, in that a plan can be gradually refined and adjusted in the course of plan execution. In the early stages of execution, a plan may be specified at a quite coarse level involving just a global course of action. The next step is to determine how the plan can be realized in the actual state of affairs and consistently with other plans that are up for execution as well. This leads to the specification of a concrete course of action that are attuned to the actual circumstances. It would

²The spectrum of research on plans is enormous: (plans play a central role e.g. in Artificial Intelligence [Pollack, 1992, Cohen and Perrault, 1986], Psychology [Hofstein, 1994], Linguistics [Kamp, 2007] Sociology [Rogoff et al., 1994], Philosophy [Bratman, 1987]) and I can only mention some of the ideas relevant to the discussion here.

be of no great use to completely specify a plan in advance, since the conditions in which it will have to be carried out are rarely known in advance and will often be only partially known once execution gets under way. Thus “settling in advance on such partial, hierarchically structured plans, leaving more specific decisions till later, has a deep pragmatic rationale.” [Bratman, 1987, p. 29].

Constraints on Plans The distinctive mental attitude attached to the concept of planning is that of rationality; in two ways. First, in the sense spelled out above, i.e. rationality serves to maintain actionability under the pressure of limited resources and coordination. Second, rationality imposes constraints of *coherence* and *consistency* on the design and execution of plans [Bratman, 1987]:

- Coherence of a plan has to do with adopting the right means in the right place. The correct means need to be filled in at the respective underspecified parts of the plans. If that does not happen, or if means are chosen that do not lead to the realization of the plan, the plan is means-end incoherent and is likely to fail.
- Consistency alludes to the beliefs (internal consistency) and practical possibilities of an agent. Intending to do something that is contradictory to one’s own beliefs about what is practically feasible leads to an execution of the plan corresponding to the intention that is likely to fail.

4.1.2 Mental attitudes

When we put the structural issues on planning aside and turn to the content of plans, the agent’s mental attitudes move into the focus of analysis and plans are to be considered as “complex mental attitudes” [Pollack, 1990] inextricably tied to the agent’s desires, beliefs and intentions. While it is indubitable that these mental attitudes play a crucial role in the motivation, formation and execution of plans, the precise connection between these attitudes and plans as abstract (not necessarily mutual) structures is subject to controversial discussion. Before I say more on that topic, I think it is useful to introduce the concepts of belief, desire and intention in more detail.

Belief

Two main options have evolved for the formal analysis of belief. One, the modal analysis focuses on the semantic content of beliefs and the other on their syntactic structure.

Semantic approaches The classic formulation of a modal analysis of belief is due to Jaako Hintikka [Hintikka, 1962]³, who used a possible worlds-approach to model belief and knowledge. The basic idea behind semantic approaches to belief is that the totality of an agent’s beliefs divides the set of all possible worlds into those worlds which are compatible with that totality and those which are not. Formally, compatibility can be modeled with the help of a relation of accessibility on the set of possible worlds, i.e. for a world w' to be compatible with the beliefs of an agent in w , the relation of accessibility is required to hold between w and w' . Consequently, an agent’s beliefs in a given world w are defined as the set of propositions which are true in all worlds accessible from w .

³DRT also employs this type of modal belief logic: [Kamp, 1990, 2003]

Syntactic approaches Besides the semantic approach to an agent's beliefs, her beliefs can also be treated on a syntactic level [Konolige, 1986]. In this case, the elements constituting an agent's 'belief base' are coded by formulas from some formalism such as predicate logic. The 'logic' of an agent's beliefs is also specified in syntactic terms e.g. in the form of a proof system based on deduction rules together with a proof manager responsible for updating the given belief set and answering questions that are put to the agent. In this conception, an agent's beliefs form a belief system, summarized in definition 8.

Definition 8 *A belief system consists of*

- *a set of deduction rules,*
- *a set of sentences representing the agent's beliefs and*
- *a set of control strategies that manage the application of inferences to the agent's belief set.*

Drawbacks of sentential and modal belief models Both the modal and the sentential approach to belief have certain drawbacks that prevent them from being theories that we can employ directly for the purposes of this thesis. On the one hand, modal doxastic logic suffers from the problem of 'logical omniscience' [Hintikka, 1975] which entails that an agent always believes all logical consequences of whatever she believes. This becomes a problem if one takes into account that humans as well as computational agents possess only limited inferential resources, which would be used up long before all logical consequences of a given belief set would be computed and stored. In addition, from a computational point of view the modal approach to belief is not as tractable as the sentential approach [Gärdenfors, 1992]. On the other hand, the sentential approach to belief does not assign semantic content to beliefs, i.e. it does not ground beliefs in their denotation but restricts analysis to the syntactic surface. Sentential theories of belief also have to face the fact that it is doubtful that beliefs are sentences - one can argue that this is not the case but that belief is directed toward the propositional content of a sentence, i.e. the corresponding 'information state'.

Hybrid belief logics All in all, from a formal point of view, a combination of both approaches would suit my needs best, i.e. a theory of belief that avoids the failure of the modal approach to account for resource constraints while grounding its (resource bound) syntactic manipulations semantically [Fagin and Halpern, 1987]. A 'hybrid' account has been developed using a combination of DRT and Computational Tree Logic in [Singh and Asher, 1993]⁴. It is this approach I want to use as inspiration for the design of a theory of belief that matches the goals of this thesis. In principle, I already have the basic ingredients for the design of a suitable theory of belief. Section 3.3 introduced a modal structure in the form of the branching-time structure of the EPS. At that time I did not say much about how to generate and interpret the modal structure of the future but now I can give a preliminary interpretation of the EPS layer as a model of an agent's beliefs. The second part of this chapter will then outline how these EPS-possibilities are generated by the agent's rational control system.

⁴Nicholas Asher's original account to belief in the framework of DRT [Asher, 1986] is hard to categorize. On the one hand, it is built upon modal semantics but on the other hand beliefs pertain to representations and not propositions.

Locating belief The leading question of the following paragraphs is about what an agent should, can or must believe. In order to answer this question, we must first examine a topic that is ignored in all the approaches to belief mentioned above. How can we determine which truth value we should assign to certain formulas in certain worlds. That is, how can we put an agent into a position to gather knowledge how to arrive at a valuation V of a certain formula ϕ in a world w . Standard modal logic does not say anything about how this could be achieved. Instead, the valuation V is supposed *to be given*. We can not rely on V being given but, as designers, we must specify how an agent can extract V from her internal and external states of affairs⁵. That is, an agent needs to be able to dynamically determine the possibilities in which she can, should or must believe. In the proposed architecture (definition 6, page 37), possibilities⁶ occur in two ways in the EPS. The probably more fundamental type of possibility is tied to the way in which the agent represents the future, i.e. her beliefs as to what might occur depending partly on how she will act. The other type of possibility is related to past possibilities that might have occurred if some other action had been chosen at an earlier time. Both these types of temporal possibility constitute what the agent *should* believe in. As we do not want the robot to be encapsulated in sceptic solipsism, I assume an agent *must* believe in what she perceives. That is, all EPSs derived from SMSs are to be considered as necessary aspects of reality⁷. The third type of belief is of a trivial nature: an agent *can* believe in everything that she imagines possible even beyond her own abilities (e.g. finding a unicorn). Summing up, figure 4.2 pictures the location of past and future possibilities in the EPS-layer.

⁵The problem of 'givenness' is handed down the other components of modal model theory, the set of possible worlds W and the accessibility relation R .

⁶This is the point to make a note picking up what has been touched on during the discussion of EPS modalities. Instead of considering all conceivable possibilities, we can drastically reduce the complexity of an agent's belief set if we assume that we only need to consider the (possible) temporal evolutions of the world the agent is situated in. Even from a philosophical point of view, it is hard to imagine a world that has absolutely no temporal or personal connection to the world we know - this concerns in particular our own identity and temporal structure - imagining a world in which oneself does not exist is only insofar possible as it is the existence of myself that imagines such a world. It may be useful for philosophical investigations to assume such possibilities, but they do not play any role for the type of reasoning that is in question here. An agent can only act in her world - thus it would not be sensible to implement reasoning in worlds completely detached from reality. Note that this does not affect temporal variants of the current world, in particular past (possible evolutions that might have occurred) and future possibilities (possible evolutions that can occur).

⁷Please notice that this does not exclude the possibility that the agent can delude herself about what she perceives, as illusions are to be treated on a higher level of mental processing than that of the SMS-data. Illusions occur if a mental representation does not match the actual percept, and they should be handled accordingly. In addition, this does not imply that an agent's history is consistent. In fact, the temporal evolution of an agent's beliefs about reality is distinguished by the occurrence of belief revisions, to which we will turn in the next section.

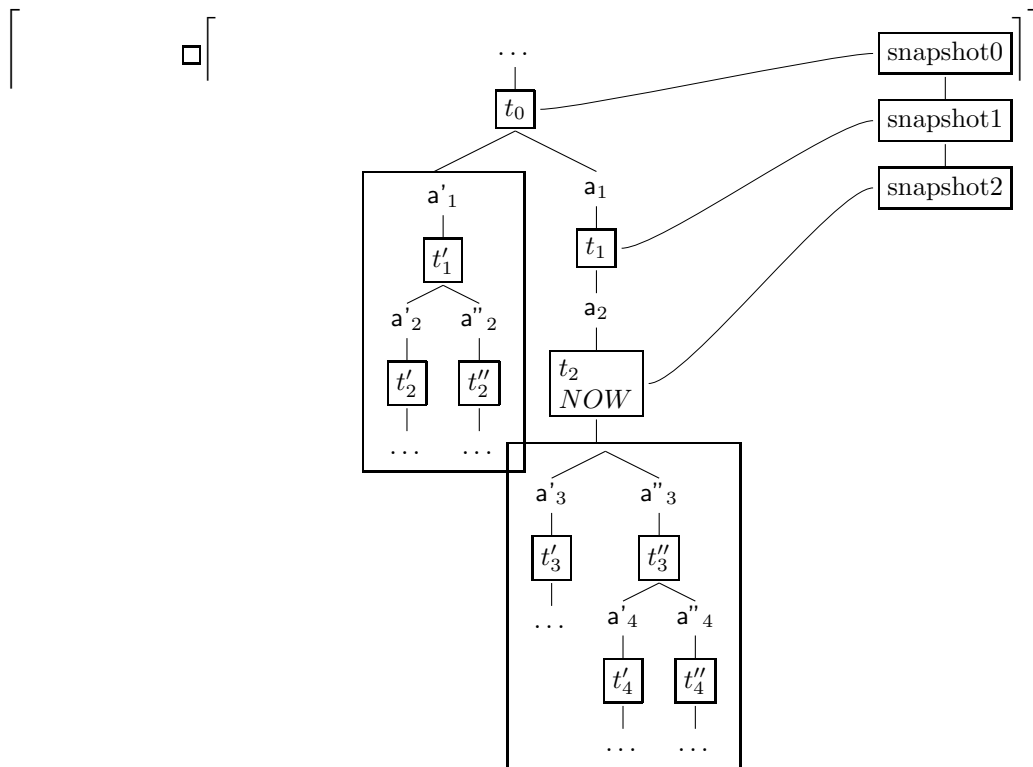


Figure 4.2: Belief location in the EPS layer. Past and future possibilities are boxed.

By the way, the above question about the nature of the objects of belief has been answered in the discussion of possibilities as constituents of the EPS. I assume that it is not sentences in which an agent believes but possible states of affairs in the EPS of which their structure is determined by their IRS-representation.

Belief and anchors A point intimately related to the above discussion of the location of belief are the varying grades of belief reliability, as beliefs can range from arbitrary to fully justified. How can we distinguish between different reliability degrees of beliefs? The cognitive architecture that has been outlined in this and the preceding chapters enables us to make some distinctions. If beliefs are represented at the IRS level, a measure of reliability of a belief at this level is given by whether or not its IRS representation is anchored. First, the IRS representation itself can be anchored in the EPS - I called such anchors internal anchors - and second, components of the EPS involved in this anchoring can in turn be anchored in the SMS. These anchors were called external anchors (see definition 7, page 41). Double anchoring - of the IRS level representation in the EPS and of the relevant EPS parts in the SMS testify to the perceptual basis of the belief representations. In this case we also say that the IRS representation is externally anchored. The reliability of such a belief is treated as absolute in the present setup; there are no options for speculating about its correctness. Representations that are EPS anchored, but where the part or parts of the EPS that functions as its anchor are not SMS anchored in turn, constitute a second, lesser degree of reliability, IRS representations without any kind of anchoring (i.e. for which the anchoring has to be established) an even lower degree. The central point here is that the use of anchors allows to distinguish between justified and unjustified belief - justified beliefs are what I call facts (and I refer to belief in facts as 'know-that'); I reserve the term 'belief' for the unjustified beliefs. Summarizing, I propose the following categorization of beliefs and facts:

Definition 9 *Belief and Facts*

- *An agent's fact base facts consists of the set of externally anchored EPSs and of the IRSs which are internally anchored in externally anchored EPSs.*
- *An agent's belief base beliefs contains*
 - *all unanchored EPSs and*
 - *all IRSs which are not anchored in externally anchored EPSs.*

The distinction between beliefs and facts at the IRS level has two aspects pertaining to the different types of individuals at the IRS level. I distinguish the belief in or know-that about the existence of thing-individuals from the belief in or know-that about propositions. This captures the difference between belief in propositions (e.g. the belief that the agent will build a corner bolting vs. the know-that that an agent built a corner bolting) and the existence of things (the difference between belief in the existence of a corner bolting and the factual existence of a corner bolting)⁸.

⁸From a logical point of view, the difference between facts and beliefs with respect to thing-individuals mirrors the distinction between de-re and de-dicto interpretation of noun phrases. Facts are anchored in the agent's presentation of reality, 'non-factual' beliefs on the other hand do not have this connection to reality.

Definition 10 *Operators for facts and beliefs*

If K is an IRS and x a thing-individual reference marker (for the entity to which the belief resp. fact is ascribed), then the following expressions are IRS-conditions:

- xK_tK (x knows that K , where K_t stands for $\text{Know}_{\text{that}}$)
- xBK (x believes K)

Definition 11 *Anchoring in facts and beliefs*⁹

If x is an IRS reference marker, then the following expressions are (variable) IRS-anchors:

- $\langle x, ! \rangle$ (x is anchored in a fact, i.e. externally anchored)
- $\langle x, ? \rangle$ (x is anchored in a belief, i.e. internally anchored)

Feedback and belief revision Given what was said about feedback loops in the beginning, I can now state a preliminary relation between belief and feedback. Once an agent has received feedback on the action she just performed, the resulting EPS affects the agent's beliefs in two ways.

- The feedback EPS may contain an updated picture of reality that is not in agreement with past presentations (e.g. an object may have changed its position). Special mechanisms are needed to prevent this from leading to an inconsistent presentation of reality. We will have to keep this point in mind when we come to the formal design of EPS building procedures: it must be ensured that the actual EPS is up to date *and* consistent.
- The revised picture of reality that results from feedback may also affect the content of future possibilities: new options may have evolved or previous options become unreachable. This point will have to be kept in mind when I will spell out the planning system in the next section 4.2. It should also be noticed that a change of future possibilities in the EPS level normally affects the evaluation of IRSs.

Desires

Once the reactive nature of the agent is taken into consideration, it is necessary to discuss how an agent's actions are motivated, i.e. for which reasons an agent performs an action. This is a heavily debated question within action theory, where the classical answer is due to Donald Davidson [Davidson, 1963]. Basically, Davidson argues that an agent's actions are caused by a combination of her beliefs and desires. It is characteristic of desires that they can be inconsistent and do not require the agent to know how to achieve them. Human desires are often caused by physical needs (such as the need for food), or they may be grounded in phantasy; (e.g. the desire to meet a live unicorn); and their emergence does not follow principles of rationality. It is only when desires are inputs to the agent's deliberation, which may result in the selection of a subset of achievable and consistent goals that they become subgoal as part of a refined conception of the process. Such desires are inputs to the agent's deliberation process, resulting in the selection of a subset of achievable and consistent goals. I assume for the sake of simplicity that desires are mutually consistent but not necessarily achievable and that it is such sets of desires that serve as motivating 'pro-attitude' toward acting.

⁹See also note 3, page 41

Intentions

Davidson's view that actions are reducible to a combination of beliefs and desires is rejected by Michael Bratman. He argues for an expansion of the belief-desire model to a belief-desire-intention model [Bratman, 1988, 1987]. In his approach, both desires and intentions play the role of 'pro-attitudes' that motivate action, where intentions play the part of the rational motivator. We already became familiar with the notion of intention in section 2.2.1. We should now refine the concept of intentions with respect to planning. What makes Bratman's considerations on intention so fruitful as starting point for formalizations of intentions is his claim that "the conception of intention is inextricably tied to the phenomena of plans and planning." [Bratman, 1987, p. 2]. From such a point of view, intentions are distinguished from desires by their motivational role; intentions are distinct from desires in that they are 'conduct-controlling pro-attitudes' whereas desires are 'potential influencers of action' [Bratman, 1987]. This means that intentions exhibit a volitional commitment to consequent action that desires lack, intentions are choice with commitment [Cohen and Levesque, 1991]. The point about commitment is that intentions have a characteristic stability in that they resist reconsideration without the addition of new information. The unifying notion of commitment as involving both conduct control and stability fits the needs of cooperation stated above, since it adds a notion of reliance that makes it possible to coordinate personal and intrapersonal interactions. For the further course of argumentation, it should be kept in mind, that this analysis of intentions considers intentions as actions. Consequently, intentions should be analyzed as events rather than states.

4.1.3 Summary and Outlook

Given the aim of this dissertation the treatment of mental attitudes needs further elaboration in two directions. First, it is necessary to spell out how desires, beliefs and intentions contribute to the rational conception and execution of complex goal-directed actions. This elaboration must capture the ways in which an agent rationally reacts to the information that reach her from the outside world or result from her internal reasoning processes. This can be said to be the 'internal' use of mental attitudes. The second direction has to do with how agents represent attitudes that they attribute to other agents with which they interact, and how they arrive at such representations. Elaborating the treatment of this second direction requires the combination of belief, desire and intention operators, which turn attitude representations into the representation of beliefs, desires and intentions that the agent attributes to others (as distinct from having those attitudes herself). The representations that result from applying those operators thus reflect an 'external' perspective on the attitudes represented by the operand representations (those to which the operators are being applied). This external perspective was already present in the discussion of explanation in chapters 2 and 3. I will proceed in the following manner: first, I will outline the architecture of the agent's rational control and I will then turn to the topic of the representation of attitudes from an external perspective.

4.2 The skeleton of active control

This section spells out an operational definition of beliefs, desires and intentions and their interaction in the motivation and control of an agent's action.

4.2.1 The main control cycle

A simple control loop

It was mentioned in the introduction that the agent layer is considered as a runtime environment that manages the agent's mental and real operations. The paradigm of such a control architecture is the simple loop given in definition 12.

Definition 12 *The paradigm control loop*

do

Sense();

Think();

Act();

until quit.

The BDI-interpreter

What follows introduces a more advanced and fine-grained control algorithm than that which is given in definition 12, the belief-desire-intention (BDI) interpreter. The architecture I adopt for presentation is the Procedural Reasoning System (PRS, [Georgeff and Lansky, 1987, Ingrand et al., 1996, Inverno et al., 2004], loosely following the simplified version given in [Singh et al., 1999]). PRS is probably one of the most popular real-time control architectures for autonomous robots and agents.

The main loop The crucial demand on the agent's control architecture is that it should be able to respond to both the evolution of reality and internal states while realizing intentions under the pressure of limited resources. That is, it must handle reaction *and* action while taking into account the dynamics of the agent's environment and her own internal states. Put in a condensed way that will be elaborated in full detail next, the BDI-interpreter deals with these demands as follows:

- Observations of internal and external events (including e.g. EPS feedback or the adoption of goals) are pushed to a storage queue
- In the next step, from this storage queue the interpreter generates options of (re-)action, depending on whether some of the events in the queue can trigger plans from the agent's plan library.
- All triggered plans are collected and sent to the deliberation function
- Deliberation chooses one of the options delivered by the generator and pushes it to an intention stack, either an existing intention stack if the option is a plans subgoal, or a new one if not.
- An updated intention stack then consists of the agent's intended means (a plan or atomic action) to react to a given situation or to realize a desired goal.
- If the next element in the currently active intention stack is a goal, it is pushed to the event queue, if it is an atomic action, this action is executed.

- The final step consists in updating the agent's presentation and representation and the evaluation of current intentions and new events
- With the updated picture of reality and internal states, the next cycle is executed.

As I am interested in a proof-of-concept rather than in detailed implementation issues, definition 13 states the outlined algorithm in terms of a pseudo programming language that can be translated to a real programming language when necessary. Similar considerations hold for the algorithm itself which can be fleshed out to a full planning system such as the distributed multi agent reasoning system [Inverno et al., 2004, dMars] if necessary.

Definition 13 *The BDI-interpreter main loop*

```

1 initialize-state;
do
2 options:=option-generator(trigger-queue,Beliefs,Goals,Intentions);
3 b-add(options, EPS(now));
4 update(IRS);
5 selected-options:=deliberate(options,Beliefs,Goals,Intentions);
6 update-intentions(selected-options,Intentions);
7 execute(Intentions);
8 get-new-SMS;
9 if f-update(SMS) then f-add(SMS,EPS(now));
10 update(IRS);
11 get-new-triggers(EPS,IRS);
12 drop-successful-attitudes(Beliefs,Goals,Intentions);
13 drop-impossible-attitudes(Beliefs,Goals,Intentions);
until quit.
```

As can be seen from definition 13, the main cycle of the BDI-interpreter involves several subroutines which in turn work on the agent's beliefs, intentions, desires as well as on the representational architecture of SMS, EPS and IRS. The functioning of the BDI-interpreter main loop, involved procedures and information structures plays a central role throughout the presentation of GDRT in this thesis. Thus the detailed discussion of the BDI-interpreter stretches across the remainder of this thesis and I give an overview of the places in which I discuss the BDI-interpreter in note 4 below.

Note 4 *Discussions related to the BDI-interpreter main loop*

- (1) Line 1 of the main loop is discussed in section 7.1.4, page 147.
- (2) Line 2 is discussed in section 4.2.3, page 66.
- (3) Line 3 is discussed in section 6.2.5, page 116.

- (4) Line 4 is discussed in detail in section 6.3.2, page 134.
- (5) Lines 5, 6 and 7 are discussed in section 4.2.3, page 68.
- (6) Lines 8 and 9 are discussed in sections 6.2.5, page 116 and 4.2.3, page 66.
- (7) Line 10 is discussed in section 5.4, page 88 and 6.3.2, page 134.
- (8) Line 11 are discussed in 4.2.3, page 67.
- (9) Lines 12 and 13 are discussed in section 4.2.3, page 68.

The BDI-interpreter operates on data structures. The data structures employed by the BDI-interpreter are listed in note 5 and introduced in more detail in section 4.2.3.

Note 5 *Data structures of the BDI-interpreter*

- *lists: linear sequences of linked elements, elements can be directly accessed*
- *queues: elements can only be accessed in the order they have been added (“first in first out”)*
- *stacks: elements can only be accessed in the reverse order they have been added (“last in first out”)*

4.2.2 Plans

The core notion of the agent’s control architecture is that of a plan (or ‘knowledge area’ [Ingrand et al., 1996]). In fact, the interpreter loop given above in definition 13 can be seen as an abstract plan for appropriate reaction to and action on events observed by the agent. This high-level plan then accesses the agent’s library of more detailed plans matching the respective events. In general, plans can be considered as compact representations of beliefs about means, options and goals, i.e. a plan can be identified with the belief that its execution will bring about the intended feedback.

Components of a plan

The above interpreter (definition 13) contains the bits of information which we need in addition to the results of section 4.1.1 (partial hierarchical structure, consistency and coherence of intentions) to give a detailed specification of the concept of a plan, as it will be needed in our further developments.

The option generator requires plans to be annotated with invocation and context conditions - stating the states of affairs under which a certain plan is triggered and a set of preconditions that specify the conditions that must hold in order to execute the plan. In addition the plan should come with a (re)presentation of the feedback that should result from the execution of the plan (where the feedback constitutes the plan’s goal). Finally, the core of the plan consists of a body specifying the course of atomic and complex actions that must be performed to achieve the intended feedback. The plan body has a tree-like structure similar to the tree-like EPS structure presented in section 3.3.3. That is, plans are considered as trees (rooted, directed and acyclic graphs) with one distinguished start EPS and possibly multiple end EPSs interconnected by a sequence of atomic or complex actions. Put in computational jargon, plans (resp. scripts or frames [Abelson, 1981, Minsky, 1972, Schank and Abelson, 1977]) are

declarative procedure specifications corresponding to complex modal formulas which I represent as stated in definition 14. Plans can set 'flags' during their execution to influence the triggering and processing of other plans. Flags set during the execution of a plan are collected in a set `set-flags`.

Definition 14 *Plan Schema*

A plan consists of the following parts:

Type:	The name of the plan
Invocation:	The triggering conditions for the execution of this plan. Acquisition of a new goal (active invocation) or a change in beliefs or representations (reactive invocation)
Context:	The presuppositions that must be satisfied for the proper launching of the plan
Feedback:	The EPS or IRS which holds after the performance of the plan
Body:	The specification of the plan in terms of the tree-like structure of actions and states of affairs. A successful branch of the plan is marked with a final leaf $END(+)$ an error branch of the plan is marked with $END(-)$

Metaplans

Metaplans extend the agent's reasoning capabilities to manipulate beliefs, desires and intentions. Typical scenarios for the invocation of metaplanning are e.g. the choice between several applicable plans, the modification and manipulation of intentions or the computation of available resources given the constraints of real-time processing. As has been adumbrated above, we can also consider the main control loop as a toplevel metaplan that guides the execution of all other planning. Metaplans have the same structure as 'simple' plans but their invocation conditions involve plans. They are handled by means of the `deliberate` procedure spelled out in detail in the next section 4.2.3. Plans and metaplans form the agent's plan library, a database containing her latent and previously activated abilities. It should be noted that I do not make extensive use of metaplanning in the analysis of examples in chapter 7.

4.2.3 System states

The structure of the main control loop requires the agent to store and manage information of different type and purpose: beliefs, observations, plans and goals as well as intentions. The combination and configuration of these sources of information constitutes the agent's system state which will be discussed now.

Belief States

Belief and option generation The fundamental information that nourishes the main control loop consists of observations of internal and external events and states of affairs. They provide the triggering conditions for the generation of options for (re)action. Following the analysis proposed in the previous sections such triggers stem from beliefs and facts encoded at the EPS level or represented as IRS (or both). In view of what has been said about the reliability of beliefs in section 4.1.2 the question arises

whether it is reasonable to feed all of an agent's actual beliefs to the option-generator: doing so would violate the constraints on plans imposed by the boundedness of resources as discussed in section 4.1.1. I restrict the generation of options to the factual subset of an agent's beliefs (definition 9), as the generation of options from beliefs would lead to a potentially infinite recursion of option generation on all of the agent's plans, where not only plans that are of actual importance are considered but also options of actions for which their preconditions are not factual or their goals do not realize the agent's desires.

Functions associated to beliefs and facts Again, we need to distinguish between facts and beliefs. The operators associated with the EPS presentation of beliefs and facts are the query operators **b-add**, **b-remove** to add and remove beliefs based on the agent's current set of options of actions. The **b-add** operation concerns the EPS structure, where it adds possible options of action generated by the agent's planning engine. In addition, the command **b-add** can also be executed in response to the processing of utterances by other agents. Once it is recognized that certain options have become irrelevant, the respective beliefs are removed by the operation **b-remove**. **f-add** adds facts to the EPS via a translation of incoming SMS data. Consequently, we have to define a lookup table that for each SMS expressions specifies how that expression is to be translated into a factual EPS expression. Each of these **f-add** operations must be counter-checked for compatibility with the the agent's existing set of EPS presented facts. If there exist incompatibilities between an existing and a new EPS conditions concerning an already registered thing (e.g. a move to a different location), the newer information replaces the old information. Similarly, new information (e.g. a thing appearing for the first time in the agent's area of vision) is translated into EPS expressions. These two types of information update are treated by the operation **f-add**, which is discussed in detail in section 6.2.5 and definition 38 on page 116 . In addition, I introduce an operation **f-update** that decides whether a given SMS expression should effect a change in the EPS - only if there is an update concerning existing or new objects, this information is added as a new fact. Recall that each EPS is tagged with a unique time index so that an EPS structure can be interpreted as a fact history.

For the IRS, I will define a function **update(IRS)** that checks for the availability of (sets of) EPS expressions that render possible the construction of IRS-individuals. This function will be spelled out in detail in section 5.4, page 88 of this thesis.

The Event Queue

The event queue is the storage for potential planning triggers: it contains observations about changes in reality or in the agent's system states that defines a potential room of action. Each time the agent makes a new observation¹⁰, this observation is pushed onto the event queue for further processing. It is noteworthy that these observations not only include beliefs or facts but also the addition, removal or manipulation of goals and intentions. The operation **g-add** pushes a goal or plan to the event queue, **g-remove** removes goals and plans from the event queue once they are either fulfilled or have been recognized as unrealizable. There are two types of goals; besides goals to bring about a certain state of affairs I also include goals that have to do with the agent's own information management. These goals are to test whether a given condition is provable from some set of other conditions. I denote such goals

¹⁰Please note that the conception of observation I use here is one of unconscious perception, mirroring the delivery of data from e.g. the eye to the brain.

as `provable(K,database)` and `?variable=value`. Once the event queue is filled with the agent's observations of internal and external events, it serves as input to the option generator.

Option generation Given the agent's event queue, the option generator iterates on the queued items using them to select plans from the agent's plan library. This is done by simple unification of invocation conditions with queued facts or goals. All the plans triggered by items on the event queue are collected together with the plans in the event queue and make up a set of the agent's options for action given the actual circumstances. At this point an important relation to the discussion on EPS structures can be established. Given the set of generated options of action, this set of actions is to be considered as the agent's personal draft of her future. With respect to her EPS structure, this set of future options of action presents possible evolutions of reality that are under the control of the agent¹¹. The point I want to make here is that the future of the agent's EPS is to be considered as reflecting the option graph generated by the BDI-interpreter. That is, each item in the set of possible options of future action makes up a substructure of the future EPS branching time structure, then presenting a belief in possible future states of affairs. It is important to keep this point in mind when we spell out the formal definitions of the EPS structure in section 6.2. The main loop of the BDI-interpreter captures the relation between generated options and the EPS. This relation is established by the call of the function `b-add(options,EPS(now));`. Discussed in detail in section 6.2.5 on page 119, this function adds the current set of options to the EPS. Definition 15 spells out the generation of options.

Definition 15 *option-generator(trigger-events)*

```
options:={};
for trigger ∈ triggers do
  for plan ∈ plan-library do
    if matches(invocation(plan), trigger) then
      if provable(context(plan), facts ∪ IRS ∪ set-flags) then
        options:=options ∪ { plan};
return(options)
```

It is obvious that the draft of the future generated by all plans matching the actual state of affairs will contain numerous options in any realistic setup. Thus the next step in the process needs to reduce the set of options to one plan, which is adopted as intention - this is the task of the deliberation procedure, which is introduced next.

Options and Intentions

Strategies to cope with real-time demands The deliberation function is responsible for the choice of a plan from the generated option graph and subsequent commitment. It is in particular this part of the architecture which has to face the demands of real-time processing. Lengthy deliberation can result in exceeding the limits of real-time response. Thus we have to design the deliberation procedure in a way that keeps the number of options under control. The process of decision-making for one or the other

¹¹Of course, the agent's EPS future has to be enriched with explanations of other temporal variations occurring in her environment. This source of options is introduced in the next section 5.3 on descriptions

option can be steered by two principles: metaplanning¹² or random choice. As the examples I discuss in the second part of this thesis do not contain situations of several concurrent plans resulting from option generation, the following definition of the deliberation function serves mainly illustrative purposes.

Definition 16 *deliberate(options)*

```

if length(options) ≤ 1 then return(options);
else metalevel-options:=option-generator(b-add(option-set(options)));
    selected-options:=deliberate(metalevel-options);
    if null(selected-options) then
        return(random-choice(options));
    else return(selected-options).

```

Intentions Once deliberation has picked up a plan, it sends this plan back to the BDI-interpreter. At this point, the characteristic commitment associated with intentions (see section 4.1.2) comes into play, in those cases where the option generator is modified with the introduction of a function *post-intention-status()*; at the end of the procedure as specified in definition 18 which prefers the realization of existing intentions over other available desires. The crucial part is the last line in definition 18. This line forces working down existing intentions by executing the topmost subplan or atomic action of the current intention stack. If the topmost item on the intention stack is a subplan, the execution of this intention recursively pushes the subplan to the event queue until an atomic action is the topmost item on the intention stack.

Definition 17 *option-generator(trigger-events)*

(replacing definition 15)

```

options:={};
for trigger ∈ triggers do
    for plan ∈ plan-library do
        if matches(invocation(plan), trigger) then
            if provable(precondition(plan), facts) then
                options:=options ∪ { plan};
post-intention-status();
return(options).

```

Definition 18 *post-intention-status()*

```

if null(Intentions) then
    for goal ∈ Goals do
        event-queue:=event-queue ∪ g-add(goal);
else for stack ∈ Intentions do

```

¹²Two main such strategies on option filtering are implemented in PRS: the hard-wired compatibility filter and the computationally more complex override filter ([Georgeff and Ingrand, 1989, Bratman, 1988]). The former type of option filtering is about checking whether the plan in question is consistent with the other plans up for execution. The override filter, typically triggered by internal conditions such as new facts or beliefs checks whether parts or the whole of plan should be suspended given the actual situation, e.g. with respect to new opportunities. From such a point of view, filters implement a certain kind of self-reflection [Ingrand and Georgeff, 1990].

```
event-queue := event-queue ∪ g-add(item-on-top-of-stack);
```

An intention thus consists of an initial plan and subordinated sub-plans. The option returned to the main loop by the deliberation function is pushed to the intention stack *Intentions*, a partially ordered structure with certain minimal elements (the roots) that contains all tasks that are chosen for execution at the current or some later time. Processing an intention stack has to obey its ordering: intentions earlier in the stack must be dropped or realized before later intentions can be executed [Ingrand and Georgeff, 1990]. Of course, metalevel plans can cause reorderings of the intention stack with regard to priorities or other relations between intentions. The next step of the main loop executes the topmost action from the intention stack, i.e. it either performs an atomic action, posts a new subgoal to the event queue or acquires new beliefs. Often, this execution results in an update of the SMS and IRS, which in turn may trigger the dropping of impossible or successful intentions. With this updated picture of reality, the loop starts again. If there is no new information in the event queue, existing intentions will be executed. This captures the resistance against reconsidering intentions once they have been formed.

Failure and reactivation A last observation on failure and reactivation of intentions. If no bounds are placed on reactivation, i.e. unlimited retrying is allowed, this will result in failure of the agent to respond or to irrational behavior. On the other hand, the impossibility of reaching a certain goal is hard to prove - thus a practical solution should allow for a limited number of attempts and if these fail, restart planning after eliminating the current option which has eluded successful realization. Note that this does not apply to intentions which are waiting for certain preconditions to become realized which are necessary for their execution.

4.2.4 The agent layer as explanatory mirror structure

The present discussion of the agent layer considered only the use of the BDI-interpreter as a mechanism for planning and controlling of the agent's own actions. This view must now be extended to include explanations of the actions of other agents. Given the discussion of the explanation of temporal variation in section 2.1.2, the recent discovery of mirror neurons in the brain structure of apes and humans [Rizzolatti and Fogassi, 1996] suggests that the same mental processes are activated both in controlling one's own behavior and in explaining the behavior of others. Roughly speaking, mirror neurons are activated either by the execution of one's actions or by the observation of such actions being performed by other agents. That is, the control structures for decision-making and action planning can be activated by either endogenous or exogenous triggers, with the difference that in case of explaining external actions, the control structures for planning are decoupled from sensorimotor control. This amounts to a kind of *simulation* of the other agent's behavior. For the setup proposed in this thesis the functional similarity between planning and explanation is captured by making the BDI-interpreter not only responsible for the planning and control of one's own actions but also for the simulation and consequent explanation of other agents' actions. While technically the simulational use of the BDI-interpreter poses no problems, the use of the BDI-interpreter to explain other agents' actions has to consider the problem of the third person (section 2.1.2): internal states of external agents cannot be directly accessed. Thus the explanation of

other agents' actions must start from one's own plan library. Given these considerations, I propose the following processing of the actions of other agents:

- When another agent enters the scene, a new SMS-decoupled instance of the BDI-interpreter is started. For this new instance, planning is decoupled from the execution of actions; to this end, a variable `mental-mode` is introduced which is set to `true`.
- If the EPS presents that the other agent exhibits perceivable temporal variation, the given circumstances are added to the event queue of the other agents' instance of the BDI-interpreter. E.g. if an agent grasps a cube and a screw, then it should be assumed on the basis of the availability of a plan for putting together a screw and a cube together that she wants to screw the cube and the screw together, thus the explaining agent should add this desire to the event-queue of the observed agent.
- Besides this implicit treatment of another agent's actions, there also exists the explicit option via the communication of goals and intentions. That is, if another agent announces that she wants to put a cube and a screw together, the interpretation of the IRS corresponding to this utterance should trigger the addition of the other agent's intention to screw the cube with the screw to her instance of the BDI-interpreter.
- Adding new goals and intentions to an agent's representations of the mutual state of another agent should be restricted to situations in which the first agent has good evidence that the second agent has these goals or intentions - e.g. in a situation where the other agent tells the first agent that she has these goals or intentions.

It should be noted that there are no ontological restrictions on the type of agent that can trigger a new instance of the BDI-interpreter. That is, a cube rolling down a sloping table is explained with the help of simulation too, but with causal forces playing the part of intentions. The actual use of simulational instances of the BDI-interpreter is discussed in the next chapter 5.

4.2.5 Know-How

Given the discussion on the structure of the agent's planning engine above, I introduce a representation of another kind of epistemic attitude besides beliefs and cases of know-that: know-how. To know how to do a certain thing states that the agent in question has the ability to bring about the 'certain thing'.

Definition 19 *Operator for know-how*

If K is an IRS and x a thing-individual, then the following expression is an IRS-condition:

- xK_hK (x knows how to do K)

4.2.6 Summary and Outlook

The last pages have introduced the agent's dynamic control architecture, constituting the operational part of the agent layer. The BDI-interpreter generates the agent's beliefs about possibilities which together with the factual input of the SMS make up the agent's EPS. Two tasks remain with respect

to this runtime environment. First, the agent must have the ability to describe her own internal states, i.e. her intentions, facts and beliefs and desires as well as changes in other objects she observes in her environment. This enterprise is tackled in the next section in the light of what we established in the discussion of explanations in section 2.1.2. Second, and this is what will make up most of the work of this part of this thesis and makes good on the promises made in the introduction, I need to integrate the processing of goal-directed multimodal communication into the given framework.

Chapter 5

On the way to discourse processing

So far I have only discussed the internal configuration of an agent with respect to perception, presentation and representation. We must now take into consideration how internal states of the agent provide a background against which she realizes her goals and intentions and how other agents in her environment can make sense of her behavior. The interpretation of external manifestations of an agent's internal states is the core concept behind the processing of goal-directed interactions as it is analyzed in this thesis. That is, it will be necessary to capture the external processes going on at the 'surface' of an interaction (such as utterances or bodily actions) as well as the internal mental 'background' of the participants involved. I will use the term 'discourse' for such interactions. 'Discourse' in the sense intended here has a wide coverage. It not only includes verbal discourse - the exchange of utterances - but also non-verbal interactions, in which any kinds of actions by any kinds of agents can serve as signs of that agent's beliefs, goals and intentions to some other agent.

5.1 The notion of discourse

A notion central to this dissertation is that of a 'discourse'. This section provides a fundamental discussion of how this term has to be understood. Such a discussion must elaborate on the surface structures of an interaction and how they relate to the internal background of beliefs, facts, intentions and desires of the agents involved. The notion of discourse I will develop in the following must therefore satisfy the following requirements.

- A discourse can manifest itself as a mental process going on inside in an agent, a series of interactions between an agent and her environment as well as a series of interactions between several agents.
- A discourse can include different types of actions such as bodily actions, thoughts and speech actions.
- A discourse is no arbitrary sequence of actions but is structured by the principle of rational goal realization.

In a second step, I need to show how the surface structure of the actions of one or more agents relates to internal states of the agents. Thus I have to discuss how internal mechanisms of representation manipulation, perception and planning justify, motivate and control an agent's observable behavior.

To make a start, consider the following example dialog 4.

Example 4 *Example dialog for teaching mode.*

	<i>The table holds a cube, a slat and a screw. The user approaches the table and stops in front of the robot. The robot looks at the user. The robot does not know the user's name.</i>
A:	Hi, my name is Clara. What is your name?
B:	My name is Tillmann.
A:	I am going to build a corner cube bolting. Do you know how to do that?
B:	No, I don't know how to do that.
A:	I will explain it to you.
A:	A corner cube bolting is a cube <i>A points to the cube</i>
A:	screwed to the corner hole of a slat <i>A points to the slat.</i>
A:	with a screw. <i>A points to the screw</i> <i>A grasps the screw.</i> <i>A grasps the cube</i>
A:	I need your assistance to screw the slat to the cube. Please hold the slat between the screw and the cube. <i>B grasps the slat and holds it between the cube and the screw. The robot screws the cube to the slat.</i>
B:	OK.
A:	Do you know how to build a corner cube bolting now?
B:	Yes.
A:	OK. That's it. Thank you.

Example 4 is typical in that the interaction between the user and the robot consists of external actions that occur in parallel and correlation with internal mental processes going on inside the two agents. A concept of discourse that does justice to this interaction of internal and external processes should not be limited to either internal or external processes¹. In other words: a discourse consists of overt cases of interaction that are observable by all participants of the discourse as well as mental processes going on inside the participants which are not observable by the other participants. I consider both external interactions and mental processes as consisting of actions and a first approach to the notion of discourse can be stated as in note 6 which can then be refined either by specifying the means of discourse (thoughts, language, gesture, actions), the structure of discourse (in particular means-end rationality and related subconcepts), the type of environment involved (objects, other persons, the self) or the location of discourse (internal, external, mixed).

Note 6 *Discourse*

A discourse is a set of actions directed towards an environment.

¹In this respect, my conception of a discourse goes beyond that of linguistics as I do not limit a discourse to communicative actions.

5.2 Discourse elements

A discourse is not just an arbitrary sequence of actions. Its identity is determined by the goal that it serves to identify and realize. That is, a discourse can be considered a complex action that serves the realization of an intention and thus as corresponding to the execution of a plan. As we saw in the discussion of actions and plans in section 4.1.1, page 53, complex actions resp. plans have a fine-grained substructure of atomic elements. The next sections discuss the atomic and non-atomic elements of discourse that play a part later on in this thesis from both the internal agent's point of view who performs actions and the external point of view of an agent who tries to explain actions performed by other agents. Finally, I will elaborate the goal-directed nature of a discourse.

5.2.1 Internal perspectives on action

From an internal agent's point of view, actions can be classified in several ways. I adopt the basic distinction between internal and external actions and that between non-communicative and communicative actions.

Internal action and external action

A basic distinction within the domain of actions is that drawn between external and internal actions. Internal actions refer to mental processes going on inside an agent and which cannot be directly monitored by an external observer. In the framework of this thesis, internal actions are primarily concerned with the construction and manipulation of representations (IRSs), presentations (EPSs) and with the processes of the BDI-interpreter. In contrast, external actions are those that can be perceived by external observers and involve the control and manipulation of sensorimotor processes. At the level of the EPS, I label an internal action a with the prefix $int:$, an external action a with the prefix $ext:$ and observed action a of an other agent d with $d:$.

Definition 20 *Presentation of actions*

From the internal perspective of an agent x , an action a is presented in the EPS as

- *int-a: iff it is an internal action of x*
- *ext-a: iff it is an external action of x*
- *d-a: iff it is an observed external action of an other agent d*
- *d-a:UTT,K iff it is an utterance UTT of an other agent d with an IRS K constructed from K .*

Sensomotoric classification of action

From an agent's internal point of view, the typology of actions in definition 20 is reflected by their sensorimotor realization:

- Bodily actions (External)
- Cognitive actions (Internal)

- Speech actions (External, Internal)

I will now elaborate on these three classes of action.

5.2.2 Bodily actions

The most basic type of action an agent can perform is that of a physical movement of her body. The role of bodily actions is both fundamental and versatile. Their versatility is illustrated by the fact that they can be either performed in the service of a communicative or non-communicative intention but also for no higher purpose whatever. Consequently, communication via bodily actions is a mixed blessing. On the one hand, bodily communication is very intuitive and immediate. On the other hand, it is difficult to transmit a communicative intention unambiguously by means of a bodily action. A simple movement such as nodding can be used to express a communicative intention but noddings don't do this unequivocally, since they can also be performed in response to some bodily need.

5.2.3 Cognitive actions

Cognitive actions (e.g. actions of planning or belief manipulation) have been discussed in great detail in the previous chapters of this thesis and they will be discussed in greater detail below - internal actions will be defined in section 6.2. The BDI-interpreter and associated functions were already discussed in section 4.2. Interpretation procedures for IRSs will be defined in section 6.3. The reader should keep this in mind while reading the next section on speech actions.

5.2.4 Speech actions

Speech actions involve bodily and cognitive activities. The utterer of a speech action intentionally puts forward to a hearer a semantic representation of a proposition and a pragmatic pressure for (re-)action via the intentional bodily production of speech and other actions. This special function of utterances will be investigated in more and more detail in the subsequent parts of this thesis. A classification of actions which is useful for the analysis of speech actions involves the distinction between non-communicative and communicative actions. Non-communicative actions are actions which do not involve a social component, i.e. non-communicative actions are instrumental in the sense that they do not result from a communicative intention. They are directed toward bringing about a change in objects, such as 'moving a cube', whereas communicative actions involve a social component and necessarily require actions on the part of other agents in order to succeed, as they are directed toward bringing about changes in other agents' representations.

5.2.5 External perspectives on action

From the external perspective on an agent's actions, the distinctions that can be drawn from an internal perspective, i.e. those between internal and external actions and communicative and non-communicative actions are problematic. Each theory of action has to face the problem that from the external viewpoint of the observer of an agent it isn't possible to say for sure whether the action performed by the agent belongs to one class or the other. Many communicative actions are *hybrid* in that they also serve a

non-communicative goal². Consequently, such actions can be classified both as communicative and as non-communicative and can thus be said to be *ambiguous* from the viewpoint of external observation. A similar difficulty arises with respect to the distinction between internal and external action which is unambiguously available only from an internal perspective. From an external perspective, internal and external actions are inseparably tied together, as the interpretation of external actions depends on the assumption of an internal background and the interpretation of internal actions requires an external manifestation from which one can conclude the internal background.

5.2.6 The structure of discourse

Now that I have introduced the actions sequences of which potentially make up a discourse, I should discuss what exactly it is that distinguishes a discourse from an arbitrary sequence of actions. That is, I have to discuss what exactly constitutes the characteristic structure of a sequence of actions such that it makes sense to use a distinguished concept of discourse. Much of what I have to say on that topic has been mentioned along the way, thus clarifying that what exactly makes up a discourse is about connecting the various lines of argumentation concerned with goal realization, planning, rationality, actions and thoughts.

Means-end rationality

Probably the most intuitive property of a discourse (in the sense of this thesis) is that its elements are sequenced so as to realize a certain goal. In turn this implies that a discourse is arranged according to some underlying plan. That is, a discourse can be considered as a surface realization of the mental processes of the participants. Considering what has been said on the structure of planning, it stands to reason that the structural principles of planning are also reflected at the surface of discursive realization. That is, a discourse realization inherits the coherence and consistency of its underlying plans, representations and other mental processes. The other way round, the use of incoherent representations or inconsistent plans that guide the execution of discourse steps lead to an incoherent or inconsistent surface structure of the discourse that will most probably fail to realize the intended goal.

Refined discourse conception

I can thus refine the conception of a discourse as follows:

Note 7 *Discourse*

A discourse is a means-end rational set of actions performed by one or more agents, the surface realization of an agent's plan, directed towards an environment.

5.3 Representing time-individuals

While the representation of thing-individuals has been introduced in section 3.2.2, the discussion on time-individuals (section 3.3) left open what exactly their representation looks like. This section introduces

²To a certain degree, this problem will probably also occur for an agent's internal perspective on her actions. However, we would require a rational agent to be in principle able to separate communicative and non-communicative aspects of an action.

representations of time-individuals in the IRS³. On account of the specific needs of this thesis, the representation of time-individuals differs from the 'traditional' logical treatment of temporal reference in many respects. To make this clear I elaborate briefly on the traditional approach to temporal entities before I introduce the representation of time-individuals in the framework of the IRS and the EPS.

5.3.1 The traditional approach

Davidson

The classical account of time-individuals (here: events) goes back to Donald Davidson's logical analysis of action sentences [Davidson, 1967], which proposes to capture the logic of natural language sentences containing action verbs by analyzing them as descriptions of events, thus acknowledging events as bona fide elements of ontology. According to this view, events are supposed to be "entities in the world with their own observer-independent grounds of existence" [Kamp, 2007]. The following example illustrates Davidson's approach to the logical form of verb phrases involving action verbs.

- (1) build a house: $\exists e.\exists x.\exists y.agent(x) \wedge house(y) \wedge build(e, x, y)$

Vendler

While the Davidsonian analysis of reference to temporal entities seems to be acceptable at a first glance, important information contained in the predicate 'build a house' is not represented in the logical forms he proposes. First of all, this concerns the observation of [Vendler, 1957], who noticed that different verbs can have very different 'temporal profiles' in that they are used to describe very different *event complexes*. Vendler proposed a classification of the event complexes described by different types of verbs, and of the verbs that describe them. Vendler's classification of verbs into achievement, accomplishment, activity and state verbs was the first of several such classification proposals in the last century, but it has held up remarkably well and is still frequently used. According to Vendler's classification, 'build' as in 'build a house' is an accomplishment verb. What is distinctive of a verb phrase like 'build a house' (and likewise of other verb phrases containing accomplishment verbs) is that the process of construction (i.e. the building) brings about the house and that it is this result - the completed house - that "casts its shadow backward" [Vendler, 1957, p. 146] in that it actually identifies the activities leading to the result as the building of a house. To do justice to the aspect of the meaning of accomplishment verbs we need at a minimum to distinguish between different parts of the event complexes a verb phrase like 'build a house' can be used to describe, viz. the 'culmination' of the house building process, the activity that leads up to this culmination and the 'result state' - the state of affairs that results from the process and starts at the culmination time. This event complex, consisting of 'preparatory phase', 'culmination' and 'result state' is what [Moens and Steedman, 1988] call 'the event nucleus'. Moens and Steedman showed that it is characteristic of accomplishment verbs that they are used to describe complete instantiations of the event nucleus. Other Vendlerian verb types are used typically to describe parts of the event

³This is probably the right time to come back to the question what the purpose of representations is. This question has been touched upon in previous chapters in the context of reductionism as well as dynamic predicate logic which deny the need for representations. But a serious theory of discursive interaction must include overt intrapersonal tasks (such as communication) which require an agent to be fully and explicitly aware (by entertaining logically sound representations) of internal and external states and processes she is involved in as well as internal mental processes of the other agents involved

nucleus. For instance, for Vendler’s achievement verbs the described event complex tends to consist just of culmination and result state; and similar characterizations are also possible for the remaining Vendler Classes.

DRT

A very simple-minded approach can combine Vendler, Davidson and the theory of event nucleus within the framework of Discourse Representation Theory [Kamp et al., 2007] as shown in the following example 5:

Example 5 (1) *build a house*⁴

x, y, e $e : build(x, y)$ $house(y)$
--

Meaning Postulate 1:

x, y, e $e : build(x, y)$ $house(y)$	⇒	s^{res} $e : build(x, y)$ $e)(s^{res}$ $s^{res} : house(y)$
--	---	--

Meaning Postulate 2:

x, y, e $e : build(x, y)$ $house(y)$	⇒	s^{prep} $e : build(x, y)$ $s^{prep} \subseteq e$ $s^{prep} : \neg house(y)$
--	---	---

Problems

While this way of representing the semantic information contained in the example comes closer to the intuition about what ‘build a house’ actually means, there are still important problems to be solved.

- First, the probably most obvious problem is associated with the adequate representation of the result of the event of building, i.e. that the house is supposed to come into existence *if* the process of building is properly finished. This is hard to capture within the standard framework of formal semantics because
 - The condition $s^{prep} : \neg house(y)$ in meaning postulate 2 does not capture the crucial point about a thing’s coming into existence. It is not the case that the referent y is no house but that y does not exist at all at this preparatory stage of building.
 - It is subject to doubt whether the existence of a house is really a logical consequence of the building and likewise a causal effect of the activities that make up the building.

⁴There exist of course more refined theories about creation verbs resp. accomplishments in DRT, e.g. [Kamp and Bende-Farkas, 2005] thus the representation pictured here certainly does wrong to the current state of the art in DRT and is only intended to serve illustrational purposes. Nevertheless, the arguments raised here apply to the revised treatment of creation verbs too.

Basically, the problems associated with 'incremental' themes [Krifka, 1998] have been tackled from two sides; (1) syntactically with the introduction of additional predicates for 'staged' existence, becoming and causation [Dowty, 1979] or the concept of thematic roles [Krifka, 1986] that specify the relation between the building and the house and (2) semantically with a non-monotonic formulation of implication [van Lambalgen and Hamm, 2004]. Both approaches have to face the fact that describing an ongoing action as 'build a house' neither logically implies nor causally forces the house to come into existence. Instead, I propose that *describing a currently* executed action as 'build a house' essentially involves reference to the intentions of the agent who performs the action in question and that it is this attitude of the agent toward the existence of the house that relates the currently ongoing activities of building to the intentionally projected existence of the house. In addition, it should be mentioned that the given preparatory and consequent states are not only distinctive for the building of a house, as there are other predicates that can describe the same constellation⁵.

- Second, the event nucleus theory relies on a notion of state as well as on the notion of an event which has to be established first.
- Third, a verbal phrase of a certain Vendler class occurring in a description can be shifted to another Vendler Class by changing the amount and type of information that specifies the temporal entity which is described. E.g. 'build a house' can be transformed to an activity if information is added as in 'build a house for an hour'. Consequently, the Vendler Classes are not distinct in the sense that there is a unique mapping between predicate and temporal profile. In turn, this makes it difficult to derive ontological claims about temporal individuals from natural language expressions.

To deal with these difficulties more adequately than earlier proposals it is necessary to improve on the model theory that [Kamp et al., 2007] give for a DRS-language containing representations such as in example 5. The clause in the truth definition of this model theory for the evaluation of the main DRS in example 5 is the one given in definition 21.

Definition 21 *Evaluation of DRS event conditions (simplified) [see Kamp et al., 2007, p. 115]*

Given a set of events and states EV structured by $<$, a Universe of individuals U and an Interpretation function I ,

- $g \models_M e : R(x_1, \dots, x_n)$ iff $\langle g(e), g(x_1), \dots, g(x_n) \rangle \in I(R)$

Where g is an assignment that maps e onto an element of EV and x_1, \dots, x_n onto elements of U .

But what does this tell us? All it tells us is that events described by occurrences of 'build a house' are events that stand in some 'build'-relation to the one who is doing the building (or the ones who are doing the building) and the thing that is built. When we try to improve on this model theory and to develop one which tells us more about the conditions that must be fulfilled in order that a situation can be described as one in which a house is built, then all the difficulties mentioned above confront us and need to be met head on. Moreover, an important further desideratum for such an improved theory is that our reasons for using a phrase like 'build a house' in a given situation is often motivated by our

⁵The use of pre- and post-states has to face a lot more problems, e.g. the amount of information contained in these states or their temporal extent.

ascribing to the agent(s) the *intention* of building a house. This is something about which the clause in definition 21 gives no information. Consequently, if we want to do better, we must review both the model theory and the representation of temporal individuals. In the following sections, I will present an account of temporal individuals that circumvents the problems of the proper treatment of temporal individuals we have just discussed.

5.3.2 Temporal entities in GDRT

Temporal individuals as mental entities

One respect in which traditional accounts of temporal reference fall short of explanatory adequacy was touched upon section 3.3.3, where it was pointed out that human beings structure their actual and projected (possible future) experiences in terms of the causal relations that they perceive between them and plan-goal structures or intentions on which they impose a certain 'explanatory' coherence [Zacks and Tversky, 2001]⁶. The conception of models which incorporate a notion of temporal succession grounded in causal connections actually accords well with one of the fundamental assumptions of DRT, namely that humans make use of mental representations (in particular when interpreting utterances). In fact, we can establish a natural relation between DRT and the psychological theory of event perception structures if we introduce intentions, plan-goal and causal structures as mental entities of representation. In addition, the dynamic backbone of the BDI-interpreter makes it possible to ground notions such as plans, intentions and causal structures in the agent's internal (re-)presentations. The central feature that connects the EPS, the BDI-interpreter and IRSs is that of anchoring, which has been introduced for thing-individuals, but is now to be extended to time-individuals. What should an anchor for a time-individual look like?

Temporal anchoring

In the following I seek to define anchoring mechanisms for time-individuals in the IRS that identify their reference in the EPS structure. Recall that an anchor consists of two parts - a floater and a source. For a floater representing a time-individual, its source must supply information on how to *identify* the temporal EPS structure in which the respective floater is anchored. But what should the information look like that makes such identification possible? This is where explanations of temporal variation come into play: the anchor source of a temporal individual specifies a temporal profile of the time-individual floater in terms of causal, plan-goal or intentional structures. The other way round, IRS representations of temporal entities prescind from the detailed structure of temporal variation by subsuming specific parts of the respective temporal variation as 'packages' of causal, plan-goal or intentional structures (the 'identification type')⁷ From the viewpoint of the linguistic surface, time-individuals represent predicates,

⁶It should be noted that while my approach to the treatment of temporal entities shares this starting point with [van Lambalgen and Hamm, 2004] the way I proceed with this psychological insight is fundamentally different. Hamm and Lambalgen develop their formalism along a strictly physical understanding of temporal variation, whereas my proposal allows for other types of segmentations such as behavior or intentions. In addition, the Hamm and Lambalgen propose a proof-theoretic treatment of planning whereas my approach involves a possible-worlds semantics.

⁷Recall the terminology we have adopted: The term temporal variation as I understand it refers to an uninterpreted, unsegmented sequence of action from which no temporal entities such as events have been extracted. Segmentation then establishes temporal profiles, i.e. packages structures of temporal variation.

where the predicate's argument structure is part of the anchor source that identifies the time-individual in the EPS. Consider the following representation of eventive IRS time-individuals.

Note 8 *IRS representation of eventive temporal entities.*

If *handle* is a handle for an IRS time-individual, *K* an IRS, *e* a temporal floater for an event and *x* a thing-individual floater representing the agent of the action, then

- $\langle e, xCAUSEK \rangle$
handle(*e*)
- $\langle e, xDOK \rangle$
handle(*e*)
- $\langle e, xINTK \rangle$
handle(*e*)

are IRS-conditions.

The representation of time-individuals fits into the proposed structure of an agent's internal states as pictured in the example of figure 5.1.

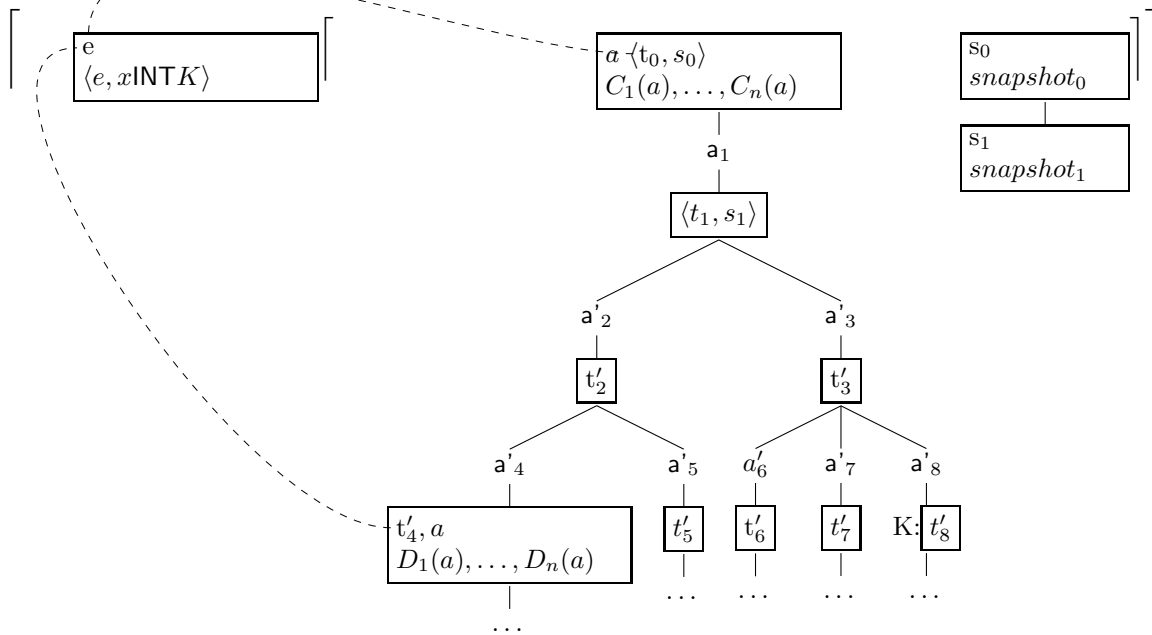


Figure 5.1: Integration of the representation of time-individuals into the proposed layer structure. This example represents an agent *x* intending to bring about a state of affairs specified by *K*. The INT-Operator states that the agent *x* has the intention to bring about *K*, i.e. that the path leading to *K* is among *x*'s intentions.

The semantics and pragmatics of time-individuals

The information contained in representations as pictured in note 8 does not suffice to identify the temporal profile of a time-individual. It is not enough to specify the type of explanation which is used to identify

the temporal profile in question but we also need more detailed information on what makes the temporal profile in question a distinct entity - we need to associate a plan with the time-individual in question. Plans are not part of the IRS but of the EPS. Thus a specification of a temporal individual must consist of two parts: a semantic IRS part and a pragmatic EPS part. It is the handle of the time individual that allows to associate IRS representations and their temporal EPS profiles. E.g. the predicate 'x build K', where x is the agent and K an IRS representing the goal of building associates with both an IRS representation as in note 8 and an EPS structure in the form of a plan for building K. A time-individual of type event thus consists of a semantic representation and a pragmatic profile in terms of EPS-structures as given in note 9.

Note 9 *The semantics and pragmatics of eventive time-individuals*

A handle *handle* for an event consists of a semantic part (*SEM*) and pragmatic part (*PRG*), where *OP* is one of the operators *CAUSE, DO, INT*:

$$SEM \left[\begin{array}{l} \langle e, xOPK \rangle \\ handle(e) \end{array} \right],$$

PRG specifies the identification conditions for *e* in terms of

- an EPS path iff *OP* = *CAUSE*
- an EPS subtree with or without intentional commitment otherwise.

The second type of a time-individual is that of a state. In the light of the discussion of states in section 3.3.3, we should distinguish two types of states. First, the basic conception of an 'instant' state as a thing having a certain property as shown with a given EPS. Second, the concept of a state as an idealized⁸ extension of the first notion of instant state. Let me start with instant states.

Note 10 *IRS representation of stative temporal individuals. If handle is a handle for a stative temporal entity, K an IRS, s a floater for a state then*

$$\bullet \left[\begin{array}{l} \langle s, K \rangle \\ handle(s) \end{array} \right]$$

is an IRS-condition.

Note 11 *The semantics and pragmatics of stative time-individuals*

A handle *handle* for a state consists of a semantic (*SEM*) and pragmatic (*PRG*) part:

$$SEM \left[\begin{array}{l} \langle s, K \rangle \\ handle(s) \end{array} \right],$$

PRG An EPS specification K_1 of the IRS *K*.

Continuant states require additional specifications on the temporal duration of the state, e.g. as in "the cube was red for an hour". Such additional constraints on the duration of states could be triggered by the lexical entry for measure predicates. As I don't make use of continuant states I don't discuss this issue in more detail.

⁸Continuant states are idealized in the sense that an abstraction from ongoing change below or above a certain threshold is applied.

5.3.3 Interactions of explanations and tense

The following incorporates the interactions of lexical aspect with tense⁹, as it is only in the light of the constant progression of time and action that the use of the 'segmentation operators' CAUSE,DO,INT makes sense. It is important to notice that I consider the term 'tense' not only a linguistic concept but in a more general sense a fundamental cognitive concept. That is, the meaning of 'tense' as I use it in the following goes beyond the linguistic notion of tense but pertains to the cognitive dimension of tense, where tense defines the terms in which an agent can reason about her past, present and future.

The contribution of tense

A decisive component in the semantic representation of tense is the IRS 'now'-constant n which always refers to the current EPS time t . The progression of EPS time and the consequent movement of n affects the temporal location of time-individuals. A temporal individual located in the future becomes present once the n -pointer reaches the location of the time-individual and is located in the past when n has passed the location of the time-individual. This must be mirrored in the IRS representation of tense in two respects. First, I need to represent the relation of a time-individual with respect to the current now and other time-individuals with the help of the relation predicates \prec (entirely precedes), $<_{beg}$ (starts before), \subset (temporally included in), \subseteq (temporally included in or equal to). In turn, the temporal location of a time-individual with respect to the present now affects the type of explanation which is to be employed. Past time-individuals are to be accounted for via causal explanation, while a present or future process is to be accounted for via plan-goal or intentional explanation. But a future time-individual that is explained with the help of intentions can be causally explained once it has become past. In the following, I illustrate this change of view on time-individuals by looking at a number of statements that can be obtained by combining the phrase 'build a corner bolting' with different tenses. We begin by having a look at the following utterances:

- (2) I build a corner bolting.
- (3) I will build a corner bolting.
- (4) I am going to build a corner bolting.
- (5) I am building a corner bolting.
- (6) I built a corner bolting.

While utterances (2) to (6) describe the same action of building a corner bolting, they present the action different as regarding the relation between the action, the current now as well as the intentions of the agent who is described as performing the action of building a corner bolting and their executions. In the first case (2), the utterer expresses her present active intention to build a corner bolting which is announced to be realized soon after the utterance. Example (3) is normally taken to express that at some time in the future the utterer will have the intention to build a corner bolting and that the actual execution of the intention will take place after this future intention has become active. In example (4),

⁹It should be noted that in the following I do not give an analysis of all tenses which are available to a speaker of English but restrict myself to those tenses which are crucial for the processing of goal-directed real-time interaction between humans and robots.

the utterer expresses that she has had the intention for some time before the utterance and is going to realize this intention at some time in the future. Utterance (5) states that the utterer is building a corner bolting at the time of the utterance with the intended result of bringing about a corner bolting. Finally, example (6) refers to an action the utterer has performed and completed successfully in the past. Table 5.1 states the different modes of tensed action presentation via a combination of the performer’s intention (e_0), the present now (n) and the realization of the intention via action (e_1) by the performer of the action in question.

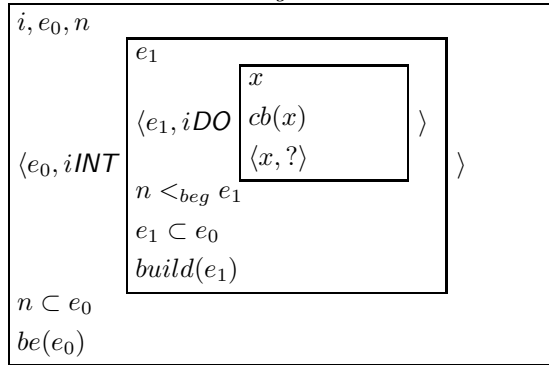
Simple present	Will-future	Going-to future	Present progressive	Simple past
$n \subset e_0 <_{beg} e_1$	$n <_{beg} e_0 <_{beg} e_1$	$e_0 <_{beg} n <_{beg} e_1$	$e_0 \subseteq n$	
$e_1 \subset e_0$	$e_1 \subseteq e_0$	$e_1 \subset e_0$	$e_1 \subset e_0$	$e_1 \prec n$

Table 5.1: Tense as a relation between an intention (e_0), the now-constant (n) and an action that realizes the intention (e_1).

The examples 6 to 10 apply the proposed analysis of tense to the treatment of temporal individuals in the IRS. It should be noted that I distinguish intentional action from intended action at the level of IRS representation via different handles for the intention: intentional action is represented as an event $be(e)$ and intended action is represented as an event $will(e)$ resp. $going - to(e)$. That intentions are represented by events rather than states is motivated by Bratman’s plan-based account to intentions introduced in section 4.1.2.

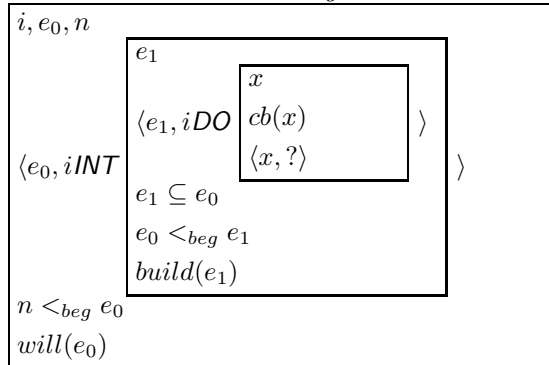
Example 6 *Simple present*

“I build a corner bolting”



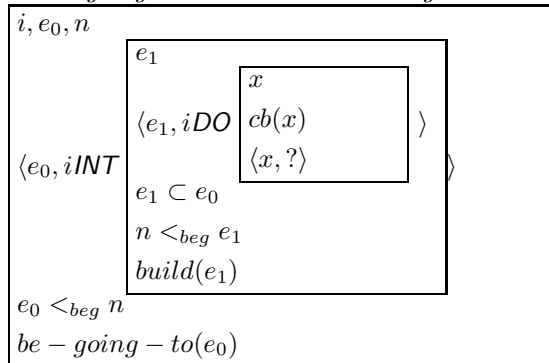
Example 7 *Will-future*

“I will build a corner bolting”

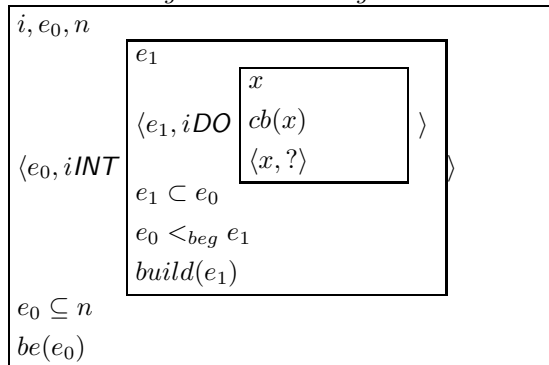


Example 8 *Going-to future*

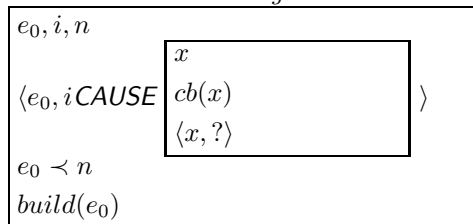
“I am going to build a corner bolting”

**Example 9** *Present progressive*

“I am building a corner bolting”.

**Example 10** *Simple past*

“I built a corner bolting”



Mindful of these examples I will develop a systematic treatment of the relation between tense, intentions, time and actions that allows an agent to properly construct and interpret IRSs from utterances or for the purpose of uttering in the next section. Before I do so, I have to discuss a final point concerning the interpretation of (tensed) time-individuals occurring in an IRS.

5.3.4 Relating utterances, thoughts, representations and actions

In this section, I discuss how an IRS derived from speech actions or an agent’s own cognitive actions relates to (sequences of) internal and external actions that should be undertaken by an agent who is able to act in response to such representations in a semantically *and* pragmatically meaningful manner. That is, we must consider the semantic and pragmatic dimension of meaning, as only the combination of semantic and pragmatic processing of IRSs enables a robot to reasonably participate in goal-directed

interaction. It is exactly this 'interface' between formal semantics and formal pragmatics that I seek to develop in this thesis. We do not only want the robot to know what a certain utterance means in terms of truth conditions but also what she should do in response to the utterance, i.e. how a proper reaction can be calculated from a given IRS constructed from the utterance. Similar considerations apply to the case where an IRS does not arise from an utterance but from the agent's own mental processes, e.g. if it is necessary to check whether a certain IRS occurring as part of the agent's reasoning processes is satisfiable given the current circumstances or not or whether further (internal and external) actions need to be undertaken to satisfy the respective IRS. In the next sections, I develop a 'dummy solution' for this problem that is tailored to the needs of this thesis and thus excludes many of the subtleties of a linguistically adequate analysis¹⁰. In particular, with respect to the limitations of this thesis (see section 1.2.3), I assume that semantic representations in the form of IRSs can be constructed from utterances and do not give further elaborations on this point. Furthermore, I assume that IRSs constructed from utterances can not contain specified thing-anchor sources but only variable thing-anchor sources. It is part of the semantic processing of an IRS (the part which I called 'semantic binding' earlier) to determine the internal and external anchors of IRS thing-individuals - but not part of the construction of IRSs.

Reactive and plain interpretation

The problem that we have to deal with when elaborating on the relation between utterances and IRSs is that the correct interpretation in terms of an appropriate reaction to an IRSs constructed from an utterance is hard to capture in a general way. Consider the following utterances:

(7) Please pass the slat!

(8) Did you pass the slat?

An appropriate reaction of a hearer of utterance (7) consists in her invocation of a plan for passing the slat. Consequently, utterance (7) has a successful *interpretation* iff the hearer of the utterance has passed the slat (an external action) - and the utterer holds the slat in her hands. In contrast, an appropriate reaction to utterance (8) consists in an answer "yes" or "no", depending on whether the hearer has passed a slat or not. The hearer must not perform any external actions besides uttering the answer for a successful interpretation of utterance (8). In the following, I distinguish these two cases of utterance interpretation as basic *modes of interpretation*. I call the primary interpretation mode associated with the appropriate reaction to utterance (7) *reactive* interpretation, as it requires further external actions in order to 'fit the world to the utterance' by changing reality in a way that satisfies the intention of the utterer. The primary interpretation mode associated with the interpretation of utterance (8) is what I call *plain* interpretation. Plain interpretation requires to 'fit the utterance to the world' by checking whether the utterance describes reality in an appropriate way. In detecting trivial satisfaction of intentions, plain interpretation is a preliminary to reactive interpretation - so in (7) plain and reactive interpretation are involved. The consequent question is whether and how one can formulate an account to utterance interpretation in terms of plain and reactive interpretation that fits this thesis' requirement for algorithmic feasibility and also applies to the case of IRS interpretation not directly related to utterances.

¹⁰I decided not to discuss empiric material here as this would involve syntactic and semantic issues that are not of central interest here. However, the empiric data this thesis is based on are from a corpus of transcriptions from the so-called tangram task [Carletta et al., Under Revision].

Given the discussion on pragmatic meaning in section 2.3.2, it is not surprising that it is not only intrinsic properties of utterances that determine the right interpretation mode but that we must consider extrinsic 'conventions' on the interpretation of utterances. In the following, I introduce some 'dummy'-conventions which I employ in the further course of this thesis.

The first convention concerns the use of a *standard interpretation procedure*. The standard interpretation procedure for IRSs that will be developed in full detail in the next chapter consists of a plain interpretation attempt followed by a reactive interpretation attempt if the plain interpretation attempt fails followed by a plain interpretation attempt to check the results of the reactive interpretation attempt. This process can be influenced by setting a 'flag' *force-plain* that prevents reactive interpretation. *force-plain* is associated with the semantic-pragmatic concepts associated with certain linguistic entities or can be set during the execution of plans.

Note 12 *Some conventions concerning interpretation mode*

- *An utterance containing the phrase 'know-that' is interpreted only in plain mode. The semantic-pragmatic concept for 'know-that' sets the flag force-plain that prevents reactive interpretation. This convention forces know-that to apply only to factual knowledge.*
- *An utterance of which the semantic representation K involves 'i' as the main agent of the time-individual(s) contained in K is interpreted in plain mode if the time-individual is located in the past. The semantic-pragmatic concept for 'i' sets the flag force-plain if 'i' is the agent of a time-individual located in the past or present. This captures the irrevocability of an agent's history.*

With the help of the conception of interpretation modes as introduced above we can foreshadow the synthesis of the semantic and pragmatic dimension of meaning (section 2.3) - keeping in mind that a formal treatment is developed in the next chapter 6.

A successful *plain* interpretation identifies the thing-individuals and the time-individuals of an IRS K in the EPS. It consequently establishes a semantic binding of K under consideration of the pragmatic profiles of the conditions and anchors occurring in K . A *reactive* interpretation of K creates the possibility of a successful plain interpretation of K in cases where plain interpretation fails and thus models the 'impact' of the pragmatic profile of K on the EPS. In other words, plain interpretation mirrors the conception of truth-conditional semantics whereas reactive interpretation mirrors the conception of action-based pragmatic meaning. But: plain and reactive interpretation are interdependent in that they require each other - just as semantic binding and pragmatic profiling do with respect to meaning. I will leave it at these adumbrations (a full specification is given in the next chapter) and turn to a final important point that we need to discuss: the algorithmic construction of IRS-individuals.

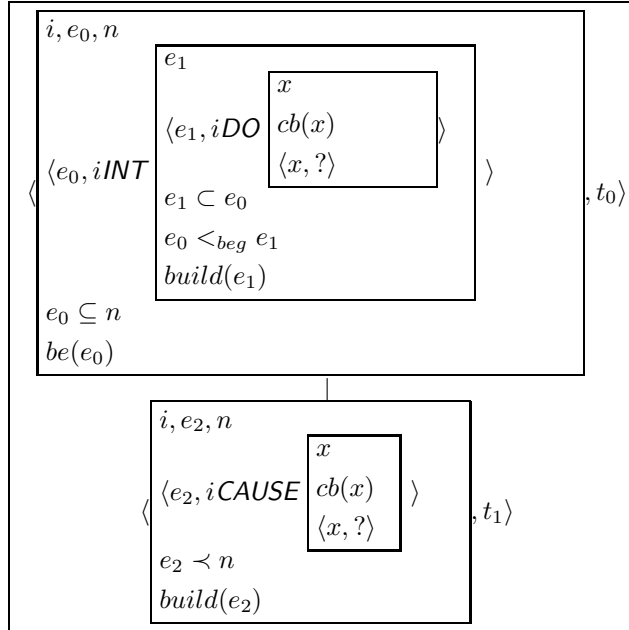
5.4 The construction of IRS-individuals

This section discusses the individuation of thing- and time-individuals and thus captures the inward direction of objective and temporal reference construction as proposed in sections 3.2.2, page 37 and 3.3.3, page 44 in a formal way. Basically, the individuation of entities in the form of IRS-conditions proceeds in the opposite direction to IRS-interpretation. Given an agent's EPS structure, the question is how a rule-based account of the extraction of IRS-entities can be achieved.

5.4.1 Representing the progress of time

Before I start with the explication of IRS individual construction, I must consider a point that has not yet been discussed in great detail. Every embodied agent (such as a robot) is situated in time and as such is subject to the fundamental restrictions imposed by the constant and unstoppable passing by of time. That is, a present EPS-time t_0 at which an agent is holding an IRS in her working memory is unique (t_0 can never be reached again) and fixed (the conditions holding at t_0 can never be revised) and so is the IRS which the agent has in mind at t_0 . In particular it is impossible for the agent at a later time t_1 to travel back in time to revise the IRS which she held in memory at t_0 , if she notices that she *had* misrepresented reality at t_0 and wants to correct this mistake *now* at t_1 . Travelling back in time is possible only by means of temporal *reference*: the only way to access a temporally earlier IRSs is to construct a new IRS at a later time that refers back to that earlier IRS¹¹. At the EPS level, temporal situatedness is modelled by the continuous addition of externally anchored EPS time-indices to an agent's EPS structure, where the permanence of temporal progress stems from the perpetual execution of the main loop of the BDI-interpreter. At the IRS level, we keep track of the progression of time with respect to the development of an agent's IRSs by anchoring each IRS in the EPS-time at which it is constructed. That is, GDRT distinguishes situatedness *in* time - implemented with the runtime environment of the BDI-interpreter - from reasoning *about* time - implemented with the concept of IRS time-individuals. These two dimension of time are illustrated in example 11, where an agent has an IRS representing her intention to build a corner bolting at t_0 . This state of mind is followed by an IRS representing the accomplished intention at t_1 .

Example 11 *Temporal anchoring of IRSs.*

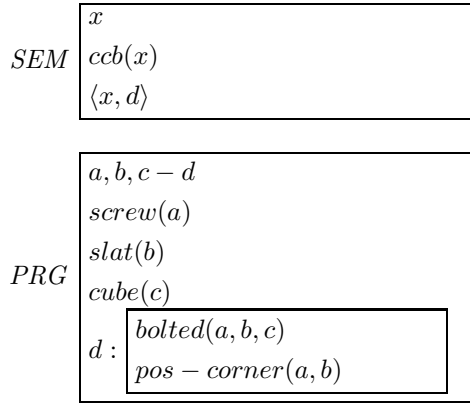


¹¹Consequently, there is no option for belief *revision* in the classical sense, but only for future *updates* of belief.

5.4.2 Thing-individuals

The construction algorithm for thing-individuals has already been touched upon in section 3.2.2, where I discussed how a corner cube bolting is individuated from a specific EPS configuration that matches the pragmatic identification conditions $[PRG]$ of the corner cube bolting. I will now render the procedure for the construction of thing-individuals more precise. First, recall the semantic-pragmatic concept for a corner cube bolting as specified in entry 2.

Sem-Prag-Concept 2 *Corner cube bolting*



In addition, assume initial conditions as specified in figure 5.2.

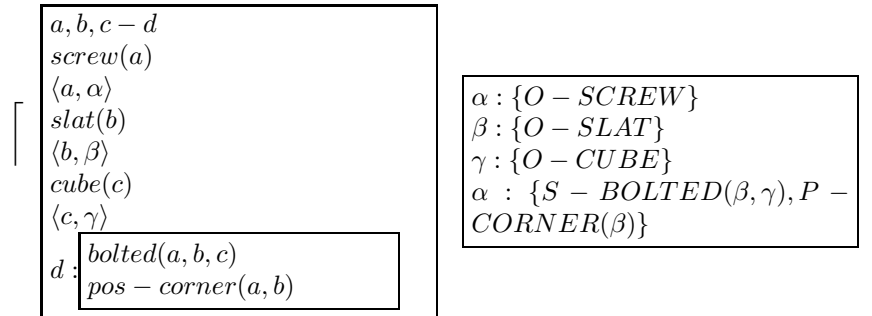


Figure 5.2: Initial EPS (left) and SMS (right) configuration

The BDI-interpreter (remember definition 13) contains a line `update(IRS)`; and its functioning is spelled out in the next paragraphs. Basically, for thing-individuals `update(IRS)`; has to check if the current EPS contains (a set of) conditions which match one of the $[PRG]$ -parts of the agent's knowledge base. In the present case of example 5.2, the EPS matches the $[PRG]$ -part of the entry for 'corner cube bolting'. The `update(IRS)`;-function will then introduce a new IRS-referent x anchored to the bolted-condition (which identifies the corner cube bolting), $\langle x, d \rangle$, and an IRS name-condition $ccb(x)$, resulting in the updated picture 5.3. The consequent question resulting from this example is how the `update(IRS)`;-function should actually be formulated. A sensible answer to this question must also take into account the construction of temporal individuals. Hence, I first discuss the construction of time-individuals before I come back to this topic.

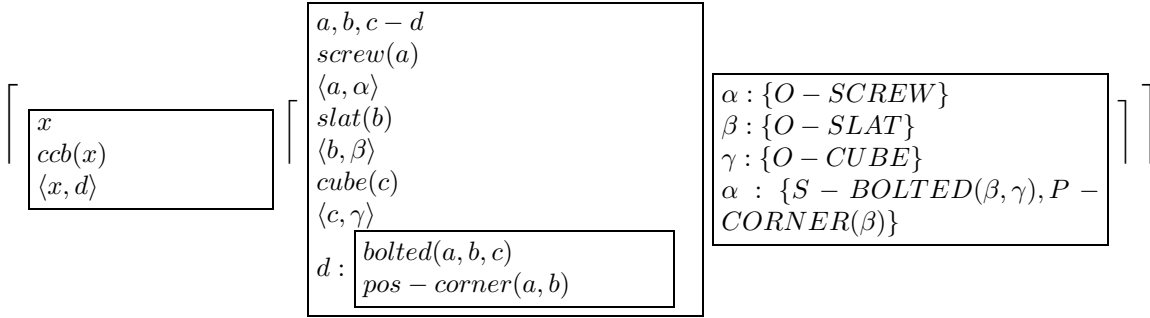


Figure 5.3: Configuration after the individuation of the corner cube bolting in the IRS.

5.4.3 Time-individuals

The construction process for time-individuals was informally discussed in section 3.3.3. The basic idea behind this approach was that time-individuals are individuated from a given EPS-structure under the assumption of causal, plan-goal or intentional structures. However, I left open how exactly this is supposed to work. This section will introduce the construction process for time-individuals in more detail. In doing so, it has to be kept in mind that the segmentation of temporal variation depends on the temporal location of the entity in question, i.e. whether the entity is located before, inside or after the current now. A segment of temporal variation located before the current now should result in a causal explanation, while present or future segments require the assumption of plan-goal structures and intentions. In addition, the construction of time-individuals requires setting the variable `mental-mode` to true in order to decouple the execution of actions from the BDI-interpreter (see section 4.2.4).

Causal temporal individuals

I start the discussion of the individuation of temporal entities with an example of causal individuation, where the action to be individuated is completed, i.e. a factual 'result' state is available in the EPS. Figure 5.4 displays the initial state of the EPS where a cube is on a table and the final state of the EPS where the cube is in the hand of the human. The human performed some unknown kind of action `d-a:unknown` between these two states of affairs.

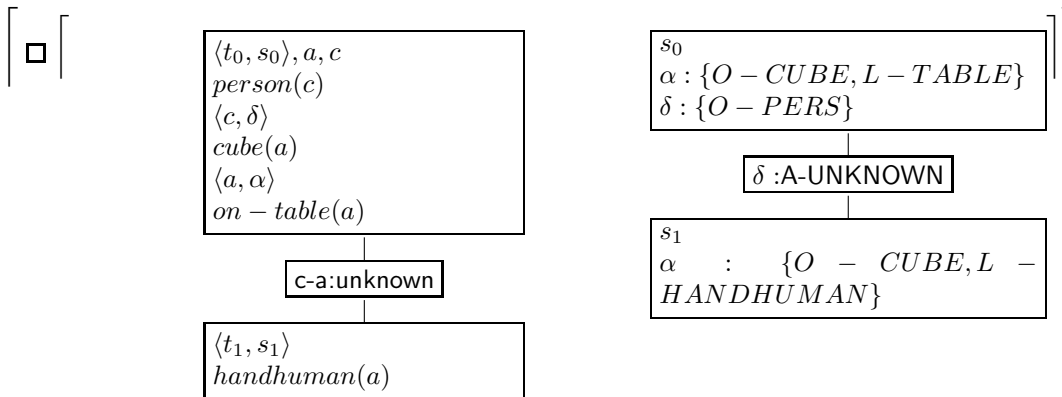


Figure 5.4: Initial configuration for the individuation of a causal temporal individual. No IRS has been constructed yet.

Type:	grasp(K)	
Invocation:	g-add(s, x $\langle s, handhuman(x) \rangle$ $hold(s)$
Context:	$\langle x, ! \rangle$	
Feedback:	a $handhuman(a)$ $\langle a, ! \rangle$	
Body:	<pre> graph TD t0[t_0] --- grasp[grasp(a)] grasp --- t1[t_1] t1 --- end[END(+)] </pre>	

Figure 5.5: Plan for grabbing K, active invocation.

The agent's perceiving of the EPS at t_1 should activate a mechanism that checks whether there exists a plan in the agent's plan library of which the plan's feedback matches the EPS at t_1 . If this is the case, the preceding action *c-a:unknown* in figure 5.4 can be explained with the help of this plan as an execution of this plan. Consequently, a new IRS-referent e_0 for a time-individual can be introduced, anchored in a causal explanation with the target state located at t_1 . The temporal profile of e is defined by the binding of the activated plan's variables to the EPS-history which was used to activate the plan. Figure 5.6 displays the resulting IRS, representing an event of building a cube bolting constructed with the help of the plan for grabbing displayed in figure 5.5.¹²

¹²This procedure is probably too simple, and several refinements could be applied. This concerns in particular the amount of information which is used to identify the temporal entity, e.g. it would be sensible to also consider the initial state of the temporal segment.

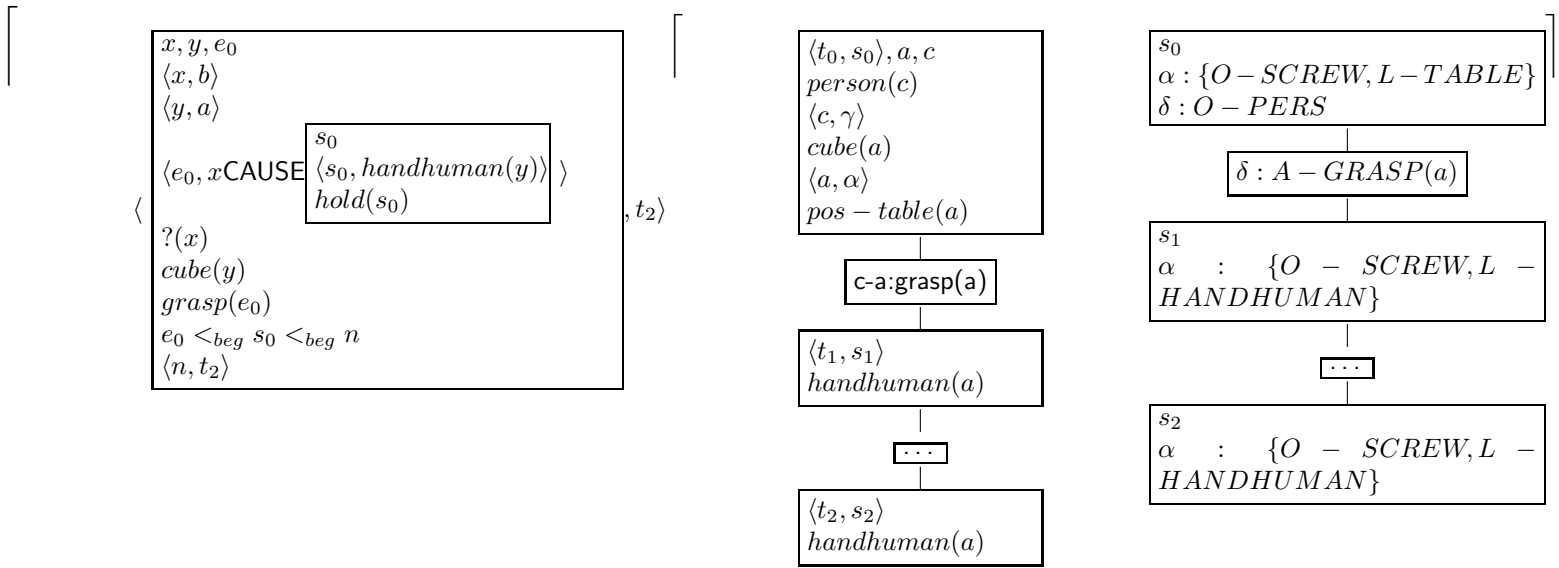


Figure 5.6: Final configuration of causal temporal individual construction, time has moved on to t_2 and the event of grasping is identified in backward-looking manner.

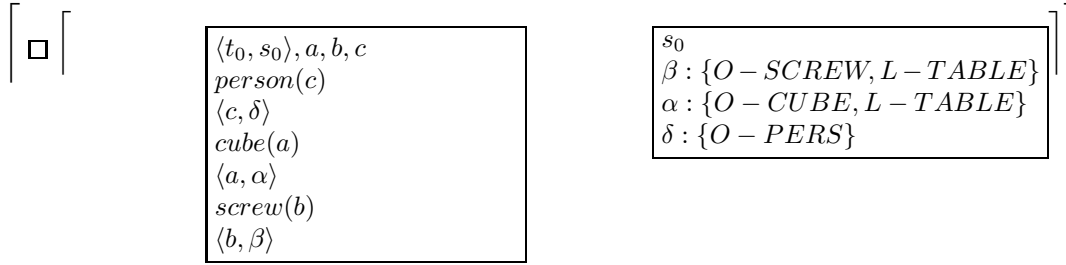


Figure 5.7: Initial configuration for intentional individuation

Type:	build(K)	
Invocation:	f-add(K=	a, b $screw(a)$ $cube(b)$)
Context:	$\langle a, ! \rangle$ $\langle b, ! \rangle$	
Feedback:		
Body:	g-add(build($a, b - d$ $d : bolted(a, b)$ $screw(a)$ $cube(b)$)

Figure 5.8: Reactive invocation of building a cube bolting

DO- and INT-individuals

The next example illustrates the use of desires and intentions for the construction of temporal individuals, i.e. temporal entities with a present or future location. In order to make any predictions about external temporal variation, the robot's BDI-interpreter unit must be used in *mental-mode* (see section 4.2.4) and the robot needs to know about possible desires that can be triggered by certain states of affairs. In the following, this is the (probably a bit artificial) desire to build a cube bolting whenever there is a cube and a screw on the table. The initial configuration of the example is displayed in figure 5.7. The plan in figure 5.8 states that the availability of a cube and a screw on the table triggers a desire to build a cube bolting. Given that the plan in figure 5.8 is activated by an instance of the BDI-interpreter simulating the human (section 4.2.4), the IRS in figure 5.7 should be constructed via the `update(IRS);`-function from the present state of the BDI-interpreter instance for the human. The EPS contains the new option associated with the plan for building a cube bolting and the current now is anchored to the time-span of this plan.

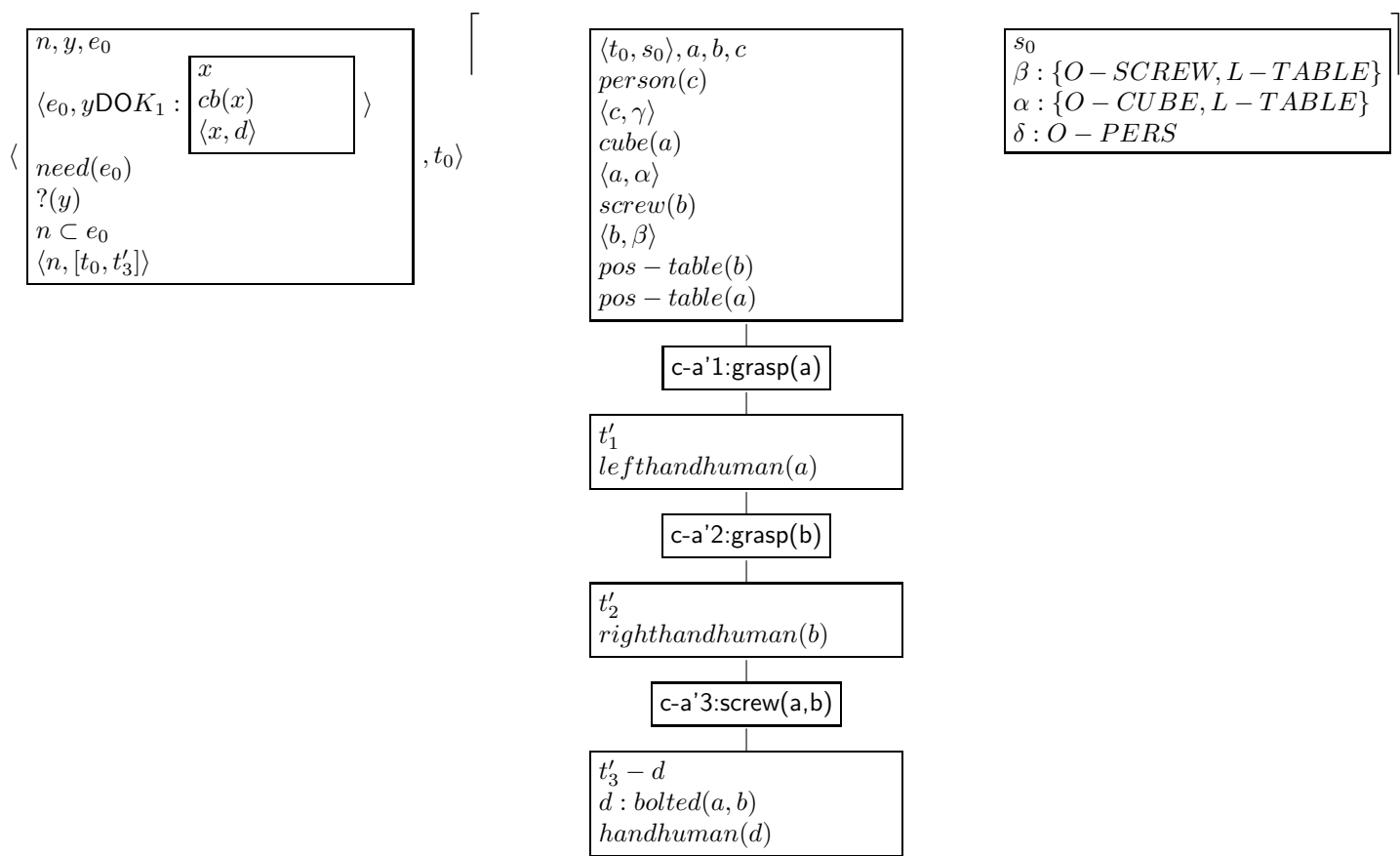


Figure 5.9: Step 2 of the intentional individuation of building a cube bolting by a human: “y needs a cb”. After the EPS has been updated, y’s actions can be explained with the ascription of an intention as represented in the IRS.

Given the scheme for intentional explanation (definition 5) and the configuration of figure 5.9 (where there is a cube and a screw on the table and a plan available for building a cube bolting) this would already suffice for the robot to assume that the human actually adopts the intention to build a cube bolting. However, this is probably a condition too weak for the ascription of intentions. Instead, I require that the human undertakes at least one step in the direction of building a cube bolting in order to justify the assumption that she does have the intention to build a cube bolting. The plan for building a cube bolting should thus be modified in a way such that it triggers the addition of the intention of the agent that is to be explained once it is recognized by the explaining agent that the agent to be explained has brought about the initial state of affairs (i.e. the EPS located at t_1 in the plan) which is part of building a cube bolting. Figure 5.10 pictures this state of affairs. Please note that the action a_1 has been individuated following the procedures of the recent example analysis of causal individuation.

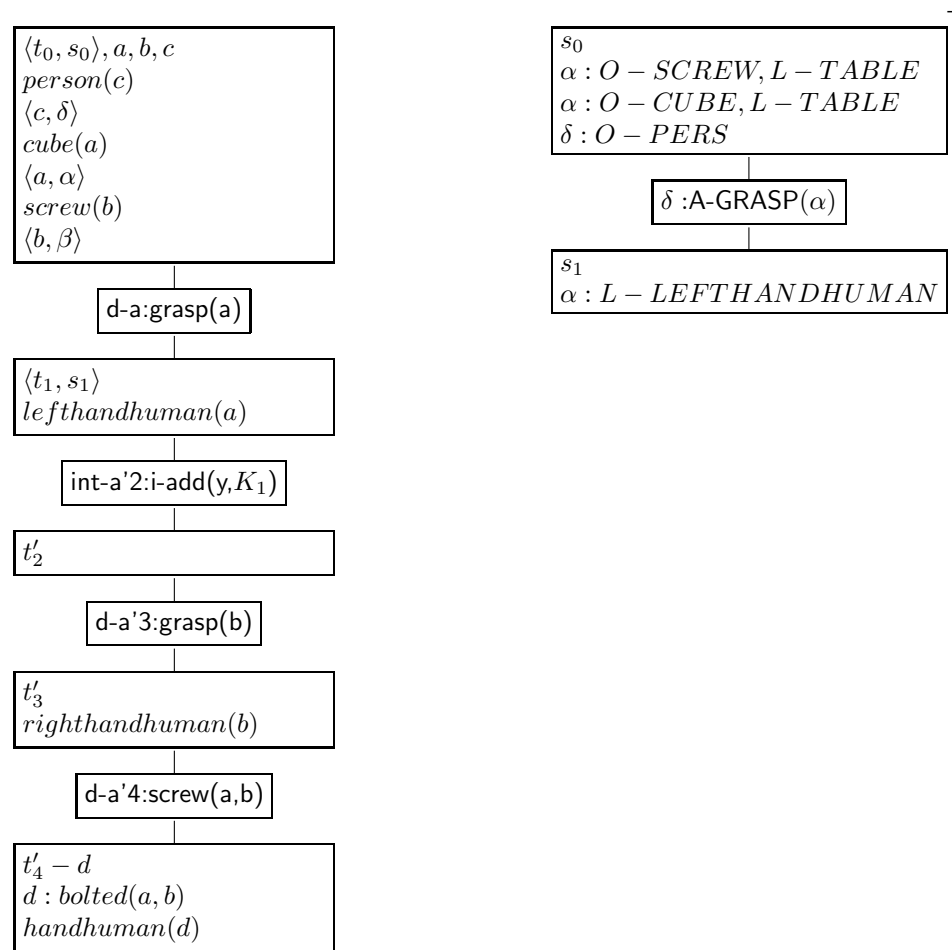


Figure 5.10: Step 3 of the intentional individuation of building a cube bolting by a human. As the human has undertaken the first step toward building a cube bolting at t_1 , the goal of building a cube bolting is added to her intention stack.

Once the the intention of the agent whose actions are to be explained has been confirmed by her initial execution of the plan for building a cube bolting and the intention to build a cube bolting has been added to her intention stack, the `update(IRS)`-function should construct an updated IRS as shown in picture 5.11. The final step of this example is reached once the cube bolting x can be anchored in a factual EPS according to the example analysis in section 5.4.2. The `update(IRS)`-function should then construct an updated IRS via causal explanation, as a factual result state is available that matches the feedback of this plan. This step is pictured in figure 5.12.

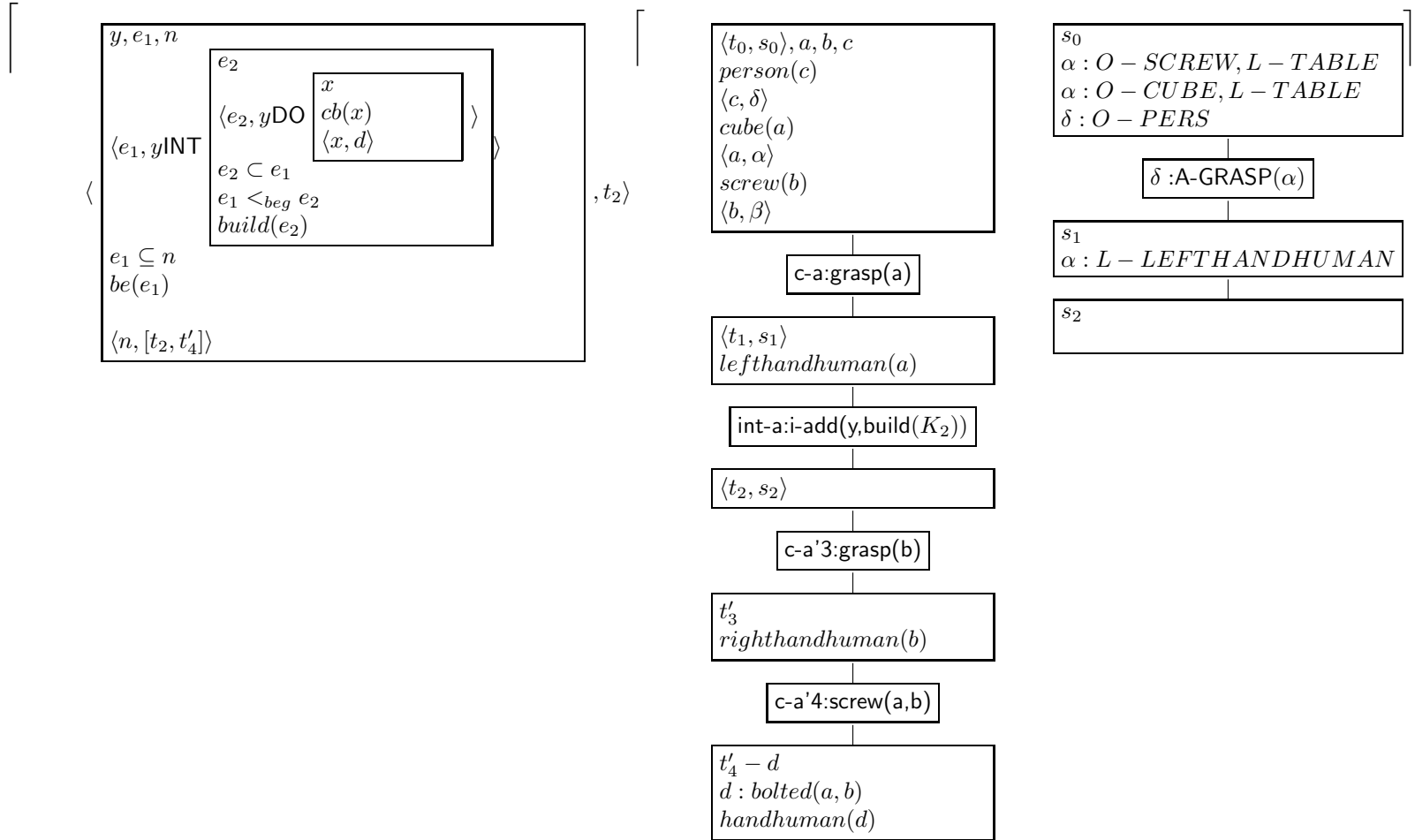


Figure 5.11: Step 4 of the intentional individuation of building a cube bolting by a human: “y is building a cb”

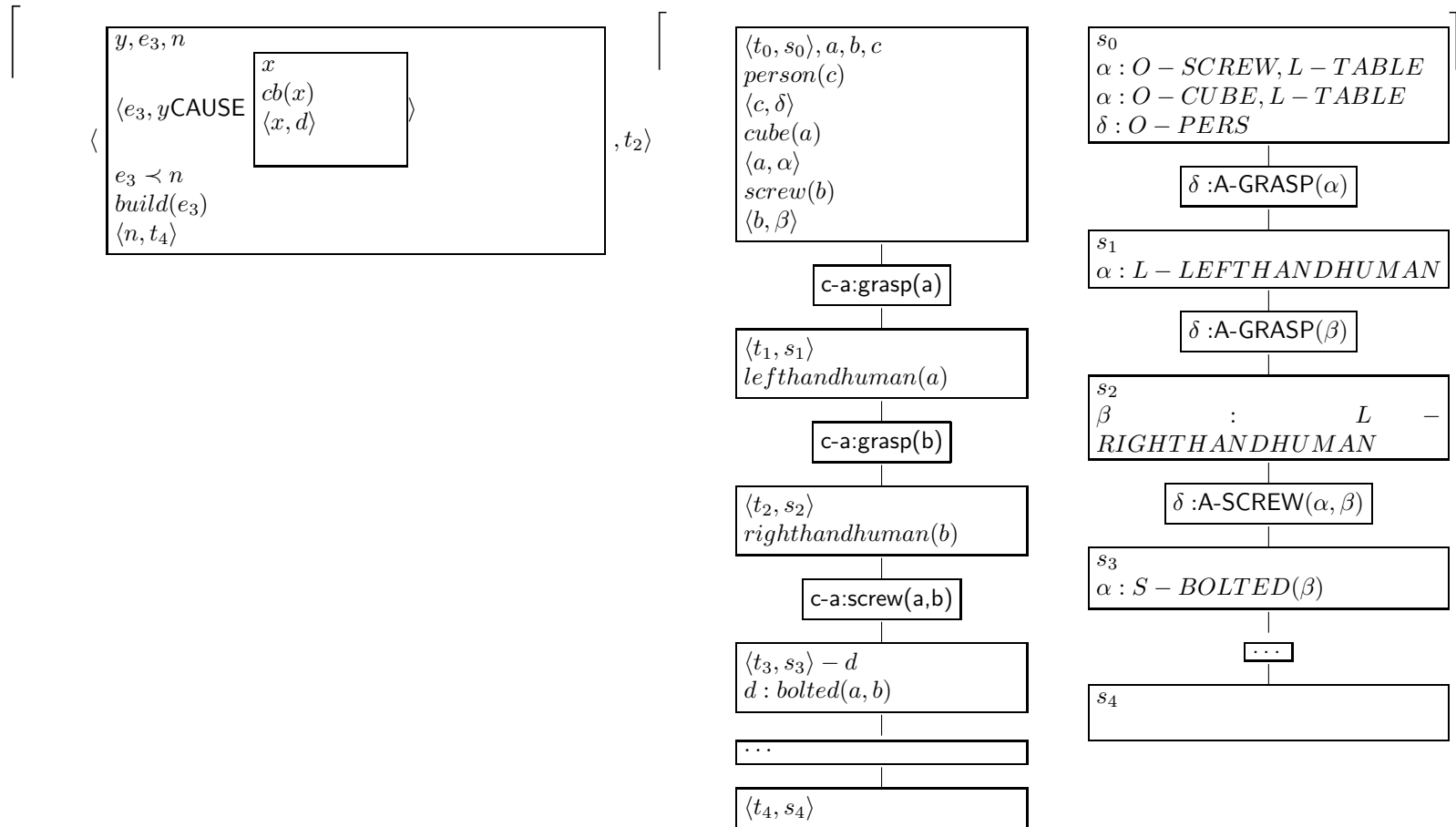


Figure 5.12: Step 5 of the intentional individuation of building a cube bolting by a human: “y built a cb”. The intention ascribed to the human has been realized with the states of affairs holding at t_3 . Consequently, the backward identification of the human’s building of a cb involves a causal explanation.

5.4.4 The `update(IRS);`-function

I now return to a more precise formulation of the construction process for IRS-entities. While I think that it is intuitively clear what the `update(IRS);`-function should do, spelling out a precise algorithmic formulation is a bit more complicated.

For the construction of thing-individuals, the algorithm should iterate through the lexicon of the explaining agent who runs the algorithm in order to check whether there exist some pragmatic identification conditions for a thing-individual in the lexicon which match the current state of the EPS structure. If this succeeds, the semantic part of the respective lexical entry is added to the IRS and an anchor between the added IRS thing-individual and the EPS entities involved is entered into the IRS.

For the construction of time-individuals, the explaining agent must distinguish between the case of 'backward' retrospective individuation by means of causal explanation of past actions and the case of 'forward' prospective individuation by means of plan-goal and intentional explanation. In the former case, the algorithm should iterate through the lexicon of the agent who runs the algorithm in order to check whether there exists a plan of which the feedback matches to the current state of the EPS structure. In the latter case, the algorithm should check whether there exists a plan of which the invocation conditions match the current state of the EPS structure. If such a plan is triggered, both forward and backward construction of time individuals must add the semantic information [*SEM*] associated with the identified time-individual to the IRS. In addition, it must be determined how the constructed time-individual relates to the current now given table 5.1, which I redisplay below.

Simple present	Will-future	Going-to future	Present progressive	Simple past
$n \subset e_0; e_0 <_{beg} e_1$	$n <_{beg} e_0 <_{beg} e_1$	$e_0 < n < e_1$	$e_0 \subseteq n$	
$e_1 \subset e_0$	$e_1 \subseteq e_0$	$e_1 \subset e_0$	$e_1 \subset e_0$	$e_1 < n$

Table 5.2:

A further point that must be considered with respect to the construction of IRS time-individuals is that the construction procedure should not only apply to an agent's own actions but also to other agents. That is, an agent performing forward construction of time-individuals must iterate through her own *and* her simulation of other agent's BDI-interpreters to gather information about the mental background (in the form of beliefs, intentions and desires) of herself and other agents which in turn must then be combined with the information of above table to yield an IRS representing the time-individual describing her or other agent's actions.

Given the current now of the explaining agent, a plan which can be activated by the explaining agent from her current set of facts and the explaining agent's simulation of the current BDI-configuration of the agent to be explained, the following cases must be considered for the forward construction of time-individuals according to table 5.4.4:

- The activated plan is an active intention, i.e. it has been added to the intention stack of the agent whose actions are to be explained before the current now. Two options exist: either the intention is currently executed or the intention will be executed at some later time. In the first case, the segmentation operator of the event to be constructed is INT and the temporal location of the event as well as the goal of the intention are placed inside the current now. In the second case, the

segmentation of the event to be constructed is also INT and the temporal location of the event is placed inside the current now but the goal of the intention is placed after the current now.

- If the plan is an intention which is not active but has been added to the intention stack at some earlier time, the temporal operator of the event to be constructed is INT, the temporal location of the event is placed before the current now and the goal of the intention is placed after the event of activating the intention.
- If the plan is a future intention, the temporal operator of the event to be constructed is INT, the temporal location of the event is placed after the current now and the goal of the intention is placed after the event of activating the intention.
- If the plan is not part of the agent's intentions but her desires, the temporal operator is DO and the temporal location of the event is placed after the current now. This case is associated with verbal phrases such as 'I need' for which I assume that they are not able to relate to intentions but only to desires and thus override the 'standard' procedure for forward time-individual construction.

Finally, the `update(IRS)`-function should also trigger the construction of IRSs from utterances. I assume that the EPS-action 'UTT' as which each perceived utterance is classified is passed on to the `update(IRS)`-function which in turn activates the module which is responsible for the construction of IRSs from utterances.

Summing up what has been said on the construction of IRSs, the following definition 22 gives a possible formulation of the `update(IRS)`-function in terms of our pseudo programming language.

Definition 22 `update(IRS)`;

```
//Start with an empty IRS
K:={};
// Construct thing- and stative time-individuals
for items in sem-prag do
  if matches (PRG(item), EPS) do
//Introduce a new EPS reference marker for the set of EPS-conditions that identify
the respective thing-individual is the EPS (see section 3.2.2)
  if PRG(item)=set-of-relations then introduce(PRG(item),EPS,new-marker);
  K:=K ∪ { SEM(item) };
  instantiate-anchors(K,EPS);
  // latest-time-of-EPS is the last EPS time anchored in an SMS snapshot.
now(K):=latest-time-of-EPS;
end do;
//Backward looking construction of time-individuals
if matches (feedback(plan), facts) then
  K:=K ∪ { SEM(time-individual(plan)) };
  temporal-operator(SEM(time-individual(plan))):=CAUSE;
  now(K):=time-of-EPS;
//Past tense
```

```

    temporal-location(SEM(verb(plan))):=precedes(now(K));
//Forward looking construction of time-individuals
for agents in BDI-interpreters do
    for plans in plan-library do
        if matches(invocation(plan), facts) then
            if invocation(plan) in intentions(agent) then;
//Add the semantics of the time-individual corresponding to the plan to K
            K:=K∪{SEM(time-individual(plan))};
            temporal-operator(SEM(time-individual(plan)):= INT;
            if currently-active(plan) then
                if currently-executed(plan) then
                    now(K):=[time-of-EPS, END(plan)];
//Present progressive
                    temporal-location(plan):=in(now(K));
                    temporal-location(goal):=in(temporal-location(plan));
                elseif
                    now(K):=[time-of-EPS, END(plan)];
// simple present
                    temporal-location(plan):=in(now(K));
                    temporal-location(goal):=after(now(K));
                if previously-added(plan) then
                    now(K):=time-of-EPS;
//going-to-future
                    temporal-location(plan):=before(now(K));
                    temporal-location(execution):=after(now(K));
                if scheduled-adoption(plan) then
                    now(K):=time-of-EPS;
                    //will-future
temporal-location(plan):=after(now(K));
                    temporal-location(execution):=after(temporal-location(plan));
                if invocation(plan) in desires(agent) then
                    temporal-operator(time-individual(plan)):= DO;
                    now(K):=[time-of-EPS, END(plan)];
                    temporal-location(plan):=in(now(K));
                    temporal-location(goal):=after(temporal-location(plan));

            end do;
        end do;
// process utterances
if UTT then
//Add semantic representations constructed from utterances
    K:=K∪{semantic-parse(UTT)}; // Set the time-anchor of the constructed IRS to the EPS-time, see section 5.4.1
    time-anchor(K):=time-of-EPS;

```

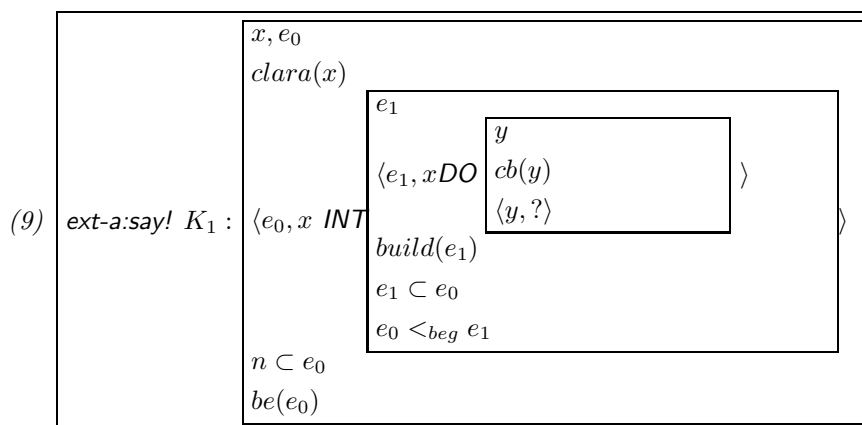
```
// Check the list of pending IRSs, see section 6.3.2
for K in pending-list do
  if resolved(K, set-of-thing-anchors) then
    remove-pending-list(K);
    interpret(K, set-of-thing-anchors);
end do;
```

5.5 A foretaste of discourse processing with GDRT

To conclude this chapter, I want to give a simple example demonstrating the processing of discursive interaction within GDRT as it has been informally developed in the last chapters. Basically, the interpretation mechanisms for IRSs which I define in the next chapter in full detail make use of the close relation between discourses and plans, an idea which goes at least back to the early days of artificial intelligence [Rumelhart, 1975, Schank and Abelson, 1977, Abelson, 1981, Bower, 1982] and discourse analysis [van Dijk and Kintsch, 1983]. These approaches to discourse understanding proposed that the processing of discourse is about recovering the general plan underlying the discourse (where the plan corresponds to the story of the discourse) by incrementally recovering the subplans (corresponding to the phrases of which the discourse is made up) involved. This idea was limited to sequences of sentences, but I will employ it similarly to the interpretation of non-verbal actions that are guided by the principles of rational action. In my implementation of GDRT, the mechanisms of discourse processing are built into the procedure of IRS interpretation as it will be formally developed in section 6.3.2. I will try to illustrate the overall function of discourse processing with GDRT via the following simple example 12.

Example 12 A simple discourse

Assume that an agent holds the following EPS-presentation of a speech action concerning the verbalization of the IRS K_1 in her mind.



Given the restriction of this thesis to the discussion of the semantics-pragmatics interface, I do not spell out how K_1 generates an utterance but assume that the execution of (11) leads to the following utterance directed at Clara:

(10) “Build a corner bolting, please!”

Clara perceives this utterance. Given my assumption of an appropriate design of Clara's lexicon (that properly translates the utterance in an IRS), she should get a similar representation to the one the agent had in mind when giving her the command except for the representation of the action's agent and the location of the now-constant:

$$(11) \quad \begin{array}{l} x, e_0 \\ clara(x) \\ \langle e_0, i \text{ INT} \rangle \\ \text{build}(e_1) \\ e_1 \subset e_0 \\ e_0 <_{beg} e_1 \\ n \subset e_0 \end{array} \left\langle \begin{array}{l} e_1 \\ \langle e_1, i \text{ DO} \rangle \\ \text{cb}(y) \\ \langle y, ? \rangle \end{array} \right\rangle$$

If Clara is cooperative, the interpretation mechanism for the constructed IRS should trigger a reactive interpretation via the plan 'build' that adds the commanded goal to her own desires.

$$(12) \quad \text{int:g-add}(\text{build}(\left. \begin{array}{l} a, b - c \\ \text{nut}(a) \\ \text{cube}(b) \\ c : \text{bolted}(a, b) \end{array} \right)))$$

Once this goal is accomplished, she should confirm the successful interpretation of the command by saying "OK".

$$(13) \quad \text{ext-a:say! "OK"}$$

5.6 Summary

This chapter has undertaken the first steps toward the processing of discursive interaction. I have introduced a general notion of discourse not limited to verbal actions that treats a discourse as a means-end rational surface realization of plans involving one or more agents and an environment. A discourse can involve different means and modalities of action, of which I have discussed bodily actions, thoughts and verbal actions. I also discussed the representation and construction of IRS-individuals. Of course, I am not done with this yet. I still have to show how discourses are actually induced, controlled and processed and dealing with this task will make up the remainder of this thesis. Before I do so, I give a complete specification of the SMS, EPS and IRS and associated procedures in the next chapter.

Chapter 6

Putting Things Together

The previous chapters have discussed the relation between an agent's mind and her reality from various angles:

- Chapter 2 introduced the notion of reference as a form of explanation, which enables an agent to make sense of EPS presentations of reality and mind by subsuming sets of distributed EPS-conditions as representations of IRS-individuals.
- Chapter 3 spelled out this idea in more detail by relating explanations to a branching-time model of reality and mind.
- Chapter 4 presented a structure of rational control which governs an agent's behavior and generates the future-related branches of her EPS.
- Chapter 5 developed a notion of discourse suitable to the needs of this study and discussed the form and the construction of IRSs.

This chapter gives the formal specifications of the theory of GDRT developed so far, including the definition of SMSs, SMS structures, EPSs, EPS structures and IRSs. This prepares the ground for the second part of this thesis, in which I apply the theoretical considerations of the first part to the analysis of examples of discourse processing within the framework of GDRT.

6.1 The SMS: Sensorimotor Structures

This section defines the SMS layer - its constituents and its structure. It should be noted that the SMS layer as discussed in section 3.2.1 does not only contain the output of the object recognition engine but is also responsible for motor control of actions as well as speech recognition and generation. In addition, we should recall that the SMS is not directly accessible to an agent but only via its EPS presentation (the 'translation' of SMS structures to EPS structures is discussed in section 6.2.5).

6.1.1 SMSs and SMS structures

A SMS is a structure mirroring the information contained in a single snapshot of the current state of affairs as delivered by the perceptual instruments of the robot. A sequence of SMS snapshots makes up a

SMS structure, which I will discuss after I have defined the structure of single SMS snapshots. Formally, a SMS is defined as in definitions 23 to 26.

Definition 23 *SMS Vocabulary*

The vocabulary of the language of SMSs contains

- A set of object terms $\{\alpha, \dots, \omega, \dots\}$ that represent objects which are recognized by the object recognition
- A set of property terms $\{\text{property}_1^a, \dots, \text{property}_n^z, \dots\}$ that represent properties which the object recognition engine is able to recognize as holding of one or more objects.
 - A property term is a m -place relation with $m \geq 1$.
 - 1-place property terms express properties holding of one object term,
 - m -place property terms with $m > 1$ express relations between several object terms.
 - The existence of object terms which occur as the arguments of property terms is guaranteed by the output of the object recognition, i.e. a relation between two objects can only be recognized if both involved objects are recognized.
- A set of SMS-snapshot indices $\{s_1, \dots, s_p, \dots\}$, where the alphabetical subscripts do neither indicate an ordering, an intrinsic meaning nor an internal structure of SMS-snapshot indices but are used only to clarify the design of SMS structure.

Definition 24 *SMS Expressions*

A SMS expression is an $o + 1$ tuple, consisting of an object term α followed by o property terms, where α is the object to which the properties $\text{property}_1^a, \dots, \text{property}_o^z$ are assigned by the object recognition.

- $\alpha : \{\text{property}_1^a, \dots, \text{property}_o^z\}$

Definition 25 *SMS snapshots and SMS sequences*

A SMS snapshot (or simply SMS) as pictured in figure 6.1 is a finite (possibly empty) set of SMS expressions annotated with a snapshot index s_i where $i = 1, 2, \dots, r, \dots$. A set of SMS snapshots as delivered by the object recognition forms a SMS sequence $\langle s_i, s_{i+1}, \dots, s_{i+h} \rangle$.

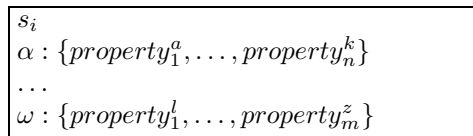


Figure 6.1: A SMS snapshot.

Definition 26 *SMS structures*

A SMS structure \mathcal{S} is a tuple

- $\langle \mathcal{S}, \text{between}, \text{transitions} \rangle$

Where

- S is a finite set of indexed SMS snapshots $S = \{S_1, \dots, S_p\}$
- 'between' a three-place relation in S defined as a direction-neutral ordering that satisfies the following properties, where free variables are assumed to be quantified universally with maximal scope:
 - (1) $\text{between}(x, y, z) \rightarrow \text{between}(z, y, x)$
 - (2) $\neg \text{between}(x, y, x)$
 - (3) $\neg \text{between}(x, y, y)$
 - (4) $\text{between}(x, y, z) \rightarrow \neg \text{between}(x, z, y)$
 - (5) $\text{between}(x, y, z) \rightarrow (\text{between}(x, z, u) \leftrightarrow \text{between}(y, z, u))$
 - (6) $(x \neq y \wedge x \neq z \wedge y \neq z) \rightarrow (\text{between}(x, y, z) \vee \text{between}(y, z, x) \vee \text{between}(z, x, y))$
- *transitions* is a set of SMS transitions (see definition 28), where each SMS transition connects a pair of adjacent SMSs (in the sense of 'between').

SMS property terms

The object recognition engine is supposed to capture different types and properties of objects. It should be noted that the following definitions of SMS property and transition terms present only a subset of the actually available output of a given object recognition engine and a given motor control - I state only those terms which are involved in the analysis of example discourses in the second part of this thesis. In the following definitions, I give a natural language description of the property term in brackets.

Definition 27 SMS property terms

- *1-place property terms*
 - *Object Types*
 - * *O-UNDEFINED* (Unknown object type)
 - * *O-CUBE* (Cube)
 - * *O-NUT* (Nut)
 - * *O-SCREW* (Screw)
 - * *O-SLAT* (Slat)
 - * *O-PERSON* (Person)
 - *Object Color*
 - * *C-RED* (red)
 - * *C-GREEN* (green)
 - *Location of physical objects*
 - * *L-TABLE* (On the table)
 - * *L-LEFTHANDROBOT* (Left hand of the robot)

- * *L-RIGHTHANDROBOT* (Right hand of the robot)
- * *L-HANDROBOT* (Hand of the robot)
- * *L-LEFTHANDHUMAN* (Left hand of the user)
- * *L-RIGHTHANDHUMAN* (Right hand of the user)
- * *L-HANDHUMAN* (Hand of the user)
- *m-Place property terms; the number of arguments is specified by the superscript m*
- *Position of physical objects*
 - * *P – BETWEEN³* (three-place property term: *x* is spatially between *y* and *z*)
 - * *POS – CORNER²* (two-place property term: *x* is at the corner of *y*)
- *State of physical objects*
 - * *S – BOLTED²* (two-place property term: *x* is bolted to *y*)
 - * *S – BOLTED³* (three-place property term: *x* is bolted to *y* with *z*)

SMS transitions

Each pair of adjacent (in the sense of the *between*-ordering) SMSs is connected by a *transition* constituting a basic motor action grounded in the robot's motor control (so we do not need an extra argument for the agent of the motor action). The set of SMS transition terms is given in definition 28.

Definition 28 SMS transition terms

The number of arguments of a transition term is indicated by the superscript n if $n > 1$. α, β, γ are object terms. In the following, 'A' stands short for action: A(CTION)-TYPE.

- *A – UNDEFINED* (Undefined action)
- *A – HOLD²(α , PROPERTY – TERM)* (Hold an object as specified by PROPERTY – TERM)
- *A – GRASP(α)* (Grasp an object)
- *A – POINT(α)* (Point to an object)
- *A – SCREW²(α , β)* (Screw two object together)
- *A – SCREW⁴(α , β , γ , PROPERTY – TERM)* (Screw three objects together to a certain configuration described by PROPERTY – TERM)
- *α : UTT :!* 'Utterance' (Utterance emitted by object α)
- *Empty transition: No action happened. (This case will rarely occur if the vocabulary for SMS transitions is sufficiently large and fine-grained.)*

6.2 The EPS: External Presentation Structures

This section introduces an agent's external presentation structure (EPS) in formal detail. In the following it is important to know that I use the term *EPS* for the presentation of a time annotated with a set of things and conditions and *EPS structure* for the branching-time structure of annotated times.

The EPS structures result from two sources of information: the SMS provides the factual part (past and present) of the EPS; the output of the BDI-interpreter generates the branching-time structure of the EPS in terms of beliefs about the future. In addition, the EPS makes use of the SMS in that its external anchor sources for thing reference markers are SMS object terms.

6.2.1 The multi-purpose function of the EPS

Within the system proposed in this thesis, the EPS plays a central and versatile role:

- (1) EPS structures provide models for the evaluation of IRSs (section 2.2). In the models provided by EPS structures, the things that are part of the EPS structure become members of the universe of the model to which IRS reference markers for thing-individuals can be anchored.
- (2) Models based on EPS structures make it possible to evaluate IRS time-individuals which are anchored in substructures of the EPS and individuated by causal, behavioral or intentional explanation (sections 2.1.2,3.3.3).
- (3) A given EPS structure specifies an agent's belief state and knowledge state at a certain time (section 4.1.2).

The way the EPS structure is related to model theory in the sense of formal logic and also to mental and cognitive models is central to the argumentation of the whole thesis and the innovation that comes with the introduction of the EPS can be appreciated fully only in the light of the discussion of different conceptions of models in section 2.2. I elaborate on this point before giving a formal definition of EPSs and EPS structures.

The EPS as dynamic mental model structure

The formal dimension of the EPS structure Let me briefly review the way an agent's EPS structure is constructed. Basically, an EPS structure is based on two sources of information, namely the SMS (presenting perceptions of reality) and the BDI-interpreter (generating possible developments of reality). This information is structured by temporal annotation as discussed in chapter 3.3, resulting in a tree-like structure of EPS-times where each time is decorated with a set of EPS things and conditions. An EPS structure can be formally interpreted as a modal model structure: it provides a set of EPS-times and an accessibility relation (the set of atomic actions ordered by a tree-like partial order $<$). The EPSs which constitute the given EPS structure then provide 'extensional models' that are indexed by the different EPS times. Given this model-theoretic view of an EPS structure, IRSs can be related to a given EPS structure in terms of a model-theoretic satisfaction relation. That is, the EPS structure provides set-theoretic models for the object language in which IRSs are formulated. The crucial point is, however, that the EPS structure itself is grounded in the output of the SMS and BDI-interpreter and

thus dynamically mirrors both the evolution of the agent's perception of her environment and of her internal states. It is this concept of anchoring that motivates my use of the term 'grounded': IRSs are grounded in the sense that they are interpreted with respect to a dynamically changing model structure (derived from the EPS structure) which in turn is grounded in uncircumventable perceptions of reality, blueprints of the future and intentional, doxastic and epistemic states of the agent.

The cognitive and mental dimension of the EPS structure Besides the notion of a formal model, the discussion of models in section 2.2 has drawn attention to two additional aspects of the term model both of which are relevant to a model theory that is suitable to our needs. First, the EPS has a cognitive dimension in that its factual part serves Craik's postulation of a small-scale model of reality. Second, this small scale model of reality allows for predictions of future developments (section 2.2.1) in terms of beliefs generated by the BDI-interpreter. These future options are structured by the agent's intentions and desires: some of the possible scenarios are preferred over others by choice and commitment. In this way the EPS plays the part of modeling the agent's mental contents and processes (section 2.2.1).

The dynamic nature of the EPS structure A point central to proper understanding of what will be said below about the use of EPS structures as models for the language of IRSs is that while an agent's EPS structure is permanently under revision through resulting from the addition of new factual states of affairs and new options of future action, an EPS structure can nevertheless provide complete models for the interpretation of IRSs *at a certain time*. We return to this point when we come to define models for IRS interpretation in section 6.3.2.

6.2.2 Syntax of EPSs

The language of EPSs incorporates vocabulary of the SMS in that it includes SMS object terms as external anchor sources for EPS thing reference markers.

Definition 29 *EPS vocabulary*

- A set T_R of EPS reference markers for things: $\{a_1, \dots, a_n, \dots\}$
- A set T_A of SMS anchor sources for things (SMS object terms) $\{\alpha, \dots, \gamma, \dots\}$ and EPS times (SMS snapshot indices) $\{s_1, \dots, s_n, \dots\}$
- A set A_{var} of EPS thing anchor source variables $\{?_1, \dots, ?_n, \dots, !_1, \dots, !_m, \dots\}$
- For each $n > 0$ a set Rel^n of n -place predicate constants for handles $\{C_1, \dots, C_m, \dots\}$
- A set **Times** of EPS-times $\{t_0, \dots, t_n, \dots\}$, where the numerical subscripts do neither indicate an ordering, an intrinsic meaning nor an internal structure of EPS-times but are used only to clarify the design of the EPS structure.

Definition 30 *Syntax of EPSs, EPS conditions and EPS anchors*

- (1) If $U \subseteq T_R \cup \mathbf{Times}$, Con a (possibly empty) set of conditions and $Anchors$ a (possibly empty) set of anchors, then $\langle U, Con, Anchors \rangle$ is an EPS

(2) If $R_1, \dots, R_m \in \text{Rel}^n$, $d \in T_R$ and $a_1, \dots, a_n, \dots, k_1, \dots, k_l \in T_R$ then d :

$R_1(a_1, \dots, a_n)$
...
$R_m(k_1, \dots, k_l)$

is an EPS-condition

(3) If $a \in T_R$ and $\alpha \in T_A$ then $\langle a, \alpha \rangle$ is an external EPS-anchor

(4) If $a \in T_R$ and $? \in A_{var}$ then $\langle a, ? \rangle$ is a variable EPS-anchor

(5) If $a \in T_R$ and $! \in A_{var}$ then $\langle a, ! \rangle$ is a variable EPS-anchor

(6) If $t \in \mathbf{Times}$ and $s \in T_A$ then $\langle t, s \rangle$ is a temporal EPS-anchor

- I call EPS-conditions for which all occurring EPS reference markers are externally anchored **Facts**.
- EPS-conditions in which EPS reference markers occur which are not externally anchored are called **Beliefs**

6.2.3 The EPS structure

The branching-time structure of a set of EPSs results from the combination of the translation of the given SMS structure to a set of EPS time indices with linear order and the BDI-interpreter's output of future options of action in the form of branching structures of unanchored EPS time indices. Consequently, an EPS structure includes two types of EPS times, where the 'now' of the agent divides the set of EPS times into a linear anchored and a branching non-anchored part. Consequently, for the definition of EPS structures we must consider both types of EPSs. In a first step, note 13 states how anchored and non-anchored time-indexed EPSs are represented.

Note 13 *Representation of time-indexed EPSs*

A time-indexed EPS is a tuple

- $\langle t, \langle t, s \rangle, \langle U, \text{Con}, \text{Anch} \rangle \rangle$

where t is an EPS time, s a SMS snapshot index and $U, \text{Con}, \text{Anch}$ are as in definition 30. In this case, I represent the EPS as

$U, \langle t, s \rangle$
Con
Anch

- $\langle t, \langle U, \text{Con}, \text{Anch} \rangle \rangle$

where t is an EPS time, s a SMS snapshot index and $U, \text{Con}, \text{Anch}$ are as in definition 30. In this case, I represent the EPS as

U, t
Con
Anch

EPS time indices are explicated for the purpose of easy presentation, but they are not directly accessible to the agent for whom we design the EPS. That is, EPS time indices (and similarly SMS snapshot indices and IRS time indices) are introduced to provide us - as designers - with a means to develop the relations between EPS, SMS and IRS structures in which an agent is embedded and which she can only access via explicit representation at the IRS level. Similar considerations hold for the functioning of the BDI-interpreter and associated data structures and algorithms. The agent does not need to know about how the BDI-interpreter is actually realized but it is sufficient for her to be able to access the processes of the BDI-interpreter via her IRS representations.

The structure of time-indexed EPSs resulting from the output of the BDI-interpreter and the SMS can be formally described in terms of a modal model structure (cf. [Singh, 1994], [Emerson, 1990]).

Definition 31 *EPS Structure*

An EPS structure is a tuple $E = \{\mathbf{T}, I, \text{Actions}\}$ of an agent x at time t , where

- $\mathbf{T} = \langle \cdot, \mathbf{Times}_A \rangle$ is a time structure of an agent x at time t as specified in definition 32, where $\mathbf{Times}_A \subseteq \mathbf{Times}$
- I associates times $t \in \mathbf{Times}_A$ with EPSs, i.e. I is a function from \mathbf{Times}_A to EPSs according to definition 13¹.
- Actions is a function from pairs $\langle t, t' \rangle$ of adjacent members of \mathbf{Times}_A to Actions as defined in definition 40.

Time structures \mathbf{T}

Definition 32 *The time structures \mathbf{T}*

We may view \mathbf{T} as a labeled directed graph with node set \mathbf{Times}_A , arc set Actions and node labels given by I . In addition, we require the graph of \mathbf{T} to be a tree. \mathbf{T} is

- acyclic provided it contains no directed cycles
- tree-like provided that it is acyclic and each node has at most 1-Actions predecessor
- a tree provided that it is tree-like and there exists a unique node - called the root from which all other nodes of \mathbf{T} are reachable and that has no Action-predecessor.

In set-theoretic terms, we can define the tree structure of \mathbf{T} as follows:

- Given a set of EPS-time points \mathbf{Times}_A
 - $\langle \cdot \subseteq \mathbf{Times}_A \times \mathbf{Times}_A$ is a strong partial ordering on \mathbf{Times}_A , i.e. $\langle \cdot$ is
- (1) Transitive: $(\forall t, t', t'' \in \mathbf{Times}_A : (t < t' \wedge t' < t'') \Rightarrow t < t'')$

¹That is, the interpretation I of an EPS-time $t \in \mathbf{Times}_A$ is a function from time indices t to EPSs as defined by a set of time-indexed EPSs $\langle t_1, \langle U_1, Con_1, Anch_1 \rangle \rangle, \dots \langle t_n, \langle U_n, Con_n, Anch_n \rangle \rangle, \dots$. Thus the 'semantics' of EPSs is either determined by an unquestionable perception of reality (a fact) or reflects the agent's planning structure (a belief). In both cases, the EPS in question must be taken as true presentation.

(2) *Asymmetric*: $(\forall t, t' \in \mathbf{Times}_A : t < t' \Rightarrow t' \not< t)$

- $<$ allows no merging of branches:

$$(\forall t, t', t'' \in \mathbf{Times}_A : ((t < t'') \wedge (t' < t'')) \Rightarrow ((t < t') \vee (t = t') \vee (t' < t)))$$

Scenarios

A useful notion is that of a *scenario* S at a time t which denotes the set of times that includes t and all future times on a branch \mathcal{B} of a given time structure T which departs from t (i.e. the set of all t' on \mathcal{B} s.th. $t \leq t'$). Thus a scenario formalizes one way in which the world could possibly evolve in the future. Formally, a scenario is defined as in Definition 33 (cf. [Singh, 1994]).

Definition 33 Scenarios

- A scenario is an EPS structure $\{\mathbf{T}, I, \text{Actions}\}$ such that $<$ is a linear ordering. Let $R = \{\mathbf{T}, I, \text{Actions}\}$ be an EPS structure. $S = \{\mathbf{T}', I', \text{Actions}'\}$ is a scenario of R iff
 - S is a scenario
 - $I' \subseteq I$
 - \mathbf{T}' is a substructure of \mathbf{T} .
- If S is a scenario of R , then there will be $t, t' \in \text{Dom}(I)$ so that \mathbf{T}' is the segment (t, t') of \mathbf{T} . t is called the starting point of S in R . Of particular interest are those scenarios S of R in which t' is a leaf of \mathbf{T} . When t is the starting point of S then we write ' $S(t)$ '.
- $S(t) \subseteq \mathbf{T}$ denotes the set of all scenarios of \mathbf{T} at t .
- The notation $[S; t, t']$ denotes an inclusive interval on a scenario S from t to t' with $t, t' \in S$ and $t \leq t'$.

Plans

Plans of some EPS structure R are like scenarios except that their time structure must not be linear.

Definition 34 Plans

- Similar to scenarios, a plan P of some EPS structure R has a starting point t . Its time structure is a subtree of \mathbf{T} with t as root. When t is the starting point of P , then we write ' $P(t)$ '.
- $P(t) \subseteq \mathbf{T}$ denotes the set of plans at t
- $[P; t, t_1]$ denotes a plan starting at t with $\text{END}(+)$ (the goal of the plan) located at t_1 .

6.2.4 Information extraction from the EPS

For the use of EPS structures as models for the interpretation of the language of IRSs, it is useful to convert the 'raw form' of the EPS structure into the logically more manageable form of sets and assignment functions. Here we make use of the function I from EPS times to EPSs (definition 31.) That is, with a given EPS structure $E = \{\mathbf{T}, I, \text{Actions}\}$ of an agent x at time t we - as designers - are provided with the following sets:

Definition 35 *EPS sets of an agent x stored in her EPS structure at t*

I write $\text{Dom}(F)$ for the domain and $\text{Ran}(F)$ for the range of a function F .

- *The set of EPS times $\mathbf{Times}_A = \{t_0, \dots, t_n, \dots\}$ (occurring in $\text{Dom}(I)$)*
- *The set of EPS things $\mathbf{Things} = \{a, b, c, \dots\}$ (occurring in the universes of EPSs in $\text{Ran}(I)$)*
- *The set of EPS properties $\mathbf{Properties} = \{p_1, \dots, p_n, \dots\}$ (occurring in EPSs in $\text{Ran}(I)$)*
- *The set of EPS atomic actions $\mathbf{Actions} = \{a_1, \dots, a_n, \dots\}$ (occurring in $\text{Ran}(\text{Actions})$)*
- *The set of EPS anchors $\mathbf{Anchors}$ (occurring in EPSs in $\text{Ran}(I)$) defines disjunct sets of*
 - *EPS facts \mathbf{Facts} and*
 - *EPS beliefs $\mathbf{Beliefs}$*
 - *where \mathbf{Facts} and $\mathbf{Beliefs}$ are EPS conditions occurring in EPSs in $\text{Ran}(I)$ according to definition 30.*

Next, I define a set of functions that assigns sets of (tuples of) agents and/or things and times to subsets of I as specified in definition 35. We have to consider that we should define these functions not only with respect to one specific agent x , but also for other agents τ (those agents for which x has started an instance of the BDI-interpreter). Thus the following functions have arguments for agents and times.

Definition 36 *EPS assignment functions*

- *A function \mathbf{B} that assigns $\mathbf{Beliefs}$ to an agent τ at t :*
 $\mathbf{B}(\tau)(t)$
- *A function \mathbf{F} that assigns \mathbf{Facts} to t :*
 $\mathbf{F}(t)$
- *A function \mathbf{T} that assigns EPS structures to an agent τ at t , i.e. the time structure of an agent at t :*
 $\mathbf{T}(\tau)(t)$
- *A function \mathbf{P} that assigns \mathbf{Plans} to an agent τ at t :*
 $\mathbf{P}(\tau)(t)$
- *A function \mathbf{S} that assigns $\mathbf{Scenarios}$ to an agent τ at t :*
 $\mathbf{S}(\tau)(t)$

- A function P that assigns **Properties** to (tuples of) EPS-things $\langle a_1, \dots, a_n \rangle$ at t :
 $P(a_1, \dots, a_n)(t)$

From a given BDI-interpreter instance for an agent x , we can define the following functions:

- A function **Attitudes** that assigns attitudes of a certain type ϕ (Desires or Intentions) to an agent x at t :
 $\mathbf{Attitudes}(\phi)(x)(t)$
- A function **Abilities** that assigns feedbacks K of the plan library available to an agent τ at t :
 $\mathbf{Abilities}(\tau)(t)$

6.2.5 SMS, EPS and the BDI-interpreter

In this section, we have to discuss in more detail how the SMS and EPS are related to the BDI-interpreter and the object recognition engine. Recall the main loop of the BDI-interpreter (definition 37).

Definition 37 *BDI-interpreter main loop*

```

1 initialize-state;
do
2 options:=option-generator(trigger-queue,Beliefs,Goals,Intentions);
3 b-add(options, EPS(now));
4 update(IRS);
5 selected-options:=deliberate(options,Beliefs,Goals,Intentions);
6 update-intentions(selected-options,Intentions);
7 execute(Intentions);
8 get-new-SMS;
9 if f-update(SMS) then f-add(SMS,EPS(now));
10 update(IRS);
11 get-new-triggers(EPS,IRS);
12 drop-successful-attitudes(Beliefs,Goals,Intentions);
13 drop-impossible-attitudes(Beliefs,Goals,Intentions);
until quit.
```

Adding facts to the EPS

I start with a discussion of the relation between the SMS and the EPS. The SMS structure is related to the EPS structure in that the factual part of the EPS structure (i.e. the linear scenario leading to the present state of affairs) is determined by the robot's perception of reality, i.e. the output of the object recognition engine in terms of SMSs and SMS transitions. The main loop of the BDI-interpreter translates new information from the SMS (remember the discussion on EPS updates in section 4.2.3) to an EPS. This is done with lines 8 and 9 of the main loop of the BDI-interpreter stated in note 14.

Note 14 *Updating the EPS with SMS data*

```

do
...
8 get-new-SMS;
9 if f-update(SMS) then f-add(SMS, EPS(now));
...
until quit.

```

f-update (introduced in chapter 4.2.3) decides whether a given SMS expression should effect a change in the EPS. In case of incompatibilities between new and old information at the EPS-level the newer information replaces the old information, if the information is completely new to the EPS-level (i.e. an object recognized for the first time) the new information is entered into the EPS. The actual addition of new information at the EPS-level is executed by the procedure **f-add(SMS, EPS(now))**; . In the next section, I spell out this procedure in more detail. For this purpose, I introduce a translation function \rightleftharpoons that specifies how SMS property terms translate to EPS conditions. We must also take care that SMS object terms are translated to EPS things together with appropriate anchors and we must ensure that the time indices of the EPS are appropriately anchored in the respective SMS snapshot indices.

Translation of SMSs to EPSs

The following definition spells out a possible formulation of the procedure **f-add(SMS, EPS, now)** of the main loop of the BDI-interpreter. Please note that the following concerns only the translation of single SMS snapshots to single EPSs and that SMS transitions are discussed in the next section.

Definition 38 *Translation of SMS snapshot indices, SMS object terms and SMS property terms to EPS times, conditions and external anchors.*

Given a SMS S with index s_i resulting from the output of the object recognition engine and a given EPS structure $\langle \mathbf{T}, I, \text{Actions} \rangle$ with distinguished *now*, translate S to a new EPS as follows:

- (1) Translate the SMS snapshot index s_i by the introduction of a reference marker for an EPS time t_m , where m increases the time index of the last EPS time $t_{m-1} = \text{now}$ by one, and add an external anchor $\langle t_m, s_i \rangle$:

$$s_i \rightleftharpoons \boxed{\begin{array}{l} t_m \\ \langle t_m, s_i \rangle \end{array}}.$$

- (2) For each object term $\alpha \in S$, introduce an EPS reference marker for a thing a and an external anchor $\langle a, \alpha \rangle$:

$$\alpha \rightleftharpoons \boxed{\begin{array}{l} a \\ \langle a, \alpha \rangle \end{array}}.$$

- (3) Given that the object terms $\alpha, \dots, \omega \in S$ have been translated in the previous step, translate SMS property terms for object types concerning an object term $\alpha \in S$ to an EPS condition as follows:

- $\alpha : O - UNDEFINED \rightleftharpoons \text{object} - \text{undefined}(a)$
- $\alpha : O - CUBE \rightleftharpoons \text{cube}(a)$
- $\alpha : O - NUT \rightleftharpoons \text{nut}(a)$

- $\alpha : O - SCREW \Leftrightarrow screw(a)$
- $\alpha : O - SLAT \Leftrightarrow slat(a)$
- $\alpha : O - PERSON \Leftrightarrow person(a)$

(4) Given that $\alpha, \beta, \gamma \in S$ are object terms that have been translated in the previous step to a, b, c with the external EPS anchors $\langle a, \alpha \rangle, \langle b, \beta \rangle, \langle c, \gamma \rangle$, introduce a new EPS thing reference marker d^2 and translate SMS property terms to an EPS condition along the lines of the following translation:

- $\alpha : C - RED \Leftrightarrow d : red(a)$
- $\alpha : C - GREEN \Leftrightarrow d : green(a)$
- $\alpha : L - TABLE \Leftrightarrow d : on - table(a)$
- $\alpha : L - LEFTHANDROBOT \Leftrightarrow d : lefthandrobot(a)$
- $\alpha : L - RIGHTHANDROBOT \Leftrightarrow d : righthandrobot(a)$
- $\alpha : L - HANDROBOT \Leftrightarrow d : handrobot(a)$
- $\alpha : L - LEFTHANDHUMAN \Leftrightarrow d : lefthandhuman(a)$
- $\alpha : L - RIGHTHANDHUMAN \Leftrightarrow d : righthandhuman(a)$
- $\alpha : L - HANDHUMAN \Leftrightarrow d : handhuman(a)$
- $\alpha : P - BETWEEN(\beta, \gamma) \Leftrightarrow d : between(a, b, c)$
- $\alpha : P - CORNER(\beta) \Leftrightarrow d : pos - corner(a, b)$
- $\alpha : S - BOLTED(\beta) \Leftrightarrow d : bolted(a, b)$
- $\alpha : S - BOLTED(\beta, \gamma) \Leftrightarrow d : bolted(a, b, c)$

Translation of SMS transitions to external EPS actions

In the next step, I define how SMS transitions are related to EPS actions. Transitions between SMSs translate to external actions of the EPS structure (one part of the set *Actions*) along the lines of the following definition 39. The other way round, the execution of a plan in which external actions occur retranslates those external actions to SMS transitions in order to trigger motor actions of the robot.

Definition 39 SMS transitions and their translation to external EPS actions

Given that α, β, γ are object terms that have been translated with the external EPS anchors $\langle a, \alpha \rangle, \langle b, \beta \rangle, \langle c, \gamma \rangle$, SMS transitions translate as follows to external EPS actions:

- $A - UNDEFINED \Leftrightarrow ext-a:undefined$
- $A - HOLD^2(\alpha, PROPERTY - TERM) \Leftrightarrow ext-a:hold(a, \Leftrightarrow PROPERTY-TERM)$
- $A - GRASP(\alpha) \Leftrightarrow ext-a:grasp(a)$
- $A - POINT(\alpha) \Leftrightarrow ext-a:point(a)$

²In the discussion of examples, I omit EPS reference markers for EPS relations and properties as far as they are not directly involved in the analysis.

- $A - SCREW^2(\alpha, \beta) \Leftrightarrow \text{ext-a:screw}(a,b)$
- $A - SCREW^4(\alpha, \beta, \gamma, \text{PROPERTY} - \text{TERM}) \Leftrightarrow \text{ext-a:screw}(a,b,c, \Leftrightarrow \text{PROPERTY-TERM})$
- $\alpha : UTT_n : 'Utterance' \Leftrightarrow a-a : UTT_n, K$ where K is an IRS representing the semantic content of UTT .
- $\alpha : UTT_n : 'Utterance' \Leftrightarrow \text{ext-a} : \text{say}(K)$, where K is an IRS from which 'Utterance' is generated.

Adding BDI-states to the EPS

The last EPS time (in the sense of $<$) which is anchored in a SMS snapshot index constitutes the division between the linear non-branching past and the (possibly) branching future - the agent's current now. In the last section, we have discussed the relation between the linear part of the EPS and the SMS. The branching part of an EPS structure is contributed by the output of possible options of future action of the BDI-interpreter. Each cycle of the BDI-interpreter main loop modifies the current branching time structure of the EPS structure according to the agent's current configuration of Beliefs, Desires and Intentions. Consider the lines 2 and 3 of the main loop of the BDI-interpreter (note 15).

Note 15 *Adding options to the EPS*

```
do
...
2 options:=option-generator(trigger-queue,Beliefs,Goals,Intentions);
3 b-add(options, EPS(now));
...
until quit.
```

With these two lines (at the beginning of the loop), the BDI-interpreter adds generated options of future action to the agent's current EPS structure at the agent's current now as determined within the previous cycle. The addition of future options does not require a special translation function, as plans and related information structures of the BDI-interpreter such as beliefs or intentions are defined in terms of the EPS (see definition 14). Impossible options of action or successfully realized intentions are removed from the BDI-interpreter's information structures and the current EPS via the lines 12 and 13 in the BDI-interpreter main loop (note 16). Consequently, an agent is always be provided with up-to-date information of her internal and external states of affairs.

Note 16 *Dropping outdated attitudes from BDI and EPS*

```
do
...
12 drop-successful-attitudes(Beliefs,Goals,Intentions);
13 drop-impossible-attitudes(Beliefs,Goals,Intentions);
...
until quit.
```

6.2.6 The set of atomic actions *Actions*

The translation of SMS transitions make up only some of the actions that are part of the EPS vocabulary. Definition 40 gives those further actions that play a role in the next part of this thesis devoted to the analysis of example discourses.

Definition 40 *The set of atomic actions $Actions = Internal - Actions \cup External - Actions$ listed in definition 40 includes those actions which occur in the analysis of examples in the second part of this thesis. In a realistic setup, the set of atomic actions will be considerably larger. For obviously internal actions such as *i-add* or *g-add*, for the sake of readability I omit the prefix *int-a*: in the presentation of examples in the next chapter.*

The set $Internal - Actions$ consists of the following items, where EPS may be either an EPS or a EPS structure.

- *b-add(EPS), b-remove(EPS) (remove and add beliefs)*
- *g-add(EPS), g-remove(EPS) (remove and add goals)*
- *i-add(EPS), i-remove(EPS) (remove and add intentions)*
- *provable(EPS , type of database) (try to unify the specified EPS with the specified database (either facts or beliefs))*
- *?variable=value //(check variable for value)*
- *yes, no //* (result of queries)
- *for x in l loop ... end loop*
- *if ... then*
- *set(variable=value), set(flag)*
- *variable=value //(I use this as a short notation for if variable=value then execute this branch)*
- *?K //Interpret K, see section 6.3.2*
- *has-interpretation(K), no-interpretation(K) //Interpretation result of an IRS K, see section 6.3.2*
- *update(IRS, EPS) // construct an IRS from the specified EPS*

The set $External - Actions$ consists of the following items:

- *undefined*
- *hold(a, \Rightarrow PROPERTY-TERM)*
- *grasp(a)*
- *point(a)*
- *screw(a,b)*

- $screw(a,b,c, \Leftrightarrow \text{PROPERTY-TERM})$
- UTT_n, K
- $say(K)$

6.3 The IRS: Internal Representation Structures

We now have all the necessary material at hand to give a formal definition of IRSs. In the following, I define an IRS language that is capable of handling the results of our discussion of

- thing-individuals (section 3.2)
- time-individuals (section 3.3.3)
- belief and knowledge (section 4.1.2)
- the temporal location of time-individuals with respect to the present now, location times and other time-individuals (section 3.3)
- anchors (section 3.2.2)

However, it should be noted that this constitutes only the basic applications of the language of IRSs. See section 8.3 for possible extensions.

6.3.1 Syntax of IRSs

The ontology of the IRS includes two sorts of entities: thing-individuals and time-individuals, where time-individuals consist of the subsorts of events and states. The IRS vocabulary thus has reference markers for each of these sorts and subsorts.

Definition 41 *IRS vocabulary*

- $Ref_{thing} = \{x, \dots, z, \dots, i\}$: reference markers for thing-individuals and the 'i'-constant.
- $Ref_{time} = \{e_0, \dots, e_n, \dots, s_0, \dots, s_n, \dots\}$: reference markers for time-individuals - events e and states s .
- The individual constant for 'now': n
- I call the set $Ref_{thing} \cup Ref_{time}$ the set of reference markers Ref .
- A set of one-place relation constants for handles: *Handles*
- A set of n -place predicate constants: Rel_n for each $n \geq 1$
- A set R_{pred} of 2-place predicate symbols: $R_{pred} = \{<_{beg}, \prec, =, \subset, \subseteq\}$
- A set of logical symbols: *Logical – Symbols* = $\{\vee, \wedge, \Rightarrow\}$
- A set of anchor sources for thing-individuals $Sources = \{a_1, \dots, a_n, \dots\}$

- A set *Timepoints* = $\{t_1, \dots, t_n, \dots\}$: temporal anchor sources for the now-constant
- A set *Source – Symbols* = $\{?_1, \dots, ?_n, \dots, !_1, \dots, !_m, \dots, ?a_1^1, \dots, ?a_l^k, \dots\}$: anchor source variables
- A set *Handle – Placeholders* = $\{?_1, \dots, ?_n\}$: placeholders for handles³
- A one-place operator \rightarrow that can be applied to a given anchor source or anchor source symbol τ as definiteness constraint $\overrightarrow{\tau}$
- A set of brackets $\{[,], (,)\}$ and question symbols $\{?_1, \dots, ?_n, \dots\}$

Definition 42 *Syntax of IRSs, IRS-conditions and IRS-anchors*

- If *Universe* \subseteq *Ref*, *Conditions* a (possibly empty) set of conditions and *Anchors* a (possibly empty) set of anchors then $\langle \text{Universe}, \text{Conditions}, \text{Anchors} \rangle$ is an IRS.
- If $N \in \text{Handle}$ and $x \in \text{Ref}$ then $N(x)$ is a condition.
- If $x \in \text{Ref}$ and $? \in \text{Handle – Placeholders}$ then $?(x)$ is a condition.
- If $ev_1, ev_2 \in \text{Ref}_{time}$, $R \in \text{R}_{pred}$ then $ev_1 R ev_2$ is a condition.
- If K is an IRS then $\neg K$ is a condition.
- If K_1 and K_2 are IRSs then $K_1 \vee K_2$ is a condition.
- If K_1 and K_2 are IRSs, then $K_1 \Rightarrow K_2$ is a condition.
- If K is an IRS, $x \in \text{Ref}_{thing}$, $s \in \text{Ref}_{time}$ then $\langle s, xBK \rangle$ is an anchor.
- If K is an IRS, $x \in \text{Ref}_{thing}$, $s \in \text{Ref}_{time}$ then $\langle s, xK_t K \rangle$ is an anchor.
- If K is an IRS, $x \in \text{Ref}_{thing}$, $s \in \text{Ref}_{time}$ then $\langle s, xK_h K \rangle$ is an anchor.
- If $R \in \text{Rel}_n$, $x_1, \dots, x_n \in \text{Ref}_{thing}$, $s \in \text{Ref}_{time}$, then $\langle s, R(x_1, \dots, x_n) \rangle$ is an anchor.
- If K is an IRS, $x \in \text{Ref}_{thing}$, $e \in \text{Ref}_{time}$, then $\langle e, x\text{CAUSE}K \rangle$ is an anchor.
- If K is an IRS, $x \in \text{Ref}_{thing}$, $e \in \text{Ref}_{time}$, then $\langle e, x\text{DOK} \rangle$ is an anchor.
- If K is an IRS, $x \in \text{Ref}_{thing}$, $e \in \text{Ref}_{time}$, then $\langle e, x\text{INT}K \rangle$ is an anchor.
- If $x \in \text{Ref}_{thing}$, $a \in \text{Sources}$ then $\langle x, a \rangle$, is an anchor.
- If $x \in \text{Ref}_{thing}$, $?, !, ?a \in \text{Source – Symbols}$ then $\langle x, ? \rangle$, $\langle x, ! \rangle$, $\langle x, ?a \rangle$ are anchors.⁴
- If $\langle x, \tau \rangle$ is an anchor, then $\langle x, \overrightarrow{\tau} \rangle$ is an anchor.

³The question mark '?' is overloaded in that it has different meaning in different contexts.

⁴Recall that in note 3, section 3.2.2, I specified the following three types of variable anchor sources for IRS thing-individuals:

- (1) $\langle \text{floater}, ! \rangle$ external anchor
- (2) $\langle \text{floater}, ? \rangle$ internal anchor
- (3) $\langle \text{floater}, ?a \rangle$ anaphoric anchor

- If $x \in Ref_{thing}$, K an IRS, then $\langle x, K \rangle$ is an anchor.
- If $x, y \in Ref_{thing}$ then $\langle x, y \rangle$ is an anchor.
- If $t \in Timepoints$ then $\langle n, t \rangle$ is an anchor.
- If $t_1, t_2 \in Timepoints$ then $\langle n, [t_1, t_2] \rangle$ is an anchor.

6.3.2 Semantics of IRSs

This section formally defines the semantic relation between given IRSs and a given EPS structure. This definition integrates the outcomes of our earlier discussions on reference, models, meaning, representations and presentations. I will briefly recapitulate these conclusions in the next paragraphs and give some general remarks on the integration of non-classical logics into the framework of GDRT.

A preliminary specification of IRS interpretations

From dynamic meaning to dynamic interpretation As discussed in section 2.3.1, the concept of meaning provided by static truth-conditional semantics does not capture the difference between meaning, informativity and successful interpretation of logical forms of utterances. DRT tries to remedy some of these shortcomings by providing a notion of informativity in terms of the contribution that meaningful utterances make to the incremental construction of semantic representations for larger units. Nevertheless, I argued in section 2.3.1, DRT does not provide a concept of dynamic interpretation: it is not part of the formalism of DRT (and this applies to newer versions of DRT too) to spell out in detail e.g. how the accommodation of presuppositional anchors is to be processed by an agent and how the transformation of preliminary to fully specified DRSs is to be achieved. It is here that the specific anchor-based relation between the IRS and the EPS comes into play. A semantic binding in terms of the incremental resolution of the variable anchor sources of an IRS constitutes one part of the dynamics of IRS interpretation. Each identification of a variable IRS thing-individual anchor source decreases the number of possible interpretations of an IRS in that it fixes the possible referents of a given variable thing reference marker. Only IRSs for which all occurring variable thing anchor sources have been resolved to at least one referent (several resolutions are possible) are passed over to the second step of IRS interpretation in terms of pragmatic profiling. Pragmatic profiling, the second contribution to dynamic interpretation, comes into play when IRS conditions involving time-individuals are identified. Two different modes for the identification of time-individuals were introduced in section 5.3.4: plain and reactive mode. Plain interpretation mirrors the traditional concept of descriptive truth-conditional semantics, whereas reactive interpretation involves a prescriptive pragmatics⁵. That is, while the dynamics of DRT only resides in the incremental construction of representations, the dynamics of IRSs also manifests itself in the incremental identification of thing- and time-individuals, thus extending the dynamics of representation to a dynamics of staged reference resolution where all the usually implicitly assumed procedures are explicitly spelled out⁶. In the technical jargon of DRT, the explicit representation and resolution of anchors allows for a dynamic conception of a 'verifying embedding' (which I call a successful anchoring) of an

⁵The two modes of interpretation are quite similar to Dov Gabbay's distinction between declarative and imperative semantics [Gabbay, 1987], a point that I will elaborate in more detail in the next section.

⁶Among others, it is in particular these points that distinguishes the approach proposed here from newer versions of DRT concerned with presupposition resolution.

IRS into a model: as the interpretation of an IRS proceeds, more and more possible 'embeddings' resp. anchorings are ruled out as variable anchor sources are resolved with respect to the pragmatic profiles of the time-individuals in which they are involved.

Variable vs. defined anchor sources In the light of these considerations, we need to decide whether to start from variable or defined thing-anchor sources when defining a semantic relation between IRSs and the EPS structure. DRT considers the case of variable anchor sources as basic for the definition of semantic truth. According to this conception, DRS reference markers are related to model-theoretic entities via a function that iterates through possible combinations of reference markers and model-theoretic entities in order to identify 'verifying embeddings' of affected conditions with respect to an interpretation function. Fixed assignments, based on resolved anchors, have somehow to be integrated into this procedure. As a consequence, the dynamics of information update and that of interpretation is mixed together, with the effect that the dynamics of reference resolution is hidden inside the treatment of information update. The given setup of IRSs containing EPS-anchors approaches the problem from the opposite direction by taking defined anchor sources as starting point, thereby separating the two interpretation steps of anchor source resolution and model theoretic evaluation.

Plain IRS interpretation In order to illustrate what is spelled out in the following - the semantic and pragmatic relation between the IRS and the EPS - it is useful to recapitulate the overall architecture of the structures we have defined so far. Given an agent x at time t_i , I call the agent's configuration of her SMS, EPS and IRS at t_i the 'cognitive structure' CS of x at t_i , $CS(x)(t_i) = \langle IRS(x, t_i), EPS(x, t_i), SMS(x, t_i) \rangle$. Suppose that $\langle t_j, K \rangle$ is an IRS belonging to $IRS(x, t_i)$ - normally t_j would be the instant 'now' and that is what I will assume here - and that x attempts to interpret K at t_i . As a first step x must find anchor sources for all those thing anchors of the anchor set $Anch$ of K where anchor sources are variable. Each such variable anchor source is of one of the three sorts: (1) internal, (2) anaphoric or (3) external. If s is a variable internal anchor source, then a (non-variable) anchor source that can replace it must be a thing reference marker from $EPS(x, t_i)$; when the variable anchor source is anaphoric, then a non-variable anchor source replacing it must be a thing-individual from $IRS(x, t_i)$; when the variable source is external, then an anchor source to replace it must be a thing reference marker from $EPS(x, t_i)$ together with an anchor $\langle a, \alpha \rangle$ belonging to $EPS(x, t_i)$ where α is an object term belonging to $SMS(x, t_i)$. If for any variable thing-anchor no suitable sources can be found, then the interpretation of K aborts. Suppose that it is possible to find a suitable non-variable anchor source to replace each of the variable thing anchor sources occurring in anchors of K . Then there will be a nonempty set G of functions g each of which is defined on the set of 'floaters' of anchors in K and maps each floater onto a suitable non-variable anchor source. Each g in G can then be used to identify the time-individuals from $\langle t_j, K \rangle$ and if this succeeds, this g is stored in a set $F \subseteq G$ to check in a final step whether K as a whole has at least one successful anchoring $H \subseteq F$ that renders possible to identify all conditions $c_1, \dots, c_n \in K$ (in particular the complex conditions) with respect to $EPS(x, t_i)$ at t_j . If there exists such a successful anchoring for K , I say that K has a successful plain interpretation.

Reactive IRS interpretation If the plain interpretation of K as described in the last section fails, i.e. no successful anchoring of K could be established, reactive interpretation comes into play. It could

be the (pragmatic) meaning of K (as discussed in section 5.3.4) that $EPS(x, t_i)$ has to be changed by the interpreter of K to $EPS(x, t_k)$ with $i < k$ in order to render possible a successful anchoring of K in $EPS(x, t_k)$. The appropriate reaction, so I argued in this thesis, is guided in particular by the time-individuals contained in K . These time-individuals specify a course of action which is to be undertaken in order to bring about the conditions that render a successful plain interpretation of K possible. That is, in response to a failed plain interpretation of K with respect to $EPS(x, t_i)$ at t_j , the interpreter of K should perform some actions which result in $EPS(x, t_i)$ being transformed to an EPS structure $EPS(x, t_k)$ which allows for a successful plain interpretation of K at t_k . Technically, this requires that we formulate a semantic *and* a pragmatic identification of time-individuals - we need to specify the conditions that identify time-individuals in plain interpretation mode (corresponding to classical truth-conditional semantics) and in addition the actions which are to be undertaken in order to make a given time-individual 'true' via an execution of reactive interpretation (corresponding to the unfolding of the pragmatic impact of time-individuals)⁷.

The logic behind the interpretation of IRSs

The default solution to IRS interpretation In the intended application scenario of this thesis, an agent x will always find herself in a situation defined by her cognitive state $CS(x)(t)$ and it is this specific situation against which she is supposed to evaluate her IRSs by default⁸. For example, we (as designers) do not want plans to be triggered on the basis of information not contained in $CS(x)(t)$. Similarly, we do not want that the agent interprets an utterance such as "Show me all red cubes" as involving quantification over an infinite set of possibly existing red cubes but as pertaining to the externally anchored red cubes provided by $CS(x)(t)$. However, the design of IRS interpretation as a staged process and the separated treatment of reference resolution and semantic evaluation supports the implementation of 'switches' between this situation-bounded IRS interpretation and non-situation-bounded IRS interpretation. At each level of IRS interpretation - thing-individuals, time-individuals and complex IRS conditions - the agent can adopt different logical attitudes towards the interpretation of IRSs. That is, depending on the situation, different logics (classical or non-classical) may be employed for the semantic interpretation of IRSs. The default attitude of situation-bounded IRS interpretation I adopt in the following is a solution that fits specifically to the application scenario of this thesis. In the next paragraphs, I give a (necessarily simplifying) overview of the logical attitudes that are in principle available to an agent for IRS interpretation and introduce the solution that is implemented with the formal definition of IRS semantics in the next section.

Classical and non-classical logic The problem we are faced with when developing a formal semantics for the IRS language is that in the application scenarios relevant to the goals of this thesis, an agent *de facto* evaluates her IRSs against EPS-based models (presentations of reality) that are *incomplete*. EPS-based models can be incomplete in several ways. First, the extensions (the referents) of IRS thing-individuals may be unknown to an agent and thus missing in the agent's EPS-based modeling of reality.

⁷The distinction between plain and reactive interpretation comes close to what [Sperber and Wilson, 1993] call the distinction between conceptual and procedural meaning. However, the actual realization of these two different conceptions of meaning proposed here takes another route than the relevance based approach of [Sperber and Wilson, 1993].

⁸In the semantic conception of evaluation against specific situations, GDRT shares similarity to the information limitation proposed by situation semantics [Barwise, 1981, Barwise and Perry, 1983]

Second, the extensions of IRS time-individuals may be unknown to an agent and thus are not contained in the agent's EPS-based models. Third, the EPS-based models against which an agent evaluates IRSs involving quantification over thing-individuals are finite with respect to the domain over which the quantifier ranges. At first sight, this situation seems to be in conflict with the fundamental assumption of bivalent formal semantics that models are *complete* in that they include all information that is relevant with respect to evaluation; with respect to complete models, a sentence evaluates to either true or false. With respect to incomplete models, there exists a third possibility: because of the information that the model lacks it is neither possible to determine the sentence as true or false. Incorporation of this third possibility in the definition of the semantic relationship between sentences and model structures leads to a non-classical, three-valued logic.

Incomplete information: thing-individuals Begin with IRS thing-individuals. It may well be the case that a thing-individual y occurring in an IRS K to be evaluated against an EPS-based model M has no model-theoretic counterpart, that y lacks an extension. The interpretation process of IRSs outlined above starts with the resolution of variable anchor sources of thing-individuals occurring in an IRS K . Either the resolution of the thing-individuals in K succeeds - then a semantic binding of the IRS to the EPS-based models of an agent has been established and each thing-individual has been provided an extension - or this task fails. In the case of failure, a plan for error correction is triggered with the goal to gather more information or to add a new EPS-entity that renders possible to resolve the anchor source of the respective thing-floater. As long as the respective variable anchor sources have not been resolved to model-theoretic entities, the interpretation process of K as a whole is interrupted. That is, only if all variable thing-individuals occurring in an IRS K are resolved and thus are provided with an extension, K is passed over to the next step of interpretation, the resolution of anchors for time-individuals. If we would like to continue the interpretation process with unresolved anchor sources of thing-individuals, in the further course we need to employ a non-classical logic. A reasonable approach would be supervaluation semantics [van Fraassen, 1966], which renders possible to deal with unknown extensions of singular nouns (here corresponding to thing-individuals). This would also drop the limitation of situation-bounded interpretation in that it considers evaluation against all possible situations in the definition of the semantic relation between sentences and model structures.

Incomplete information: time-individuals Once all thing-individuals occurring in an IRS K have been resolved, in the second step of IRS interpretation, the time-individuals of K are interpreted via the resolution of their anchors. The resolution of the anchor of a time-individual may succeed or not. In the case it does not succeed, two further options exist, depending on the interpretation mode of the IRS. If the IRS is interpreted in plain mode, the IRS containing the time-individual has no interpretation and there is no possibility to bring about an interpretation as in the case of reactive IRS interpretation, where the interpreting agent seeks to extend her EPS structure such that a plain interpretation is rendered possible. As with thing-individuals, only if all the anchors for IRS time-individuals have been resolved, the processing of the respective IRS proceeds. Otherwise the IRS interpretation process is interrupted until the necessary manipulations of the EPS structure have been executed. That is, only IRSs with resolved time-individuals (and resolved thing-individuals) are passed over to the next step of interpretation. If we would like to continue the process of IRS interpretation with unresolved time-individuals (but resolved

thing-individuals) and without restriction to specific situations, then a non-classical logic such as the three-valued logic proposed by [Kleene, 1952] could be employed to deal with unknown extensions of predicate constants (here: time-individuals).

Incomplete information: complex IRS conditions The third step of IRS interpretation is reached iff all anchors of the IRS to be interpreted have been resolved. With this step, we reach the level of complex IRS conditions (corresponding to the level of sentences of a logical language) involving the interpretation of quantification and the logical constants. With the set of resolved anchors, we have a *complete* specification of the thing-individuals and time-individuals occurring in an IRS K and consequently can make use of a classical logic for the semantic definition of the logical constants and quantification. This default solution - allowing only fully resolved IRS to pass through the interpretation process - can be altered if necessary. But if we drop the requirement that only fully resolved IRSs are passed over to the third step of interpretation, no longer can we make use of a classical logic but must employ a non-classical logic. The approach to evaluation against incomplete information implemented in GDRT can be thought of as a filter that sorts out those IRSs for which the information required for interpretation is incomplete. A final case of incomplete information has to be discussed. By default, the domain of IRS quantifiers is restricted to the closed world of an agent's finite EPS structures. However, the agent should in principle be able to interpret the transcendental nature of quantification over potentially infinite sets. In this case, the default assumption of a closed world restriction on the quantifier domain must be dropped in favor of a potentially infinite domain of quantification. This logical attitude of switching between finite and infinite quantification domains is implemented in GDRT with the definition of IRS quantifier semantics, where we make use of an inflated model $\uparrow M$ obtained from a finite EPS model M by adding a finite or infinite set of things Unk (for Unknown) to the set of EPS *Things*: $\uparrow Things = Things \cup Unk$ (where an agent does not need to know anything more about Unk besides the fact that it contains an infinite set of things). Another possibility, not implemented here, would be to make use of the ontologically neutral definition of quantification with parametric extensions [Bonevac and Kamp, 1987].

Interpretation of IRSs

This section defines the set-theoretic formal concepts involved in the interpretation process of an IRS. I start with the definition of models for IRS interpretation.

Models for IRS interpretation In defining the models for IRS interpretation we have to consider an important point that distinguishes the models for IRS interpretation from the models for e.g. the language of DRSs. As the models an agent can employ for IRS interpretation are derived from the agent's *current* EPS, those models only present the agent's current information about the state of affairs. The 'indexed' nature of the models for IRS interpretation is captured by recording the agent from whose EPS the model was derived and the time at which this was done. Mindful of this consideration, a model $\uparrow M$ for the semantic definition of the language of IRSs could be defined as follows.

Definition 43 *A model $\uparrow M$ at a time t of an agent x is a 10-tuple*

- $\uparrow M(x)(t) = \langle P, S, T, P, F, B, PRG, \uparrow Things, Attitudes, Abilities \rangle$

The sets **Attitudes**, **Abilities** and functions $P, S, \mathbf{P}, \mathbf{F}, \mathbf{B}, \mathbf{T}, \mathbf{PRG}$ of $\uparrow M$ are those of an agent's EPS at t as defined in section 6.2.4 and the agent's knowledge base as defined in chapter 7 and $\uparrow \mathbf{Things}$ is the inflation of the set of EPS things **Things**. In the following, I omit the arguments x and t of a model $\uparrow M$ and suppose that it is always the current model of the interpreting agent which is used for the interpretation of IRSs.

Successful anchoring $\uparrow M$ differs from traditional models (e.g. those for the interpretation of DRSs) However, the overall account of dynamic semantics in DRT is adopted in the following.) in that there is no 'interpretation function' included, i.e. a predefined function that maps predicates and individual constants to their model-theoretic counterparts. Instead, the concept of an 'interpretation function' is replaced by two components. First, this is the **PRG** function, which contains the pragmatic profiles associated with the handles of reference markers in the respective lexical entries. The identification conditions provided by the $[PRG]$ part of the lexical entry for a individual guide the transformation of an IRS with variable thing anchor sources to an IRS with a set of possible thing individual anchor source resolutions. The other component is the identification procedure for time-individuals, which determines at the runtime of the interpretation algorithm the 'extension' of time-individuals in that it identifies sets of thing individual anchor sources among the set of possible thing anchor resolutions that satisfy the identification conditions $[PRG]$ of the time-individuals with respect to $\uparrow M$. Finally, it must then be determined which of the possible anchorings render possible a successful anchoring of the IRS as a whole⁹. In addition, we have to consider the fact that adding an IRS to the main IRS may result in a revised set of referents, conditions and anchors. I thus have to define the notion of a successful anchoring with respect to an existing anchoring. That is, suppose an IRS K consisting of

- a set of reference markers $U_K \subseteq Ref$
- a set of conditions $Con_K = \{C_1, \dots, C_n\}$
- a set of anchors $G = \{A_1, \dots, A_m\}$

is given. An update to K will result in the sets

- U_K^{update}
- Con_K^{update}
- G^{update}

where G^{update} is a successful anchoring of K with respect to M iff G^{update} extends G in a way that it identifies the floaters in U_K^{update} with EPS entities given the pragmatic identification conditions $[PRG]$ associated with the floaters. Formally, a successful anchoring is defined as follows:

Definition 44 *Successful anchoring of an IRS K*

Given sets of possible anchorings G, H , a model $\uparrow M$ and an IRS K

- $\langle\langle G, H \rangle\rangle \vDash_{\uparrow M} K$ iff $G \subset_D H$ and for all $\gamma \in Con_K : H \vDash_{\uparrow M} K$, where $A \subset_U B$ reads as: 'the domain of A is a subset of the domain of B '

⁹The set of anchors which make up a successful anchoring corresponds to the DRT concept of a verifying embedding of a DRS K .

- $G \vDash_{\uparrow M} K$ reads as: G successfully anchors K in $\uparrow M$ and
- $\langle\langle G, H \rangle\rangle \vDash_{\uparrow M} K$ reads as: H extends G to a successful anchoring of K in $\uparrow M$.

Definition 45 *Successful interpretation of an IRS K*

- An IRS K has a successful interpretation in a model $\uparrow M$ iff there exists a successful anchoring G for K in $\uparrow M$ that extends the empty anchoring ξ .
- I write $\vDash_{\uparrow M} K$ iff there exists a successful anchoring G such that $\langle\langle \xi, G \rangle\rangle \vDash_{\uparrow M} K$.
- When $G \vDash_{\uparrow M} \gamma$, where γ is an IRS-condition, I say that G identifies γ in $\uparrow M$.
- In addition, interpretations can be determined with respect to a time, a scenario, a plan and a model, which will be written as $\vDash_{\uparrow M, S, P, t} K$.
- In the EPS (e.g. as part of a plan), an interpretation attempt of an IRS K can be triggered by the command $\text{int-}a :?K$. The algorithmic specification of this command is given in section 6.3.2.
- The EPS constituents which were identified as a successful interpretation of an IRS K with respect to a model and a time are denoted by $[K]_{\uparrow M, t}$. If the respective EPS constituents have not been identified yet, $[K]_{\uparrow M, t}$ triggers an interpretation attempt of K , $\text{int-}a :?K$.

Identification and interpretation of IRS-Conditions

In section 2.1, I argued that the processing of reference is about *identifying* intended entities from a given modeling of context and that it is the pragmatics $[PRG]$ of the IRS individual in question which provides the necessary information for identifying the individual. Consequently, the identification conditions for thing-individuals are straightforward as in definition 46.

Definition 46 *Identification of thing-individuals*

- $\langle x, a \rangle \vDash_{\uparrow M, t} \text{handle}(x)$ iff $PRG_{\text{handle}}(x) \in \mathbf{P}(a)(t)$

Semantic Binding: Identification of variable anchor sources This section deals with the issue of resolving variable thing-anchor sources. The process of accommodating variable thing-anchor sources is part of the ‘semantic binding’ of an IRS K in that it delivers the set of possible resolutions of variable thing-anchor sources. Variable thing anchor sources emerge in several cases. First, as discussed in section 2.3, it is part of the semantic processing of *utterances* that objects of reference intended by the speaker are determined. An IRS representation of an utterance comes with no such determination but the anchor sources of the involved thing-anchors have to be resolved to constituents of the hearer’s model viz. her EPS. That is, an IRS K constructed from an utterance has no defined anchor sources for the thing-anchors which occur in K (see section 5.3.4). Instead, it is part of the semantic processing of utterances to determine these bindings. Second, variable thing-anchor sources can occur in *non-utterance related* IRSs as they occur e.g. as part of an agent’s planning processes.

In note 3, section 3.2.2, I specified the three following types of variable anchor sources for thing-individuals:

- (1) $\langle \text{floater}, ! \rangle$ The source is not yet identified, but the floater must be externally anchored, i.e. in a fact.
- (2) $\langle \text{floater}, ? \rangle$ The source is not yet identified, but the floater must be internally anchored, i.e. in a belief.
- (3) $\langle \text{floater}, ?a \rangle$ The source is not yet identified, but the floater must be anaphorically anchored in an existing IRS-individual.

I propose to treat the resolution of variable thing-anchor sources with the help of plans which are triggered by occurrences of such anchors with variable sources in a given IRS. Basically, such a plan consists of an attempt to identify the lexical pragmatics [PRG] associated with the handle of the anchor floater in the current EPS and IRS configuration of the agent (IRS if the floater is an anaphor, EPS otherwise). What kind of plan can be used to resolve a variable anchor source depends on the type of the source; it is the type of the source which determines whether it must be anchored to a fact, a belief or a previously used floater. Note that each of the plans may return several possible resolutions - resulting in sets of possible anchor source resolutions. There are three procedures for resolving a variable anchor source:

- (1) *Anaphoric resolution*: the floater in question must be anchored to a previous occurrence of the floater with the same identification conditions [PRG] - if not specified otherwise, anaphoric floaters inherit their identification conditions from their anchor source. Anaphors are distinguished by their lexical form and it is their lexical entry which triggers the anaphoric resolution procedure. This procedure is called *resolve-anchor-anaphora*. It is pictured in figure 7.18, section 7.4.
- (2) *Resolution to beliefs*: the [PRG]-conditions associated with the floater in question must be identified in the EPS. This can be done by trying to find a unification of [PRG] with a substructure of the EPS, where the bindings of the unification constitute the floater's source. If no such unification is possible, the [PRG] of the floater is to be added to the beliefs of the agent. The corresponding procedure is called *resolve-anchor-belief*. It is pictured in figure 7.20, section 7.4.
- (3) *Resolution to facts*: the [PRG]-conditions associated with the floater in question must be identified in the factual part of the EPS. This can be done by trying to unify [PRG] with the set of facts, where the bindings of the unification constitute the floater's source. If no such unification is possible, a procedure for error correction must be triggered. The corresponding plan is called *resolve-anchor-fact*. It is pictured in figure 7.21, section 7.4.

Resolution of definite anchor sources An anchor source of the type $\langle x, \overrightarrow{\text{source}} \rangle$ requires the source to be unique, i.e. the anchor source must be uniquely identified by the pragmatic profile of the corresponding name of the anchor floater or an accompanying action.

- $\langle x, \overrightarrow{\text{source}} \rangle \vDash_{\uparrow M, t} \text{handle}(x)$ iff there is exactly one *source* with which $\text{PRG}_{\text{handle}(x)}$ can be identified in $\uparrow M$ at t . If this is not possible, a plan must be activated to uniquely resolve the *source*: *resolve-ambiguous-anchor(K)*. The plan is given in figure 7.22, section 7.4.

Pragmatic Profiling - Identification of time-individuals This paragraph spells out the conditions for the identification of time-individuals with respect to a model $\uparrow M$ (definition 43). It should be noted that I discuss the integration of the IRS interpretation process into the BDI-based planning approach pursued in this thesis in the next subsection 6.3.2. In the following, we have to take into account that the identification of a time-individual ev depends on

- the resolution of the thing-anchor sources which occur in the anchor source of ev (see section 6.3.2)
- the temporal location of ev with respect to the now-constant of the IRS in which ev occurs and (see section 6.3.2 below)
- the interpretation mode of ev (see sections 5.3.4 and 6.3.2)

Given a time-individual $handle(ev)$ and the anchor for ev occurring in an IRS K , the mode of interpretation for ev and the temporal location for ev with respect to the now-constant of the IRS in which ev occurs, definition 47 states the conditions for the identification of ev in $\uparrow M, t$ (definition 43), where t is determined by the time-index $\langle K, t \rangle$ of the IRS K in which ev occurs (section 5.4.1). It should be noted that not only plain interpretation is defined recursively but that reactive interpretation is supposed to work in a recursive way too. That is, once an intention has been added by reactive interpretation it is this intention which further interpretation attempts act on. The following definition 47 states one of the possible formulations of the identification conditions for time-individuals¹⁰. The identification conditions for IRS time-individuals have the following structure:

- An anchor of a time-individual $ev \models_{\uparrow M, t}$ The handle of ev
 - plain: the set-theoretic conditions that must be satisfied for a plain interpretation of ev .
 - reactive: the actions that are triggered by a reactive interpretation of ev .

The identification conditions for IRS time-individuals make use of the interpretation function I for EPS-times (definitions 33, 34, 31, 35) as formally defined with IRS-models 43. The interpretation algorithm for IRSs makes use of the following definition 47 as a function

`identify(ev, interpretation-mode, temporal-location, time-index(K), set-of-thing-anchors).`

Definition 47 *Identification of time-individuals*

- If $n \subseteq ev$ then (present)
 - $\langle s, R(x_1, \dots, x_n) \rangle \models_{\uparrow M, t} handle(s)$
 - * plain: iff $\exists G = \{\langle x_1, a_1 \rangle, \dots, \langle x_n, a_n \rangle\}$ sth. $PRG_{handle}(a_1, \dots, a_n) \in \mathbf{P}(a_1, \dots, a_n)(t)$;
 - * reactive: $\mathbf{b-add}(x, t, PRG_{handle}(a_1, \dots, a_n))$
 - $\langle s, xK_h K \rangle \models_{\uparrow M, t} name(s)$
 - * plain: iff $[K]_{\uparrow M, t} \in \mathbf{Abilities}(x)(t)$
 - * reactive: $\mathbf{add} - \mathbf{ability}(x, [K]_{\uparrow M, t})$

¹⁰More fine-grained identification conditions for time-individuals could also consider other patterns of reaction depending e.g. on the trust relationship with the agent whose utterance triggers the interpretation attempt.

- $\langle s, xBK \rangle \models_{\uparrow M, t} \text{handle}(s)$
 - * *plain*: iff $[K]_{\uparrow M, t} \in \mathbf{B}(x)(t)$
 - * *reactive*: **b-add** (x, t, K)
- $\langle s, xK_tK \rangle \models_{\uparrow M, t} \text{handle}(s)$
 - * *plain*: iff $[K]_{\uparrow M, t} \in \mathbf{F}(t)$
 - * *reactive*: prevented by the force-plain constraint associated with 'x know that K'.
- $\langle e, xCAUSEK \rangle \models_{M, S, t} \text{handle}(e)$
 - * *plain*: iff $\exists[S; t, t_1] \in \mathbf{S}(x)(t)$ sth. $PRG_{\text{handle}}(e) \in S$ and $\models_{\uparrow M, t_1} K$;
 - * *reactive*: **b-add** $(x, t, PRG_{\text{handle}}(e))$
- $\langle e, xDOK \rangle \models_{\uparrow M, S, P, t} \text{handle}(e)$
 - * *plain*: iff $\exists[S; t, n] \in \mathbf{S}(x)(t)$ and $\exists[P; n, t_1] \in \mathbf{T}(x)(n)$ sth. $(S \cup P) \in PRG_{\text{handle}}(e)$ and $\models_{\uparrow M, t_1} K$ and $[K]_{\uparrow M, t} \in \mathbf{Attitude}(Do, x, t)$;
 - * *reactive*: **g-add** $(x, PRG_{\text{handle}}(e))$
- $\langle e, xINTK \rangle \models_{\uparrow M, S, P, t} \text{handle}(e)$
 - * *plain*: iff $\exists[S; t, n] \in \mathbf{S}(x)(t)$ and $\exists[P; n, t_1] \in \mathbf{T}(x)(n)$ sth. $(S \cup P) \in PRG_{\text{handle}}(e)$ and $[K]_{\uparrow M, t} \in \mathbf{Attitude}(Int, x, t)$;
 - * *reactive*: **i-add** $(x, PRG_{\text{handle}}(e))$
- If $n < ev$ then (future)
 - $\langle s, R(x_1, \dots, x_n) \rangle \models_{\uparrow M, t} \text{handle}(s)$
 - * *plain*: iff $\exists G = \{\langle x_1, a_1 \rangle, \dots, \langle x_n, a_n \rangle\}$ sth. $PRG_{\text{handle}}(a_1, \dots, a_n) \in \mathbf{P}(a_1, \dots, a_n)(t)$;
 - * *reactive*: **b-add** $(x)(t)(PRG_{\text{handle}}(a_1, \dots, a_n))$
 - $\langle s, xK_hK \rangle \models_{M, t} \text{handle}(s)$
 - * *plain*: iff $[K]_{\uparrow M, t} \in \mathbf{Abilities}(x)(t)$
 - * *reactive*: **add – ability** $(x, [K]_{\uparrow M, t})$
 - $\langle s, xBK \rangle \models_{\uparrow M, t} \text{handle}(s)$
 - * *plain*: iff $[K]_{\uparrow M, t} \in \mathbf{B}(x)(t)$
 - * *reactive*: **b-add** $(x, t, [K]_{\uparrow M, t})$
 - $\langle s, xK_tK \rangle \models_{\uparrow M, t} \text{handle}(s)$
 - * *plain*: **g-add** $(\text{error-correction}(\text{handle}(s)))$
 - * *reactive*: prevented by the force-plain constraint associated with 'x know that K'.
 - $\langle e, xCAUSEK \rangle \models_{\uparrow M, S, t} \text{handle}(e)$
 - * *plain*: iff $\exists[S; t, t_1] \in \mathbf{S}(x)(t)$ sth. $S \in PRG_{\text{handle}}(e)$ and $\models_{\uparrow M, t_1} K$;
 - * *reactive*: **b-add** $(x, t, PRG_{\text{handle}}(e))$
 - $\langle e, xDOK \rangle \models_{M, S, P, t} \text{handle}(e)$
 - * *plain*: iff $\exists[P; t, t_1] \in \mathbf{T}(x)(t)$ sth. $PRG_{\text{handle}}(e) \in P$ and $\models_{\uparrow M, t_1} K$ and $[K]_{\uparrow M, t} \in \mathbf{Attitude}(Do, x, t)$;

- * reactive: $g\text{-add}(x, PRG_{handle}(e))$
- $\langle e, xINTK \rangle \models_{\uparrow M, t} handle(e)$
 - * plain: iff $\exists [P; t, t_1] \in \mathbf{T}(x)(t)$ sth. $PRG_{handle}(e) \in P$ and $\models_{\uparrow M, t_1} K$ and $[K]_{\uparrow M, t} \in \mathbf{Attitude}(Int, x, t)$;
 - * reactive: $i\text{-add}(x, PRG_{handle}(e))$
- If $ev \prec n$ then (past)
 - $\langle s, R(x_1, \dots, x_n) \rangle \models_{\uparrow M, t} handle(s)$
 - * plain: iff $\exists G = \{\langle x_1, a_1 \rangle, \dots, \langle x_n, a_n \rangle\}$ sth. $PRG_{handle}(a_1, \dots, a_n) \in \mathbf{P}(a_1, \dots, a_n)(t)$;
 - * reactive: $b\text{-add}(x)(t)(PRG_{handle}(a_1, \dots, a_n))$
 - $\langle s, xK_h K \rangle \models_{\uparrow M, t} handle(s)$
 - * plain: iff $[K]_{\uparrow M, t} \in \mathbf{Abilities}(x)(t)$
 - * reactive: $add - ability(x, [K]_{\uparrow M, t})$
 - $\langle s, xBK \rangle \models_{\uparrow M, t} handle(s)$
 - * plain: iff $[K]_{\uparrow M, t} \in \mathbf{B}(x)(t)$
 - * reactive: $b\text{-add}(x, t, [K]_{\uparrow M, t})$
 - $\langle s, xK_t K \rangle \models_{\uparrow M, t} handle(s)$
 - * plain: iff $[K]_{\uparrow M, t} \in \mathbf{F}(t)$
 - * reactive: prevented by the force-plain constraint associated with 'x know that K'.
 - $\langle e, xCAUSEK \rangle \models_{\uparrow M, S, t} handle(e)$
 - * plain: iff $\exists [S; t, t_1] \in \mathbf{S}(x)(t)$ sth. $S \in PRG_{handle}(e)$ and $\models_{\uparrow M, t_1} K$;
 - * reactive $b\text{-add}(x, t, PRG_{handle}(e))$
 - $\langle e, xDOK \rangle \models_{\uparrow M, S, t} handle(e)$
 - * plain: iff $\exists [S; t, t_1] \in \mathbf{S}(x)(t)$ sth. $S \in PRG_{handle}(e)$ and $\models_{\uparrow M, t_1} K$ and $[K]_{\uparrow M, t} \in \mathbf{Attitude}(Do, x, t)$;
 - * reactive: $b\text{-add}(x, t, PRG_{handle}(e))$
 - $\langle e, xINTK \rangle \models_{\uparrow M, S, t} handle(e)$
 - * plain: iff $\exists [S; t, t_1] \in \mathbf{S}(x)(t)$ sth. $PRG_{handle}(e) \in S$ and $\models_{\uparrow M, t_1} K$ and $[K]_{\uparrow M, t} \in \mathbf{Attitude}(Int, x, t)$;
 - * reactive: $b\text{-add}(x, t, PRG_{handle}(e))$

Identification of complex conditions I restrict the identification of disjunction, quantification and negation to plain interpretation. However, for the negation of time-individuals it should be intuitively clear that a reactive interpretation requires the adoption of a 'maintenance' intention that prevents a certain event or state from occurring (E.g. "Don't touch the cube"). The reactive interpretation of quantification over time-individuals (E.g. "All your attempts to build a corner bolting will fail") would require to integrate the quantifiers of CTL* [Emerson, 1990] for quantification over EPS scenarios into the given basic language of IRSs.

Definition 48 *Identification of complex conditions*

- $G \vDash_{\uparrow M, t} K_1 \vee K_2$ iff there is some H such that $\langle\langle G, H \rangle\rangle \vDash_{\uparrow M, t} K_1$ or there is some H such that $\langle\langle G, H \rangle\rangle \vDash_{\uparrow M, t} K_2$
- $G \vDash_{\uparrow M, t} K_1 \Rightarrow K_2$ iff for all possible resolutions of sources $? \text{ resp. } !$ such that $\langle\langle G, H \rangle\rangle \vDash_{\uparrow M, t} K_1$ there is some I such that $\langle\langle H, I \rangle\rangle \vDash_{\uparrow M, t} K_2$
- $G \vDash_{\uparrow M, t} \neg K$ iff there exists no H such that $\langle\langle G, H \rangle\rangle \vDash_{\uparrow M, t} K$

Integrating IRS interpretation into the framework of BDI-based planning

This section spells out how the concept of IRS interpretation introduced above integrates into the overall framework of BDI-based planning conducted by the agent layer. I start with an overview of the parts and functions of the BDI-interpreter discussed so far.

The BDI-interpreter revisited Recall the main loop of the BDI-interpreter (definition 13, page 64) which I redisplay below.

Note 17 *Recall: the BDI-interpreter main loop*

```

1 initialize-state;
do
2 options:=option-generator(trigger-queue,Beliefs,Goals,Intentions);
3 b-add(options, EPS(now));
4 update(IRS);
5 selected-options:=deliberate(options,Beliefs,Goals,Intentions);
6 update-intentions(selected-options,Intentions);
7 execute(Intentions);
8 get-new-SMS;
9 if f-update(SMS) then f-add(SMS,EPS(now));
10 update(IRS);
11 get-new-triggers(EPS,IRS);
12 drop-successful-attitudes(Beliefs,Goals,Intentions);
13 drop-impossible-attitudes(Beliefs,Goals,Intentions);
until quit.
```

The various places in which we discussed the functions of the BDI-interpreter main loop in the course of this thesis are compiled in note 18 below.

Note 18 *Discussion of the BDI-interpreter main loop*

- (1) Line 1 of the main loop will be discussed in section 7.1.4 in the next part of this thesis.
- (2) Line 2 was discussed in section 4.2.3.

- (3) Line 3 was discussed in section 6.2.5.
- (4) Line 4 will be discussed in detail in this section.
- (5) Lines 5, 6 and 7 were discussed in section 4.2.3.
- (6) Lines 8 and 9 were discussed in sections 6.2.5 and 4.2.3.
- (7) Line 10 was discussed in section 5.4, and is discussed in more detail below.
- (8) Line 11 was discussed in 4.2.3.
- (9) Lines 12 and 13 were discussed in section 4.2.3

Information limited Interpretation: $\downarrow M$ The reader should have noticed that the BDI-interpreter main loop does not handle the interpretation of IRSs. IRS interpretation is executed via the internal EPS action $\text{int-a:?}K$. In the following, the functioning of $\text{int-a:?}K$ is spelled out. In definition 45, I stated the notion of a successful interpretation of an IRS K with respect to the inflated model $\uparrow M$. But for the intended application of this thesis, it is reasonable to constrain the agent's models for IRS interpretation to closed-world models $\downarrow M$ (as discussed in section 6.3.2). The model $\downarrow M$ of an agent x at time t is given in definition 49.

Definition 49 A model $\downarrow M$ at a time t of an agent x is a 10-tuple

$$\bullet \downarrow M(x)(t) = \langle P, S, T, P, F, B, PRG, \downarrow \mathbf{Things}, \mathbf{Attitudes}, \mathbf{Abilities} \rangle$$

The sets **Attitudes**, **Abilities** and functions P, S, P, F, B, T, PRG of $\downarrow M$ are these of an agent's EPS at t as defined in section 6.2.4 and the agent's knowledge base as defined in chapter 7 and **Things** is the finite set of EPS things **Things**. In the following, I omit the arguments x and t of $\downarrow M$ and assume that $\downarrow M$ is the current model of the interpreting agent.

Triggering an interpretation attempt An interpretation attempt of an IRS K with respect to $\downarrow M$ is triggered at the EPS level by the internal action $\text{int-a:?}K$, which occurs in certain plans of the agent (e.g. the plan for IRS resolution) or as part of the overall processing of utterances. The internal action $\text{int-a:?}K$ triggers the procedure **preprocess-anchors**(K) (definition 50) which pushes K on a list of pending IRSs if it contains variable thing-anchor sources. If all thing-anchors of K are already resolved, K and its set of possible thing-anchors are passed over to the main interpretation procedure **interpret**($K, \text{set-of-thing-anchors}$).

Definition 50 states the procedure for preprocessing an IRS K .

Definition 50 *preprocess-anchors*(K);

if contains-variable-thing-anchors(K) *then*
 add-pending-list(K);
elseif interpret($K, \text{set-of-thing-anchors}$)

The list of pending IRSs is processed via the `update(IRS)`-function (which was discussed in detail in section 5.4). In definition 51, I redisplay the lines of `update(IRS)` concerned with the processing of pending IRSs. Each execution of `update(IRS)` by the main control loop checks the list of pending IRSs `pending-list` whether there are IRSs on this list for which all the variable thing-anchor sources have been resolved by the plans invoked by the occurrence of a variable anchor source (see section 6.3.2). If this is the case, the respective IRS is removed from `pending-list` and the IRS and its set of possible thing-anchors `set-of-thing-anchors` is passed over to the main interpretation procedure `interpret(K, set-of-thing-anchors)` (definition 52).

Definition 51 *The part of `update(IRS)`; concerned with IRS interpretation*

...

```
for K in pending-list do
  if resolved(K, set-of-thing-anchors) then
    remove-pending-list(K);
    interpret(K, set-of-thing-anchors);
end do;
```

...

The main procedure for IRS interpretation The main procedure for IRS interpretation `interpret(K, set-of-thing-anchors)` acts on an IRS K and a set `set-of-thing-anchors` of possible resolutions of the thing-anchors of K . Note that at this level, all anchors for thing-individuals are resolved to (at least) one entity of the EPS.

The first step of `interpret(K, set-of-thing-anchors)` preprocesses the time-individuals $ev_1 \dots ev_n \in K$ in that it determines

- The temporal location of the time-individuals $ev_1 \dots ev_n \in K$ with respect to the now of K . This function is not spelled out in detail. However, it should be easy to implement this function by making use of the usual logical properties of the constants in $R_{pred} = \{<_{beg}, <, =, \subset, \subseteq\}$.
- The interpretation mode of each time-individual $ev_1 \dots ev_n \in K$. The default interpretation mode of a time-individual is plain interpretation, i.e. if no interpretation mode is defined for a certain time-individual, plain mode is assigned to the time-individual. Otherwise, the mode assigned to the time-individual is used for interpretation. See section 5.3.4 for a more detailed discussion of interpretation modes.

Second, given K , its set of possible thing-anchors, the information about the interpretation mode, the temporal location of the time-individuals $ev_1 \dots ev_n \in K$ and the time-index of K , the identification conditions for time-individuals as stated in definition 47 are applied to the time-individuals in K . For nested occurrences of time-individuals, this may result in a recursive call of the identification procedure. The sets of thing-anchors which render possible an identification of a time-individual ev are attached to ev and returned to the BDI-interpreter as a partial interpretation of K with respect to ev . If no identification of a time-individual ev was retrieved, several options exist to proceed with the identification of ev :

- If a time-individual was identified in reactive interpretation mode, this time-individuals must be reconsidered in the next BDI-interpreter cycles, as the results of reactive interpretation do not instantly appear in $\downarrow M$ (but only within the next cycles of the BDI-interpreter). E.g. if the reactive interpretation of a time-individual adds an intention to the agent's BDI-interpreter (which in turn would render the identification of the time-individual referring to the intention possible), this can only be considered in the next cycle of evaluation. In this case, K is pushed to the list of pending IRSs and reconsidered in plain mode in the next cycle of IRS interpretation.
- If a time-individual was identified in plain interpretation mode, it may be possible that a reactive interpretation of this time-individuals renders possible an identification in the next interpretation cycles. Thus K is returned to the list of IRSs to be reconsidered in reactive mode in the next cycle of the BDI-interpreter main loop.
- If a time-individual is tagged with the flag *force-plain* and the time-individual has no identification, this can not be changed by reconsidering the IRS K containing the respective time-individual in reactive mode. However, the future development of reality may bring about an identification at a later time independent from the agent. Thus K is pushed back to the pending-list of IRSs for later reconsideration with forced plain interpretation mode.

Third, given the IRS K , its time-index and the sets of thing-anchors which identify the time-individuals of K and its sub-IRSs, the complex conditions of K have to be processed along the definition for the identification of complex conditions as specified in definition 48.

Finally, the results of interpretation are processed. If K has a successful interpretation (see definition 45), this result is returned as the internal action *has-interpretation(K)* to the BDI-interpreter. If K has no interpretation, the internal action *no-interpretation(K)* is returned to the BDI-interpreter.

Definition 52 states an algorithmic formulation of the above considerations.

Definition 52 *interpret(K, set-of-thing-anchors);*

```
// Process time-individuals of K
for ev in K do
  determine-temporal-location(ev, now(K));
  if undefined(ev, interpretation-mode) then
    identify(ev, plain, temporal-location, time-index(K), set-of-thing-anchors);
  elseif
    identify(ev, interpretation-mode, temporal-location, time-index(K), set-of-thing-anchors);
  if identified(ev) then
    set(has-identification(ev, identifying-thing-anchors));
  //Process no-identification of time-individuals
elseif no-identification(ev, interpretation-mode) then
  set(no-identification(K));
if force-plain(ev) then
  set(no-identification(K));
elseif no-identification((ev, reactive) then
  set(mode(ev, force-plain));
```

```

    elseif no-identification((ev,plain) then
        set(mode(ev,reactive));
    //Return partial interpretation results for K
    return(partial-interpretation(K(ev)))
    end do;
//Add IRSs to be reconsidered to the pending list
add-pending-list(K);
//Identify the complex conditions of K with respect to the determined sets
of thing-anchors that identify the time-individuals of K
identify(K,complex-conditions,identifying-thing-anchors)
//Return the result of interpretation to the BDI-interpreter
if successful-interpretation(K) then return has-interpretation(K);
elseif return no-interpretation(K);

```

Accessibility of IRS-referents

A core concept of the representational dynamics of DRT is that of the accessibility of referents that governs the resolution of anaphoric reference. We can mirror this concept in the given framework without much effort following the standard definition of accessibility in terms of subordination. The modification concerns the accessibility of occurrences of floaters nested within time-individual anchor sources. This can be easily treated via an addition of these anchor sources to the set of conditions for which accessibility is defined. i.e. anchor sources follow the same principles of accessibility as 'normal' conditions do. The syntactic concept of accessibility is employed by the plan for the resolution of anaphoric thing-anchor sources (see section 6.3.2). In order to access IRS referents from temporally distributed IRSs, we can consider the set of IRSs constructed along the processing of a discursive interaction as being contained in a 'top-level' IRS. I do not display this top-level IRS in the following examples, as it exceeds the limitations of A4-paper size.

Definition 53 Subordination

K_1 subordinates K_2 , iff

- K_1 immediately subordinates K_2 ; or
- There is an IRS K sth. K_1 subordinates K and K subordinates K_2
- That is, the subordination relation is the transitive closure of the 'immediately subordinates' relation.

Definition 54 Immediate Subordination

K_1 is an immediate sub-IRS of K , $K_1 <_{sub} K$, if any of the following conditions holds:

- $\neg K_1 \in Con_K$
- There is an IRS K sth. $K_2 \vee K \in Con_{K_1}$ or $K \vee K_2 \in Con_{K_1}$

- There is an IRS K_2 sth. $K_1 \Rightarrow K_2 \in \text{Con}_K$ or $K_2 \Rightarrow K_1 \in \text{Con}_K$

Definition 55 *Accessibility of IRSs*

Given IRSs K and K_1 , K is accessible from K_1 ; $K \text{acc} K_1$ iff

- $K_1 \leq_{\text{sub}} K$; or
- there exist IRSs K_2 and K_3 , sth. $K_2 \Rightarrow K_3$ and $K \text{acc} K_2$ and $K_3 \text{acc} K_1$

Definition 56 *Accessibility of IRS floaters*

Given IRSs K and K_1 , and referents x and y , x is accessible from y ; $x \text{acc}_y$ iff $x \in U_K, y \in U_{K_1}$ and $K \text{acc} K_1$.

Part III

Practical Application

Chapter 7

Application to examples

This chapter applies the developed mechanisms of GDRT to the analysis of examples for the processing of discursive interaction between robots and other agents.

7.1 Preliminaries

7.1.1 The limitations of the 'paper-and-pencil' approach

When going through the application of the proposed formalism of GDRT to examples in the next sections, the reader will notice that even the analysis of simple examples is complex and that much space is needed to present all relevant details. In fact, the actual 'power' of the formalism of GDRT as defined in the previous chapter 6 can hardly be displayed in a 'paper-and-pencil' approach but only in an actual implementation as proposed with the example setup in section 1.2.5. The way in which I present the analysis of the examples considered in the present chapter reflects this: for reasons of space I can not display the complete cognitive configuration of an agent participating in a discursive interaction at each an any time step. Instead, I picture the development of an agent's cognitive state from the perspective of successfully executed interaction. That is, I state 'a discourse history' in the form of a linear sequence of IRSs, EPSs and SMSs.

7.1.2 Example discourses

Given the limitations mentioned above, I intend to illustrate the application of GDRT in a proof-of-concept manner via an analysis of typical examples for interactions between a robot and human.

- The first example (section 7.2) illustrates 'single mode' and demonstrates the robot's execution of the plan for building a cube bolting without the participation of other agents. It introduces the reader to the processing of discursive interaction in the proposed framework of Grounded Discourse Representation Theory.
- The second example (section 7.5) illustrates 'assistance mode' in that it adds the interaction with other agents to the picture, where a human assists the robot in constructing a corner cube bolting.

This example intends to familiarize the reader with the treatment of multi-party interaction in GDRT.

- The third example (section 7.6) illustrates 'collaboration mode' and gives an example of how the robot distributes work to be done between herself and the collaborator. This example introduces more complex representations and is a first step toward the processing of free interaction.
- The fourth example (section 7.7) illustrates 'teaching mode' and extends assistance mode in that it requires the robot to verbalize the actions she performs and to explain what she does. This example demonstrates the use of context-dependent plans, as teaching mode is triggered by a lack of know-how of the human collaborator.
- All the examples mentioned above employ a more or less 'traditional' approach to the processing of discursive interaction in that the robot controls the interaction via the execution of plans in which all the necessary turns and actions are hard-coded. This changes with the final example for 'free interaction' (section 7.8), where the human initiates and controls the interaction and the robot must participate in the interaction via the interpretation of the humans utterances and gestures.

7.1.3 Examples of semantic-pragmatic concepts

This section gives examples for some of the semantic-pragmatic concepts that I employ in the analysis of examples in chapter 7. As this thesis is focused on a specific part of the analysis of discursive interaction between humans and robots, what follows is limited in the sense of the restrictions discussed in section 1.2.3, more specifically

- Semantic-pragmatic concepts do not specify syntactic properties of the concepts in question but only present the contribution to the semantics-pragmatics (IRS-EPS) interface of abstract concepts as I assume them to underly the processing of linguistic and non-linguistic information structures (such as IRSs or EPS structures).
- It is assumed that the construction and generation of IRSs with respect to utterances (the syntax-semantics interface) is contributed by a component external to the scope of this thesis. Consequently, the relation between utterances and IRSs is not discussed or specified in great detail.¹

It should be noted that the semantic-pragmatic concepts that are given in the following as well as the example plans serve merely illustrative purposes and thus can and should be refined in many respects. I am sure that the reader who goes through the following list of lexical entries, plans and example analysis will be able to think of improvements and refinements - this is what the instruments developed within GDRT intend to provide: formal and precise instruments that allow to investigate elusive problems located at the interface between semantics and pragmatics. The set of semantic-pragmatic concepts specified in the following together with the set of example plans specified in section 7.4 makes up Clara's *'knowledge base'* that she employs in the example interactions.

¹As GDRT is in its definition 'downwards-compatible' to DRT, the same algorithms that are employed for the construction and generation of DRSs with respect to utterances are applicable to IRSs.

Thing-individuals with simple anchor sources**Sem-Prag-Concept 3** *tillmann*

The semantic-pragmatic concept for the thing-individual 'tillmann' consists of an IRS specifying the name of the handle for the thing-individual 'tillmann' and an anchor source that specifies the identification of 'tillmann' in the EPS structure. As 'tillmann' is the only person involved in the examples to follow, the EPS-condition that 'tillmann' is a person is sufficient to identify 'tillmann' in the EPS.

SEM $\begin{array}{l} x \\ \text{tillmann}(x) \\ \langle x, d \rangle \end{array}$

PRG $\begin{array}{l} d \\ \text{person}(d) \end{array}$

Sem-Prag-Concept 4 *slat*

SEM $\begin{array}{l} x \\ \text{slat}(x) \\ \langle x, a \rangle \end{array}$

PRG $\begin{array}{l} a \\ \text{slat}(a) \end{array}$

Sem-Prag-Concept 5 *cube*

SEM $\begin{array}{l} x \\ \text{cube}(x) \\ \langle x, a \rangle \end{array}$

PRG $\begin{array}{l} a \\ \text{cube}(a) \end{array}$

Sem-Prag-Concept 6 *screw*

SEM $\begin{array}{l} x \\ \text{screw}(x) \\ \langle x, a \rangle \end{array}$

PRG $\begin{array}{l} a \\ \text{screw}(a) \end{array}$

Sem-Prag-Concept 7 *you*

This semantic-pragmatic concept specifies a thing-individual with a definite anchor source (requiring a unique resolution). It represents singular personal pronouns such as 'you'.

SEM $\begin{array}{l} x \\ \langle x, \vec{a} \rangle \end{array}$

PRG $\begin{array}{l} \text{person}(a) \end{array}$

Thing-individuals with complex anchor sources**Sem-Prag-Concept 8** *corner cube bolting*

The semantic-pragmatic concept for 'corner cube bolting' involves a complex identification in the EPS structure, i.e. the anchor source for 'ccb' consists of the EPS-conditions 'bolted' and 'pos-corner' holding for a screw, a slat and a cube.

SEM	x $ccb(x)$ $\langle x, d \rangle$		
PRG	$a, b, c - d$ $screw(a)$ $slat(b)$ $cube(c)$ $d :$ <table style="border: 1px solid black; padding: 2px; margin-left: 20px;"> <tr> <td>$bolted(a, b, c)$</td> </tr> <tr> <td>$pos - corner(c, b)$</td> </tr> </table>	$bolted(a, b, c)$	$pos - corner(c, b)$
$bolted(a, b, c)$			
$pos - corner(c, b)$			

Sem-Prag-Concept 9 *cube bolting*

SEM	x $cb(x)$ $\langle x, d \rangle$
PRG	$a, b - d$ $screw(a)$ $cube(b)$ $d : bolted(a, b)$

Thing-individuals affecting IRS interpretation**Sem-Prag-Concept 10** *clara*

The semantic-pragmatic concept for 'clara' specifies an IRS for 'clara' (with a self-referential anchor source) and comes with a command to be executed by the BDI-interpreter when an IRS containing the semantic contribution of 'clara' is processed (see sections 5.3.4 and 6.3.2).

SEM	i $clara(i)$ $\langle i, i \rangle$
-----	---

PRG $set(force\text{-}plain(ev))$ for time-individuals ev involving agent i if ev is located before or at n

Sem-Prag-Concept 11 *i*

SEM	i $\langle i, i \rangle$
-----	-------------------------------

PRG $set(force\text{-}plain(ev))$ for time-individuals ev involving agent i if ev is located before or at n

Thing-individuals with placeholders for IRS handles

Sem-Prag-Concept 12 *The following semantic-pragmatic concept specifies an internally anchored thing-individual for a person with unknown name. This entry enables the robot to construct IRSs involving persons for which she has not yet a unique name. The placeholder for the IRS-handle ? triggers a plan for the resolution of the handle of x .*

SEM	$\begin{array}{l} x \\ ?(x) \\ \langle x, d \rangle \end{array}$
-----	--

PRG	$\begin{array}{l} d \\ person(d) \end{array}$
-----	---

Thing-individuals with variable anchor sources

Sem-Prag-Concept 13 *This semantic-pragmatic concept specifies a thing-individual with a variable internal anchor, i.e. the anchor source is unknown and has to be determined. Such cases of variable anchor sources occur as part of the processing of utterances of other agents, where it is part of the semantic binding to resolve variable anchor sources.*

SEM	$\begin{array}{l} x \\ handle(x) \\ \langle x, ? \rangle \end{array}$
-----	---

PRG the [PRG] conditions associated with the handle of x .

Sem-Prag-Concept 14 *This semantic-pragmatic concept specifies a thing-individual with a definite variable internal anchor, i.e. the resolution of the variable anchor source must be unique.*

SEM	$\begin{array}{l} x \\ handle(x) \\ \langle x, \vec{?} \rangle \end{array}$
-----	---

PRG the [PRG] conditions associated with the handle of x .

Sem-Prag-Concept 15 *This semantic-pragmatic concept specifies a thing-individual with an anaphoric variable anchor. This is my simplifying analysis for the 'that' in "Do you know how do that?".*

SEM	$\begin{array}{l} x \\ \langle x, ?a \rangle \end{array}$
-----	---

PRG $g\text{-add}(\text{resolve-anchor-anaphora-that}(?a))$

Stative time-individuals

Sem-Prag-Concept 16 *x is between y and z*

SEM x, y, z, s
 $\langle s, \text{between}(x, y, z) \rangle$
 $\text{between}(s)$
 $\langle x, a \rangle$
 $\langle y, b \rangle$
 $\langle z, c \rangle$

PRG a, b, c
 $\text{between}(a, b, c)$

Sem-Prag-Concept 17 x is at the corner of y

SEM x, y, s
 $\langle s, \text{pos} - \text{corner}(x, y) \rangle$
 $\text{at} - \text{corner}(s)$
 $\langle x, a \rangle$
 $\langle y, b \rangle$

PRG a, b
 $\text{pos} - \text{corner}(a, b)$

Sem-Prag-Concept 18 x is cornerbolted to z with y

SEM x, y, z, s
 $\langle s, \text{cornerbolted}(x, z, y) \rangle$
 $\text{cornerbolted}(s)$
 $\langle x, a \rangle$
 $\langle y, b \rangle$
 $\langle z, c \rangle$

PRG a, b, c
 $\text{bolted}(a, b, c)$
 $\text{pos} - \text{corner}(b, c)$

Sem-Prag-Concept 19 x knows how to do K

SEM x, s
 $\text{know} - \text{how} - \text{to} - \text{do}(s)$
 $\langle s, xK_hK \rangle$

Sem-Prag-Concept 20 x knows that K

SEM x, s
 $\text{know} - \text{that}(s)$
 $\langle s, xK_tK \rangle$

PRG $\text{set}(\text{force} - \text{plain}(s))$

Eventive time-individuals**Sem-Prag-Concept 21** x build K

SEM	x, e $build(e)$ $\langle e, xOPK \rangle$
-----	---

PRG $build(K_1)$, where K_1 is an EPS expressing the feedback of the plan for building build, K is an IRS representing K_1 and OP one of the temporal segmentation operators CAUSE,DO,INT as determined by the construction algorithm for time-individuals.

Sem-Prag-Concept 22 x take K

SEM	x, e $take(e)$ $\langle e, xOPK \rangle$
-----	--

PRG $grasp(K_1)$, where K_1 is an EPS expressing the feedback of the plan for grasping grasp, K is an IRS representing K_1 and OP one of the temporal segmentation operators CAUSE,DO,INT as determined by the construction algorithm for time-individuals.

7.1.4 The initialize-state procedure

In order to present a complete picture, I need to say something on the function `initialize-state` that precedes the execution of the main BDI control loop. Once the robot recovers from standby-mode (e.g. by switching on power), the BDI main control loop (definition 13) is initialized as stated in definition 57. First, the initial data of the object recognition is translated to an EPS, capturing the initial state of affairs. It should be noted that in the current setup, the initial state of affairs is determined by the output of the object recognition engine. That is, the initial context is limited to the 'toy world' on the table in front of the robot and approaching people². Second, it is checked whether IRS-individuals can be constructed from the given EPS. The third step checks whether clara has built-in default desires. Fourth, presets for interaction modes are set (this is needed to prefer either collaboration or assistance mode). Finally, it is checked whether new individuals resulting from the acquaintance of desires in step three require an IRS update with respect to time- and thing-individuals.

Definition 57 `initialize-state()`;

`f-add(SMS, EPS); // SMS-EPS translation`

`update(IRS); // Update of IRS`

`query(default-desires); // Check for built-in 'default' desires`

`reset-interaction-modes; // Reset all interaction mode variables`

`preset-assistance; // Flag for presetting assistance`

`update(IRS); // Update the IRS with the current state of the BDI-interpreter`

²In more complex setups, the initialization procedure would also have to determine the right, i.e. relevant context with respect to the available intentions and desires

Type:	build(K)		
Invocation:	f-add(K=	a, b $screw(a)$ $cube(b)$)
Precondition:	$\langle a, ! \rangle \langle b, ! \rangle$		
Feedback:	c $c : bolted(a, b)$		
Body:	g-add(build($a, b - c$ $c : bolted(a, b)$ $screw(a)$ $cube(b)$))

Figure 7.1: Reactive invocation of building a cube bolting. If the fact is added that there is a screw and a cube available, the desire to bolt the screw with the cube is added.

7.2 Building a cube bolting

The basic type of interaction between Clara and her environment for which I give an example in the following is what I called 'single mode' earlier. Here, a single agent (viz. Clara) performs actions that change her environment, in this case, the construction of a cube bolting. Single mode discourse does not involve other agents besides Clara. So nobody can help Clara in the case of missing knowledge or incomplete abilities. Consequently, single mode presupposes Clara's knowledge and abilities to be sufficient with respect to the realization of her desire to build a cube bolting. Example 13 gives the setup for single mode.

Example 13 *Example setup for single mode*

- *A cube and a screw on the table in front of the robot.*
- *The robot is in standby mode.*
- *The desire to build a cube bolting (abbreviated 'cb') is invoked by the factual availability of a cube and a screw as pictured by the plan given in figure 7.1. Note that this plan has a reactive invocation where the agent responds to a new configuration of her environment.*

Given the situation described in example 13, the robot has at first to wake up from standby mode by executing the `initialize-state-procedure` (as stated in definition 57) of the BDI-interpreter. This can be achieved e.g. by switching on power. The first step of initialization is an update of the EPS with the prevailing data from the SMS (via the command `f-add(SMS, EPS)`) which gives rise to an EPS/SMS configuration as pictured in figure 7.2.

The next initialization step checks whether any individuals can be constructed from the present EPS. In the given situation, no such individuals can be constructed. For the example of single mode, no presets have to be set. Execution of the BDI-interpreter main loop then activates the desire to build a cube bolting. Finally, the acquired desire is pushed to the event-queue as a new goal. While not necessary for the present purpose, an IRS of the current internal configuration can be stated as in figure 7.3.

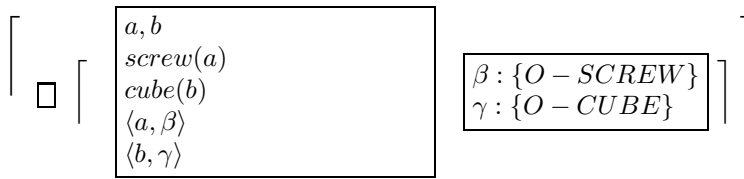


Figure 7.2: Result of $f\text{-add}(\text{SMS}, \text{EPS})$, the first step of $\text{initialize-state}()$; . The EPS is pictured at the left, the SMS at the right.

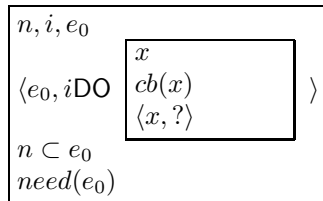


Figure 7.3: Representation of the result configuration of executing $\text{query}(\text{desire})$, $\text{g-add}(\text{desire})$; “I need a cube bolting”. This is all of the agent’s IRS at this point.

When the goal of ‘having’ a cube bolting has been added, the BDI-interpreter iterates through the agent’s plan library to find the plan for building a cube bolting. Figure 7.4 displays a possible formulation of the plan for building a cube bolting. This plan is then adopted as new future option of action in the EPS. Given that the plan for building a cube bolting is the only available option, the plan is pushed to the intention stack which in turn leads to an updated EPS where the plan has been recorded as a new intention. In turn, this leads to an initial EPS configuration as pictured in figure 7.5. In addition, the figure shows an IRS representing the current EPS configuration of the agent. Once the plan for building a cube bolting has been successfully executed, figure 7.6 pictures the final configuration of the agent.

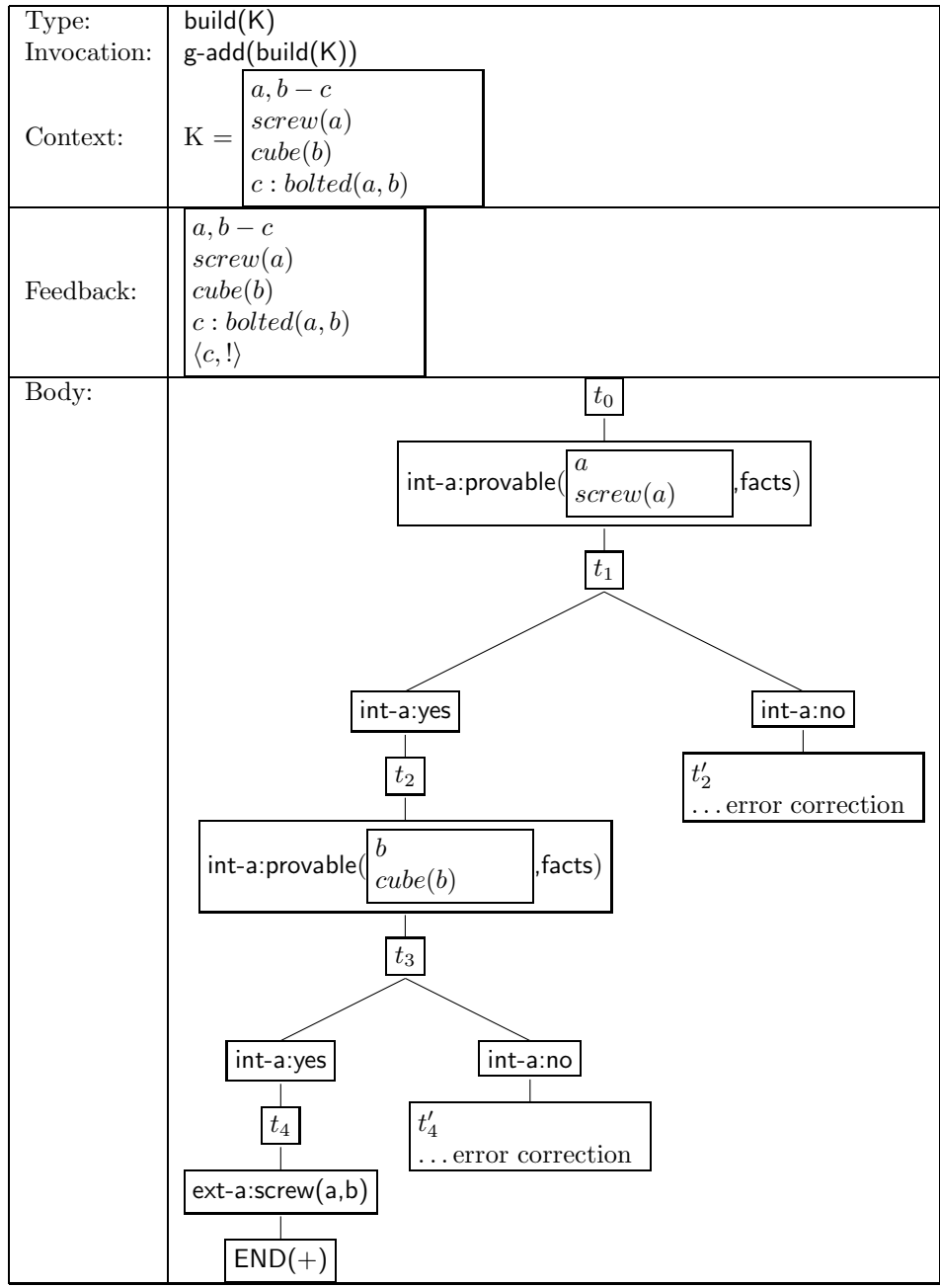


Figure 7.4: Example plan for building a cube bolting with active invocation. The plan checks for the factual availability of the screw and the cube before putting them together via the external action `screw`.

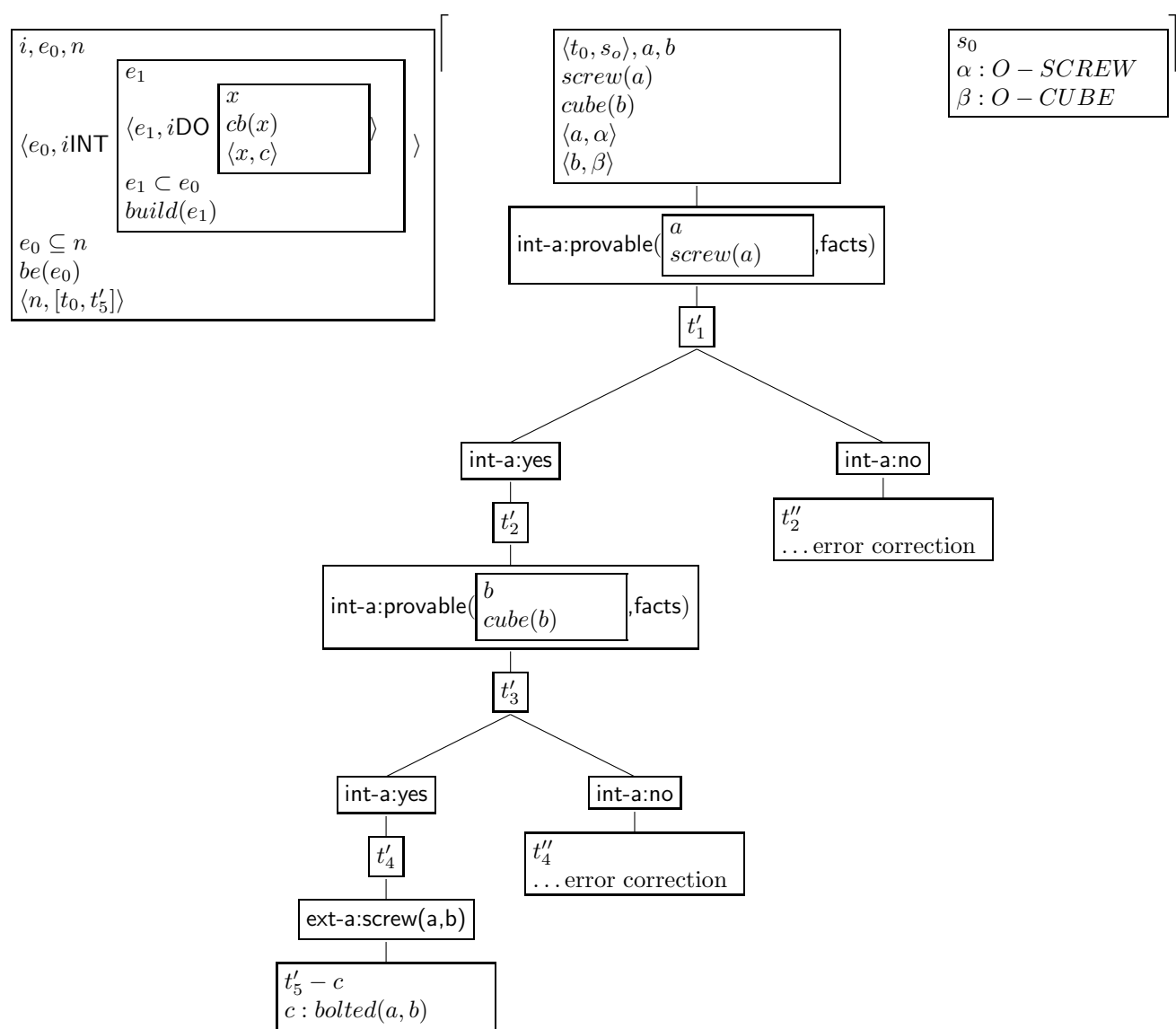


Figure 7.5: Initial state of the example for single mode. The IRS describes the current configuration of the intention stack: “I am building a cb.”. Note the extended now.

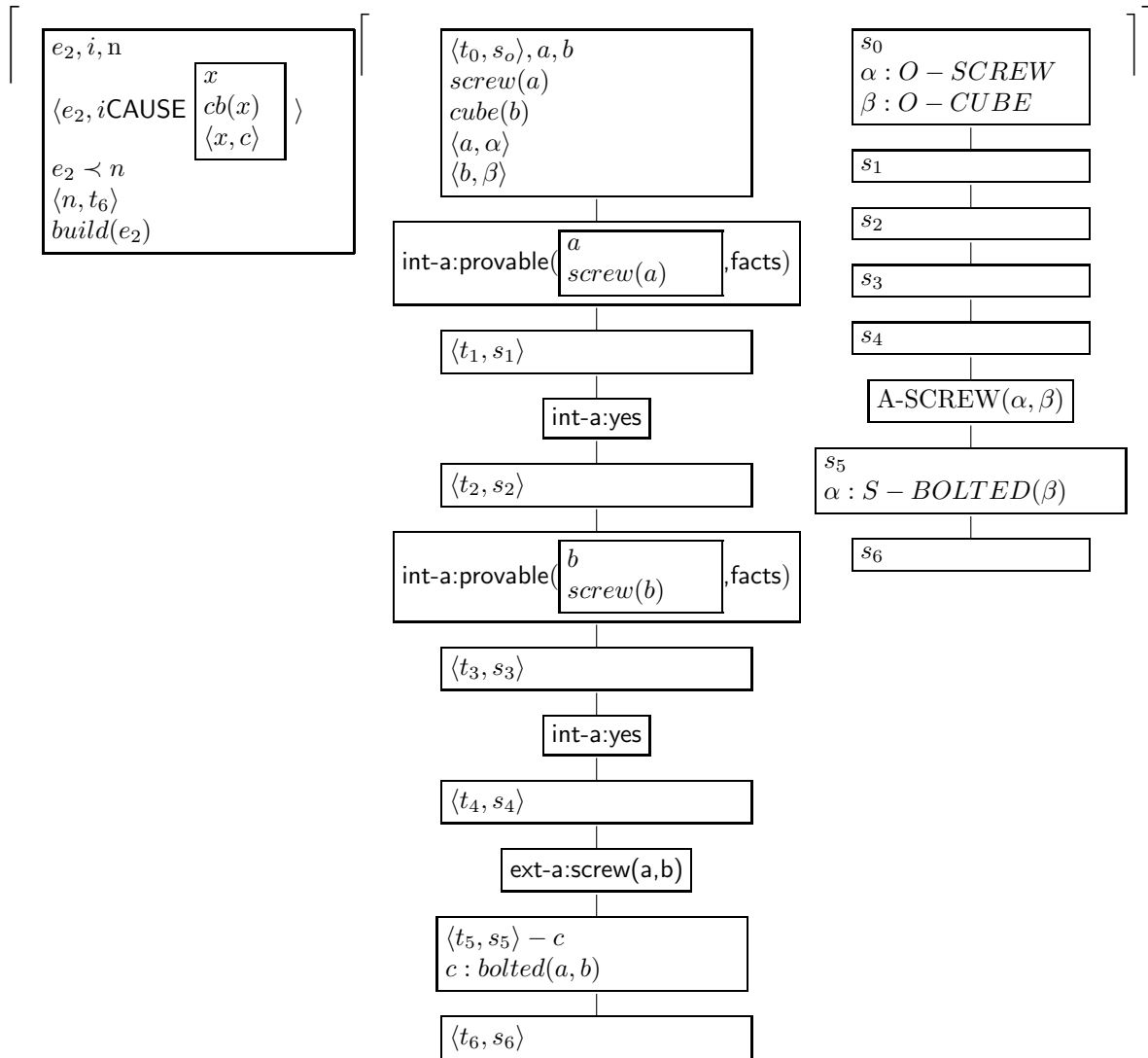


Figure 7.6: Final state of the example for single mode. The IRS describes the configuration of the EPS at t_6 “I built a cb.” Note the contracted now.

7.3 Building a corner cube bolting

In the next examples, the interaction with a human agent is taken into consideration. The robot needs help to build a corner cube bolting as she can hold only two things in her two grippers (remember that a corner cube bolting consists of a cube screwed to the corner hole of a slat). Thus she has to ask a human assistant for help to screw together slat, screw and cube to the configuration of a corner cube bolting.

An interaction that involves other agents' assistance requires procedures for the opening and closing of a discourse. The plan for opening a discourse ensures that the other participating agents are willing and able to participate in the discourse. My proposal for an opening plan executes the uttering of a greeting phrase, shares the main goal of the discourse (via the subplan `share-plan(K)`, figure 7.11) and checks whether the other agents can help in realizing the main goal of the discourse (subplan `check-mode(K)`, figure 7.15). The plan for closing a discourse mainly serves politeness considerations in that it informs the other agents about the result of the interaction. An example formulation of opening and closing plans is stated in the plans in figures 7.10 and 7.12.

After the execution of the plan for opening a discourse, the main part of the example for building a corner cube bolting consists of an execution of the plan for building a corner cube bolting which I display in figures 7.7, 7.8 and 7.9. This complex plan calls the plans for opening (`init(K)`) and closing (`finish(K)`) of a discourse with goal K at the beginning resp. end of the plan. The plan for building a corner cube bolting branches with respect to the type of assistance that can be expected from the other agents participating in the discourse. The decision which branch is to be executed depends on the know-how of the other agents as well as the mode of discourse that has been preset in the `initialize-state-procedure`. The plan `check-mode` (figures 7.14 and 7.15) collects the information necessary to decide which branch should be executed.

In addition, the plan for building a corner cube bolting has to consider that certain preconditions must be realized in order that the plan for screwing together three objects (pictured in figure 7.16) can be launched. This concerns in particular the correct position of the slat which the assisting human agent must hold between the cube and the screw such that the robot can screw it to the cube. The preconditions are realized via the verbal resolution of an IRS representation of the actions which must be performed so that the preconditions for screwing together three objects become a factual part of the EPS. The plan for the resolution of IRSs, `resolve(K)`, is pictured in figure 7.19.

Of course, errors may occur during the execution of plans. The resolution of errors is handled by a set of plans which are stated in section 7.4.5.

In addition, for the plans and the following analysis of examples, I employ work-arounds to deal with the limitations of this thesis with respect to the generation of utterances from IRSs and the construction of IRSs from utterances, which I did not specify in any detail. However, to display the role utterances play in the proposed framework, I assume that such generation and construction modules exist. If utterances occur, they are displayed as follows. When an IRS K is supposed to be constructed from an utterance UTT , this is represented as an EPS action `d-a:UTT,K`, where d is the utterer and a stands for an action. UTT is displayed at the SMS level as a sequence of words. The other way round, the generation of an utterance is represented as an external EPS action `ext-a:say utterance`, where `utterance` is either an IRS from which the utterance is to be generated or a sequence of words to be uttered. At the SMS level, the sequence of words corresponding to `utterance` is pictured.

7.4 Plans involved in the processing of examples

This section states the plans which are involved in the processing of example discourses for building a corner cube bolting. I redisplay the general structure of plans in definition 58. The role of plans was discussed in section 4.2.2, plans were formally defined as part of the EPS-formalism in section 6.2.

Definition 58 *Plan Schema*

A plan consists of the following parts:

Type:	The name of the plan
Invocation:	The triggering conditions for the execution of this plan. Acquisition of a new goal (active invocation) or a change in the EPS structure or a new IRS (reactive invocation)
Context:	The presuppositions that must be satisfied for the proper launching of the plan
Feedback:	The EPS or IRS which holds after the performance of the plan.
Body:	The specification of the plan in terms of a tree-like structure of actions and states of affairs. A successful branch of the plan is marked with a final leaf $END(+)$ an error branch of the plan is marked with $END(-)$

7.4.1 The main plan for building a corner cube bolting

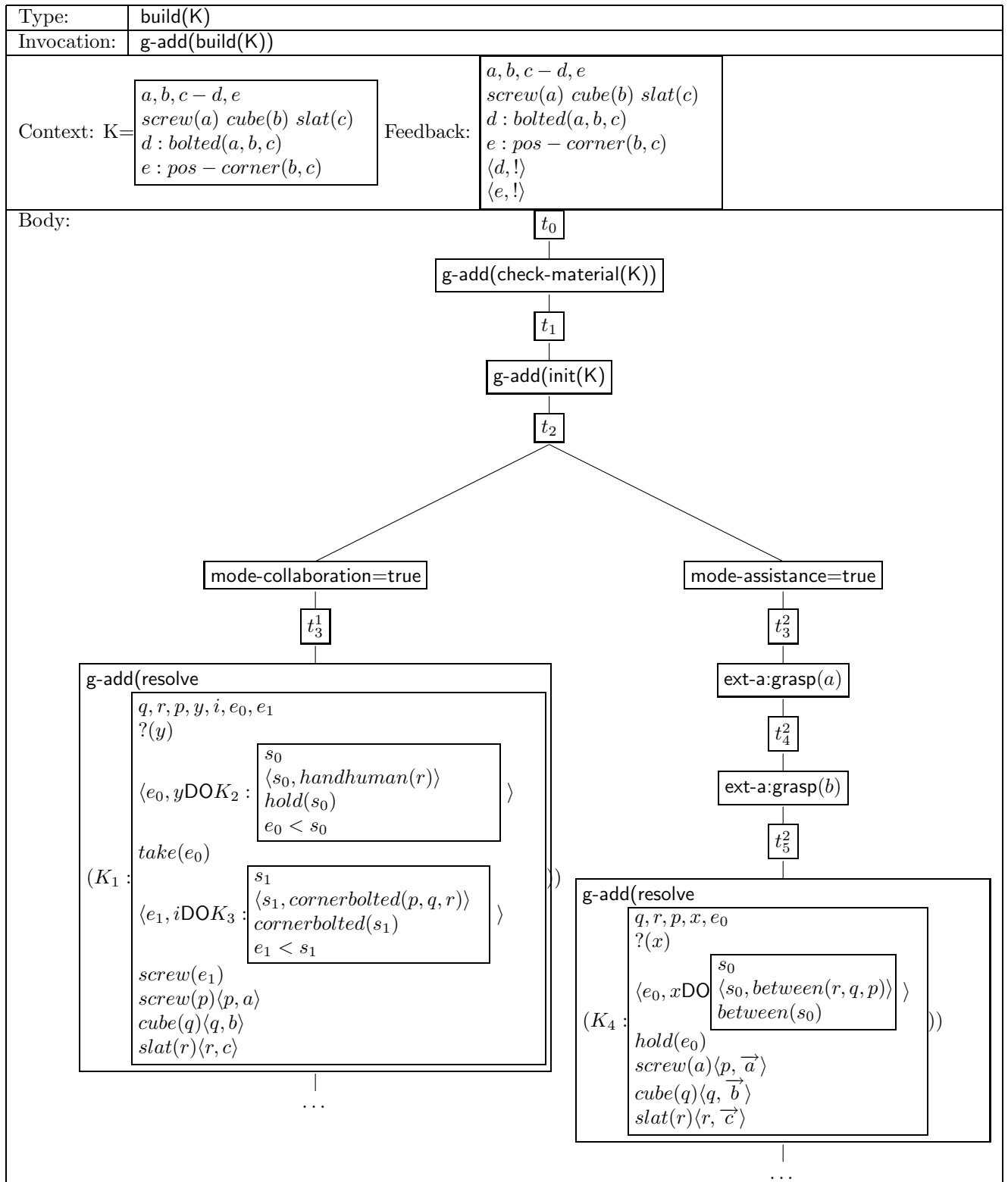


Figure 7.7: Plan for building a corner cube bolting part 1. The plan is invoked by the availability of a goal K , the main discourse goal. After the execution of initialization procedures, the plan branches with respect to the mode of interaction - either assistance or collaboration mode. Next, the preconditions for screwing together a cube and a slat are realized. I indicate the prime annotation of times with numbers.

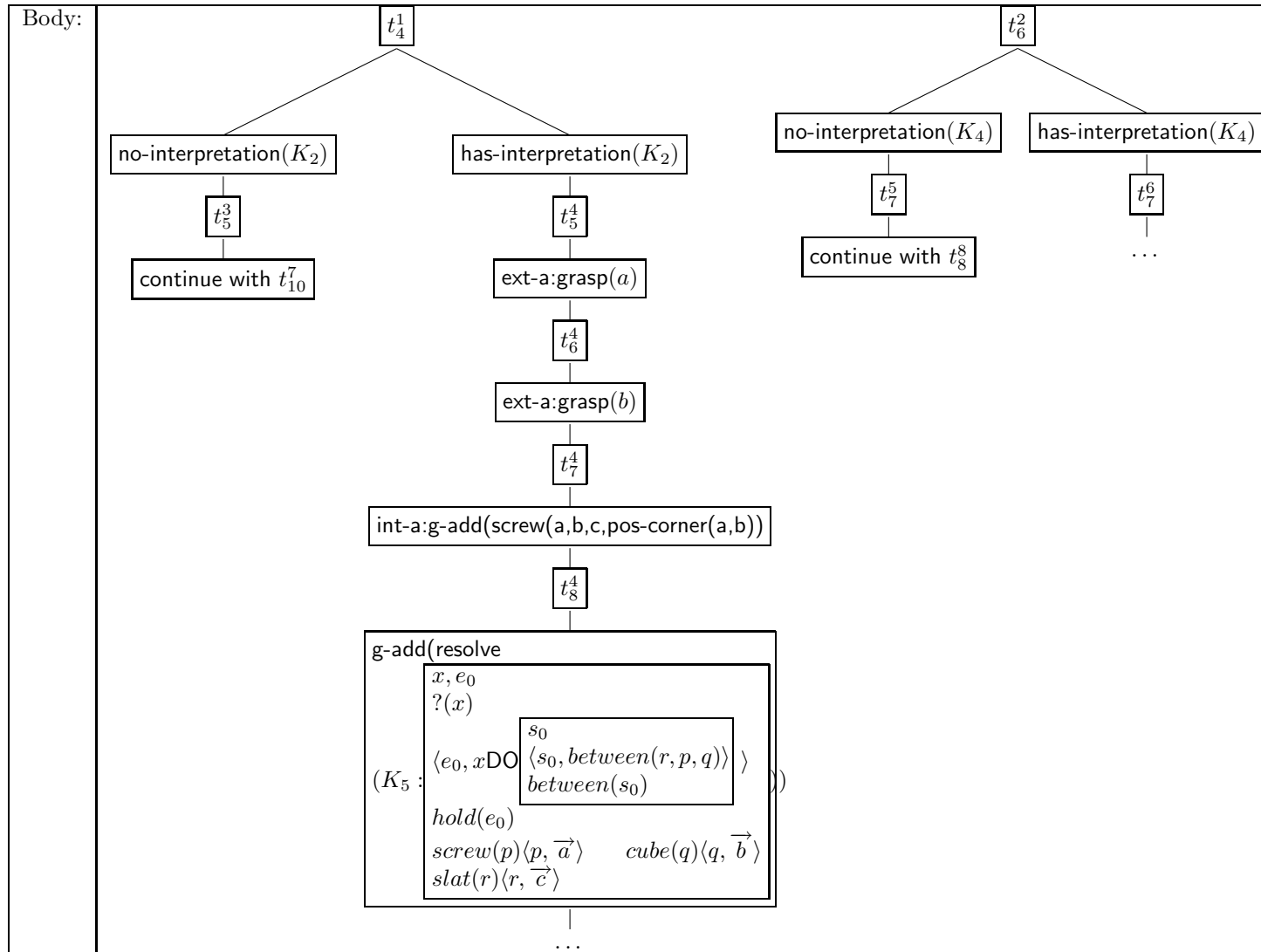


Figure 7.8: Plan for building a corner cube bolting part 2. Two branches for assistance (left) and collaboration mode (right). Assistance requires no know-how of the other agent, collaboration does require know-how.

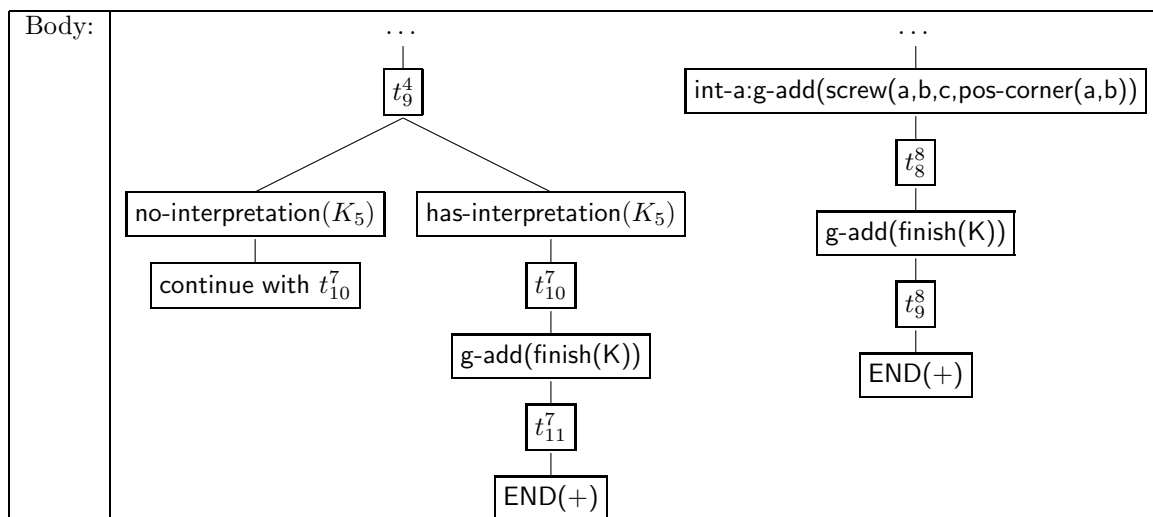


Figure 7.9: Plan for building a corner cube bolting part 3. If the preconditions for screwing together three objects are fulfilled, the actual action of screwing is executed. Finally, execution of the closing plan.

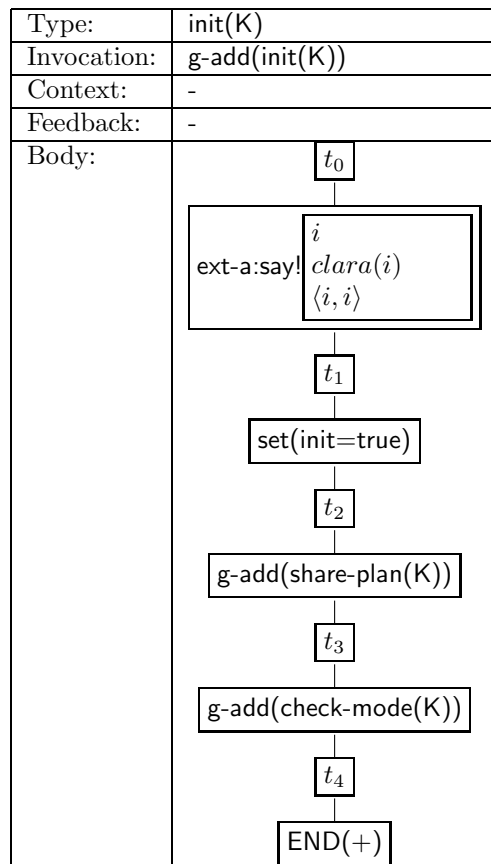


Figure 7.10: Plan for initializing a discursive interaction. The init-flag is necessary to prevent the execution of plans that require the assistance of other agents before the discourse is initiated.

7.4.2 Initialization and finishing of a discourse

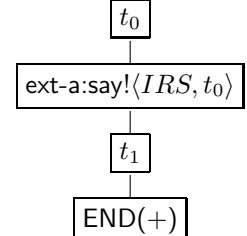
Type:	share-plan(K)
Invocation:	g-add(share-plan(K))
Context:	-
Feedback:	-
Body:	 <pre> graph TD t0[t0] --> say[ext-a:say!(IRS, t0)] say --> t1[t1] t1 --> end[END(+)] </pre>

Figure 7.11: Share the initial representation of plans and intentions corresponding to the overall discourse goal K . For the sake of simplicity, I assume that the discourse goal K is represented by the IRS at t_0 constructed by the `initialize-state` procedure.

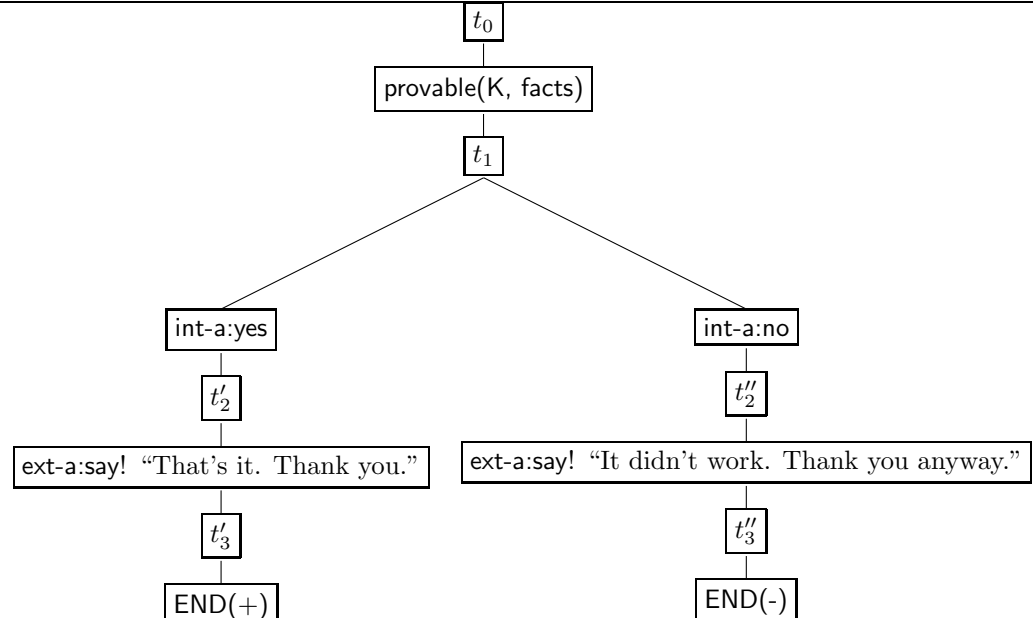
Type:	finish(K)
Invocation:	g-add(finish(K))
Context:	-
Feedback:	
Body:	 <pre> graph TD t0[t0] --> provable[provable(K, facts)] provable --> t1[t1] t1 --> int_yes[int-a:yes] t1 --> int_no[int-a:no] int_yes --> t2p[t'2] int_no --> t2m[t''2] t2p --> say_yes[ext-a:say! "That's it. Thank you."] t2m --> say_no[ext-a:say! "It didn't work. Thank you anyway."] say_yes --> t3p[t'3] say_no --> t3m[t''3] t3p --> end_plus[END(+)] t3m --> end_minus[END(-)] </pre>

Figure 7.12: Plan for finishing a discourse. The result of the interaction with respect to the main discourse goal K chooses the correct closing phrase based on the factual realization of the discourse goal.

Type:	grasp(K)	
Invocation:	g-add(K)	
Context:	$K_1 =$	s_1, x $\langle s_1, handrobot(x) \rangle$ $hold(s_1)$ $\langle x, a \rangle$ $\langle x, ! \rangle$
Feedback:	-	
Body:	<pre> graph TD A["t0]]"] --- B["grasp(a)"] B --- C["t1"] C --- D["END(+)"] </pre>	

Figure 7.13: Plan for grasping K , active invocation by the addition of a goal state IRS K . K_1 is an IRS representing K .

7.4.3 Plans involved in the construction of a corner cube bolting

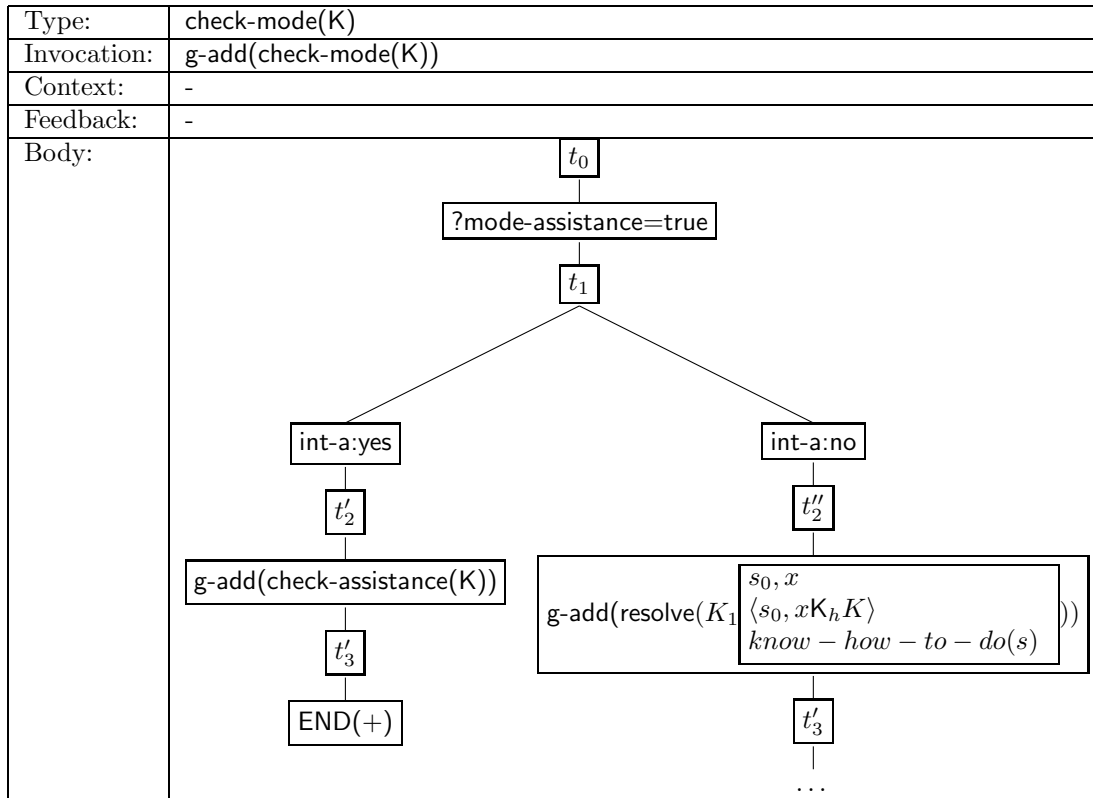


Figure 7.14: Plan for determining the mode of interaction part 1. Three possibilities are captured: plain assistance if preset by the initialization procedure and depending on the know-how of the other agent plain collaboration mode or collaboration mode with teaching.

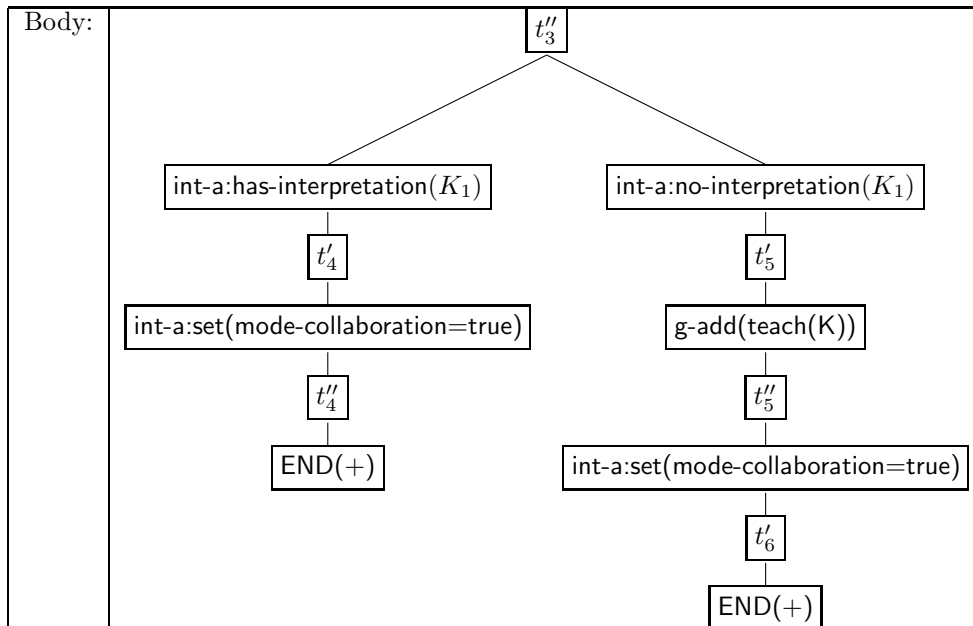


Figure 7.15: Plan for determining the mode of interaction Part 2. Depending on the know-how of the collaborator, either collaboration or teaching mode is activated.

Type:	screw(K)	
Invocation:	g-add(screw(a,b,c,pos-corner(a,b)))	
Context:	K=	$a, b, c - d$ $d : \text{between}(c, a, b)$ $\text{cube}(a)$ $\text{screw}(b)$ $\text{slat}(c)$ $\langle d, ! \rangle$
Feedback:		$a, b - e, f$ $e : \text{pos} - \text{corner}(a, b)$ $\langle e, ! \rangle$ $f : \text{bolted}(a, b, c)$ $\langle f, ! \rangle$
Body:	<div style="text-align: center;"> t_0 ext-a:screw(a,b,c,pos-corner(a,b)) t_1 END(+) </div>	

Figure 7.16: Plan for putting together three objects to the configuration of a corner cube bolting. The context condition requires that the slat is placed between the cube and the screw which the robot holds in her hands.

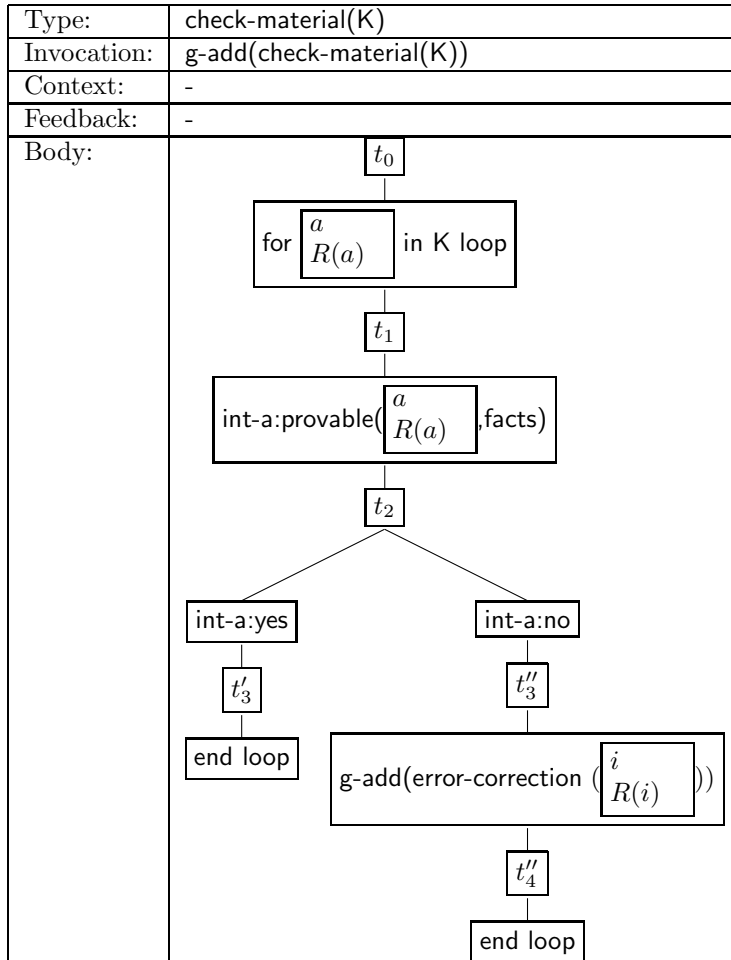


Figure 7.17: Plan for checking the factual availability of the objects involved in a plan. This plan iterates through the non-relational referents of the EPS K (the left side of the dash '—' in an EPSs universe) and checks whether they are externally anchored.

7.4.4 Resolution of IRSs, anchors and errors

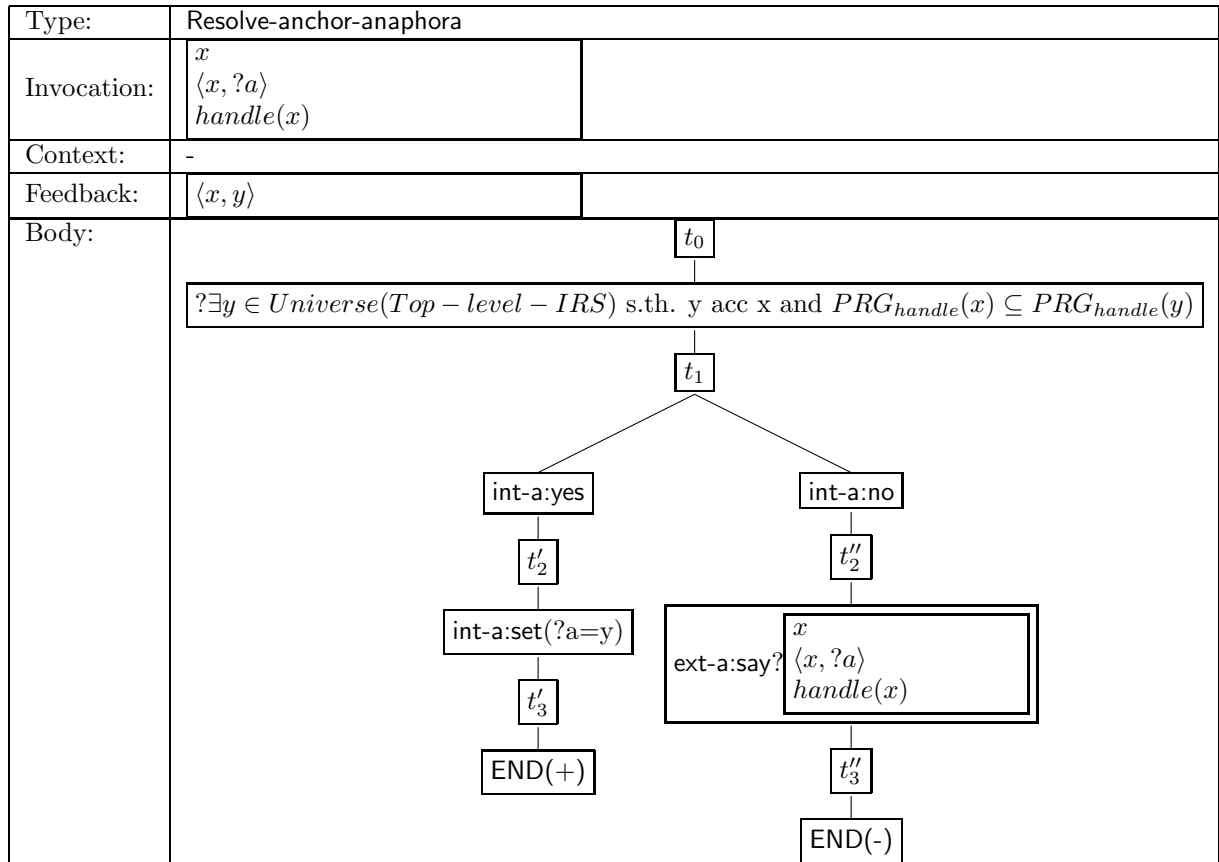


Figure 7.18: Identify an anaphoric variable anchor source. The plan searches for an accessible antecedent in the universe of the top-level IRS and contained IRSs that matches the identification conditions of anaphoric referent x .

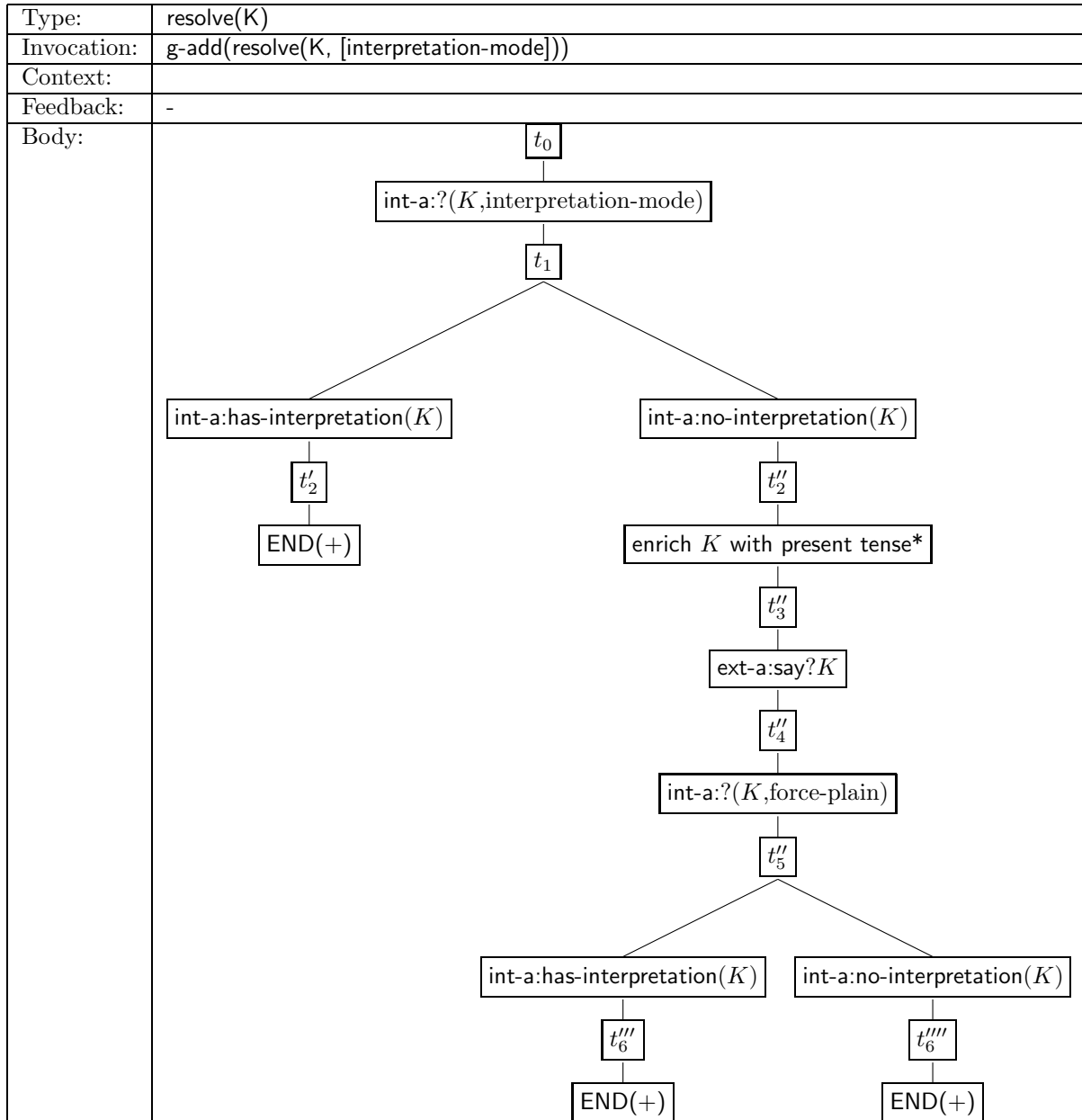


Figure 7.19: Resolve an IRS K . The plan has an optional argument to force plain interpretation mode. The step indicated with a star (*) is a somewhat tricky point. In the examples discussed later I, the resolution of IRSs concerns only present state of affairs, so it is sufficient to simply turn any IRS to be resolved into an IRS where the contained time-individuals are arranged according to the definition of present tense. However, if IRSs concerning the past have to be resolved too, (*) has to be revised into a more complicated algorithm.

Type:	Resolve-anchor-belief(K)
Invocation:	x $\langle x, ? \rangle$ $handle(x)$
Context:	-
Feedback:	$\langle x, a \rangle$
Body:	<pre> graph TD t0[t0] --> B1[int-a:provable(PRG_handle(a), beliefs)] B1 --> t1[t1] t1 --> B2[int-a:yes] t1 --> B3[int-a:no] B2 --> t2p[t'2] t2p --> B4[set(<x, a>)] B4 --> t3p[t'3] t3p --> B5[END(+)] B3 --> t2m[t''2] t2m --> B6[b-add(PRG_handle)] B6 --> t3m[t''3] t3m --> B7[END(+)] </pre>

Figure 7.20: Identify a variable belief anchor source. If it is not possible to identify the source in the set of **beliefs**, the floater's identification conditions are added to the set of **beliefs**. If several source candidates for x can be identified, they should be returned as a set of possible sources.

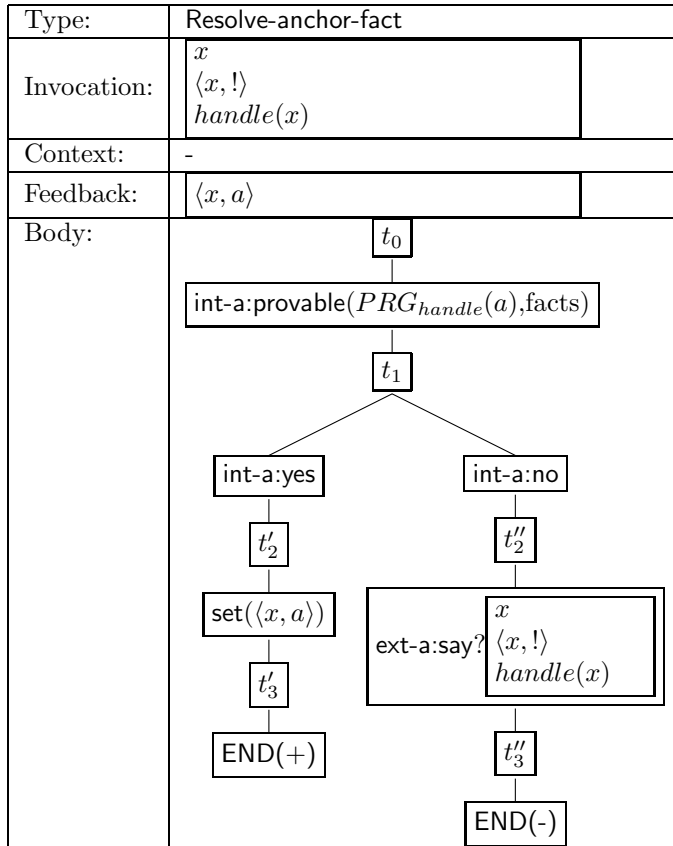


Figure 7.21: Resolve a variable fact anchor source. If it is not possible to identify the floater with an external anchor chain, the problem is uttered in the hope that some other agent can solve it.

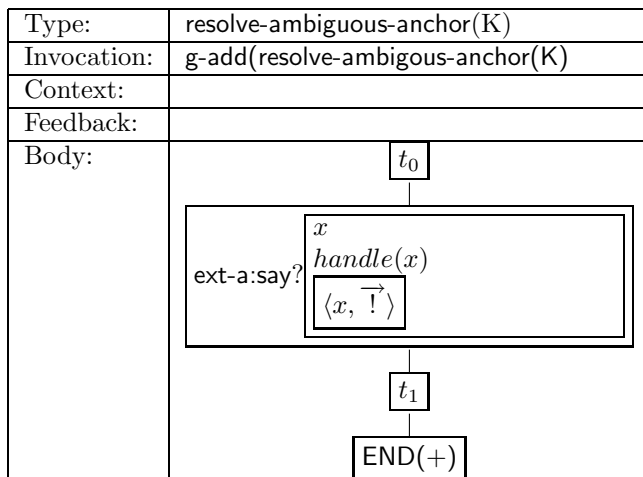


Figure 7.22: Resolve an ambiguous definite anchor source via an utterance that describes the problem.

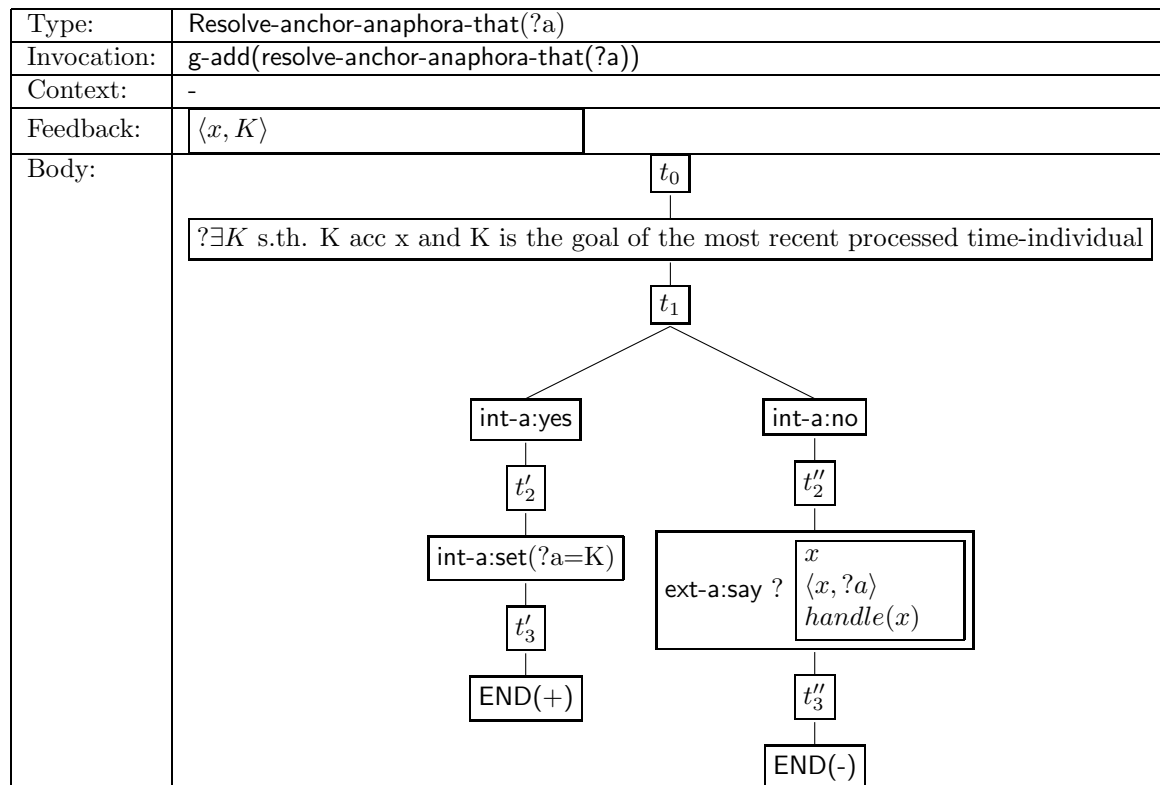


Figure 7.23: Resolve an anaphoric anchor 'that'. This is a dummy solution for which I guess that it can handle at least the cases of interest to this thesis, e.g. "Do you know how to do that?".

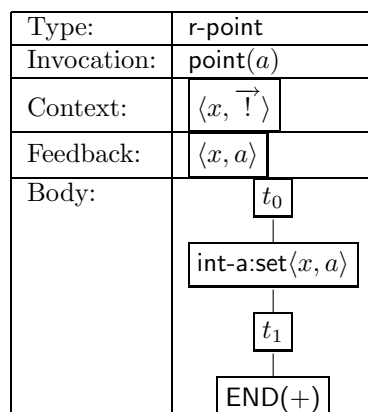


Figure 7.24: Plan for resolving a definite variable anchor source with a pointing gesture.

Type:	error-correction	
Invocation:	error-correction(K)	
Context:	K =	$\begin{matrix} a \\ R(a) \end{matrix}$
Feedback:	-	
Body:	<div style="text-align: center;"> t_0 <div style="border: 1px solid black; padding: 5px; display: inline-block;"> <i>ext-a:say!</i> $\langle e, iDOK \rangle$ <i>need(e)</i> </div> t_1 END(+) </div>	

Figure 7.25: Error correction for a missing object via an utterance that describes the problem.

Type:	error-correction	
Invocation:	error-correction(handle(e)) or error-correction(handle(s))	
Context:		
Feedback:	-	
Body:	<div style="text-align: center;"> t_0 <div style="border: 1px solid black; padding: 5px; display: inline-block;"> <i>ext-a:say!</i> "I don't think so. Can you explain that to me?" </div> t_1 END(+) </div>	

Figure 7.26: Error correction for a mismatch in interpretation. Again, this is a dummy solution that handles cases such as the incorrect reference to facts.

7.4.5 Resolution of errors

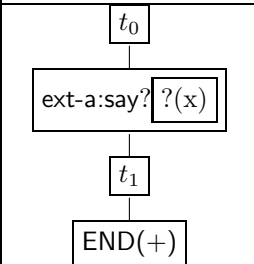
Type:	resolve-handle(?(<i>x</i>))
Invocation:	?(<i>x</i>)
Context:	init=true; $\langle x, a \rangle$
Feedback:	-
Body:	

Figure 7.27: Error correction for a missing handle, active resolution. In the present setup, thing-individuals can be constructed without knowing their name (e.g. for persons). Missing handles must be resolved before further interpretation can take place.

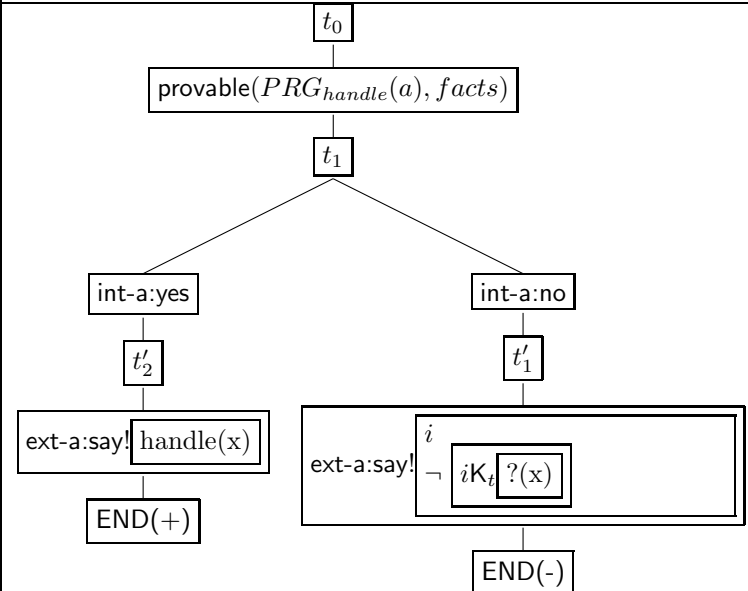
Type:	resolve-handle(?(<i>x</i>))
Invocation:	?(<i>x</i>)
Context:	$\langle x, a \rangle$ occurs in an IRS constructed from an utterance
Feedback:	
Body:	

Figure 7.28: This plan with reactive invocation handles occurrences of missing handles in an IRS constructed from an utterance, e.g. “What is your name?”.

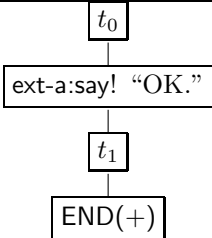
Type:	confirm!
Invocation:	has-interpretation(K)
Context:	K stems from a non-interrogative utterance: !UTT
Feedback:	
Body:	 <pre> graph TD t0[t0] --> say[ext-a:say! "OK."] say --> t1[t1] t1 --> end[END(+)] </pre>

Figure 7.29: Plan for OK-Feedback. This plan is invoked by a successful interpretation of an utterance !UTT.

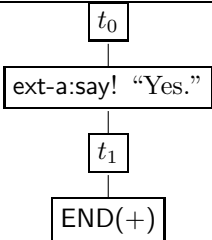
Type:	confirm?
Invocation:	has-interpretation(K)
Context:	K stems from an interrogative: ?UTT
Feedback:	
Body:	 <pre> graph TD t0[t0] --> say[ext-a:say! "Yes."] say --> t1[t1] t1 --> end[END(+)] </pre>

Figure 7.30: Plan for Yes-Feedback. This plan is invoked by a successful interpretation of a question ?UTT.

7.4.6 Feedback

Type:	confirm!
Invocation:	has-interpretation(K)
Context:	\neg UTT
Feedback:	
Body:	<pre> graph TD t0[t0] --> say[ext-a:say! "OK."] say --> t1[t1] t1 --> end[END(+)] </pre>

Figure 7.31: Plan for OK-Feedback. This plan is invoked by a successful interpretation of an IRS K which is not derived from an utterance. This can be considered as a soliloquy but also informs other discourse participants about own internal states.

Type:	meta-collect-confirm
Invocation:	confirm!
Feedback:	
Body:	<pre> graph TD t0[t0] --> if["if next-action(Intentions) is confirm! then delete confirm!"] if --> t1[t1] t1 --> end[END(+)] </pre>

Figure 7.32: Metaplan for collecting OK-Feedback. This plan deletes manifold occurrences of `confirm!` feedback on the intention stack e.g. if the result of an action renders possible the successful interpretation of several IRSs.

7.5 Assistance mode

The example interaction for assistance mode is stated in example 14 and this section gives an analysis of how this example is processed within the framework of GDRT.

The main plan underlying the analysis of examples in the next sections is the plan for building a corner cube bolting. This plan is invoked by the adoption of a main discourse goal K represented with the

following EPS:

$a, b, c - d, e$
$screw(a) \text{ cube}(b) \text{ slat}(c)$
$d : bolted(a, b, c)$
$e : pos - corner(b, c)$

. The preliminary steps of goal acquisition and initialization

were discussed in section 7.1.4. Note that in this example, the initialization procedure presets the discourse mode to **mode-assistance** to prefer assistance mode over collaboration mode even if the other agent has know-how.

Example 14 *Example interaction for assistance mode*

A:	<i>The table holds a cube, a slat and a screw. The robot is in standby mode. The user approaches the table and stops in front of the robot. The robot opens her eyes and looks at the user. The robot does not know the user's name. Plain assistance is preset and the discourse goal is about building a corner cube bolting.</i>
B:	Hi, my name is Clara. What is your name?
A:	My name is Tillmann.
A:	OK!
A:	I am going to build a corner cube bolting. Can you assist me with that, please?
B:	Yes.
A:	OK!
A:	<i>A grasps the cube and the screw.</i>
A:	Please hold the slat between the cube and the screw!
A:	<i>B grabs the slat and holds it between the cube and the screw.</i>
A:	OK!
A:	<i>The robot screws the cube into the corner hole of the slat.</i>
A:	OK!
A:	That's it. Thank you.

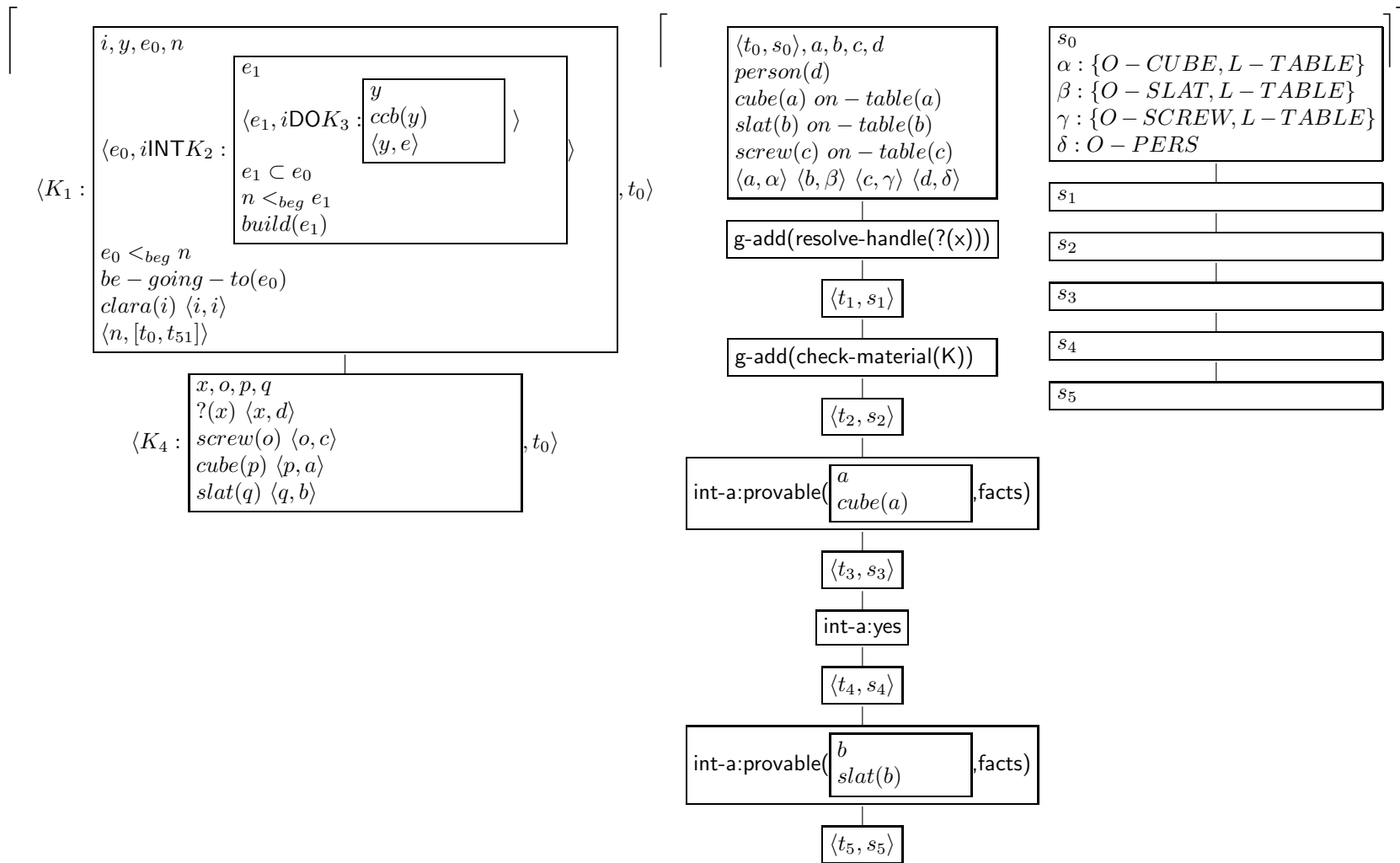


Figure 7.33: Discourse analysis of assistance mode, Part 1. K_1 (representing the current internal state of clara) and K_4 (displaying the current external state for clara) picture the result of `initialize-state`. The main goal K is about building a corner cube bolting (as instantiated with the main plan for building a ccb), its IRS representation is K_3 . The interaction starts with an execution of the plan `check-material(K)` that checks the availability of the needed materials. The resolution of the missing handle $\?x$ is delayed until the discourse is initiated.

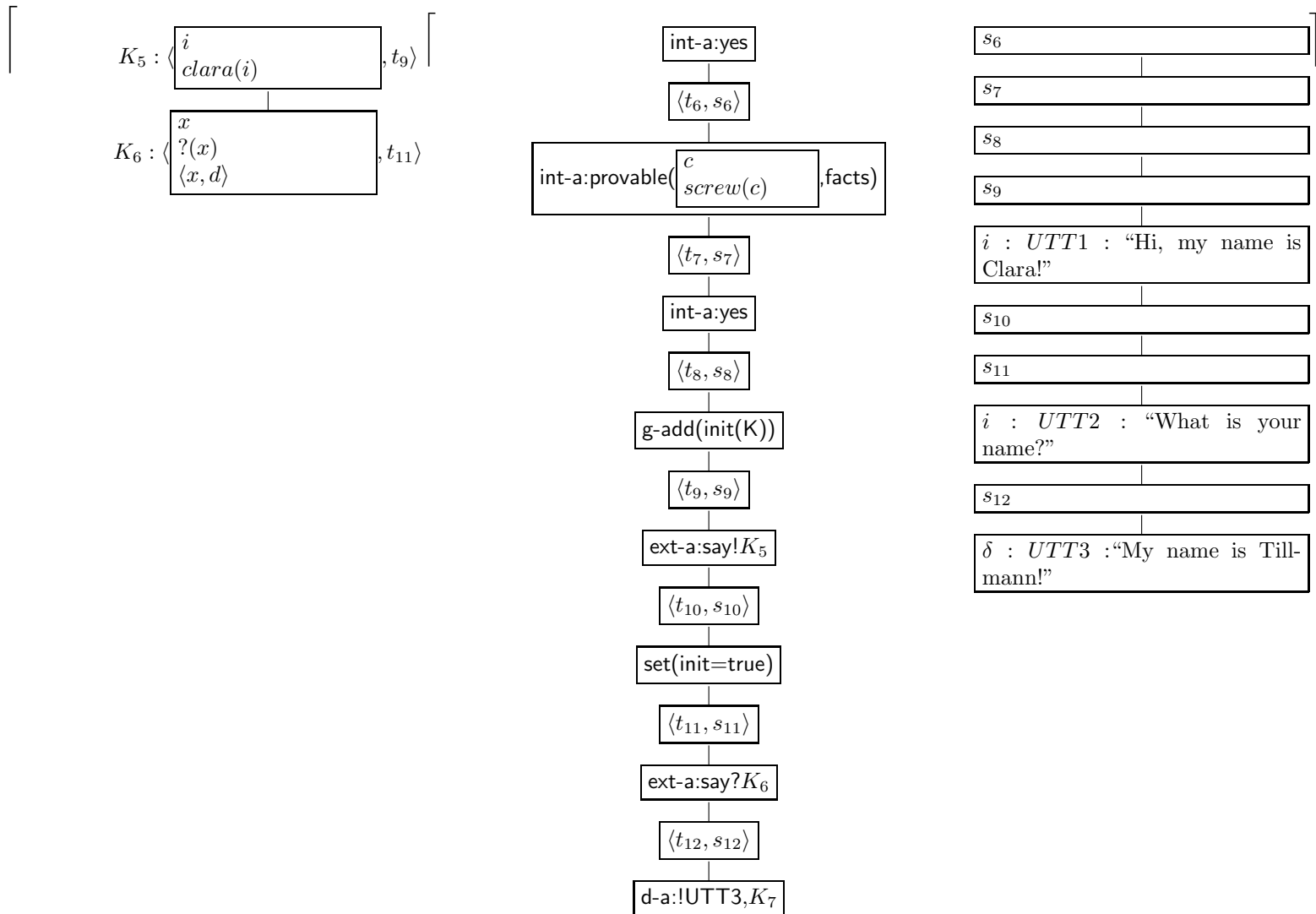


Figure 7.34: Discourse analysis of assistance mode, Part 2. After the successful execution of the material-check, the init-plan for the discourse is invoked that sets the init-flag to true, which allows to execute the resolution of the missing handle $?(x)$ for Tillmann.

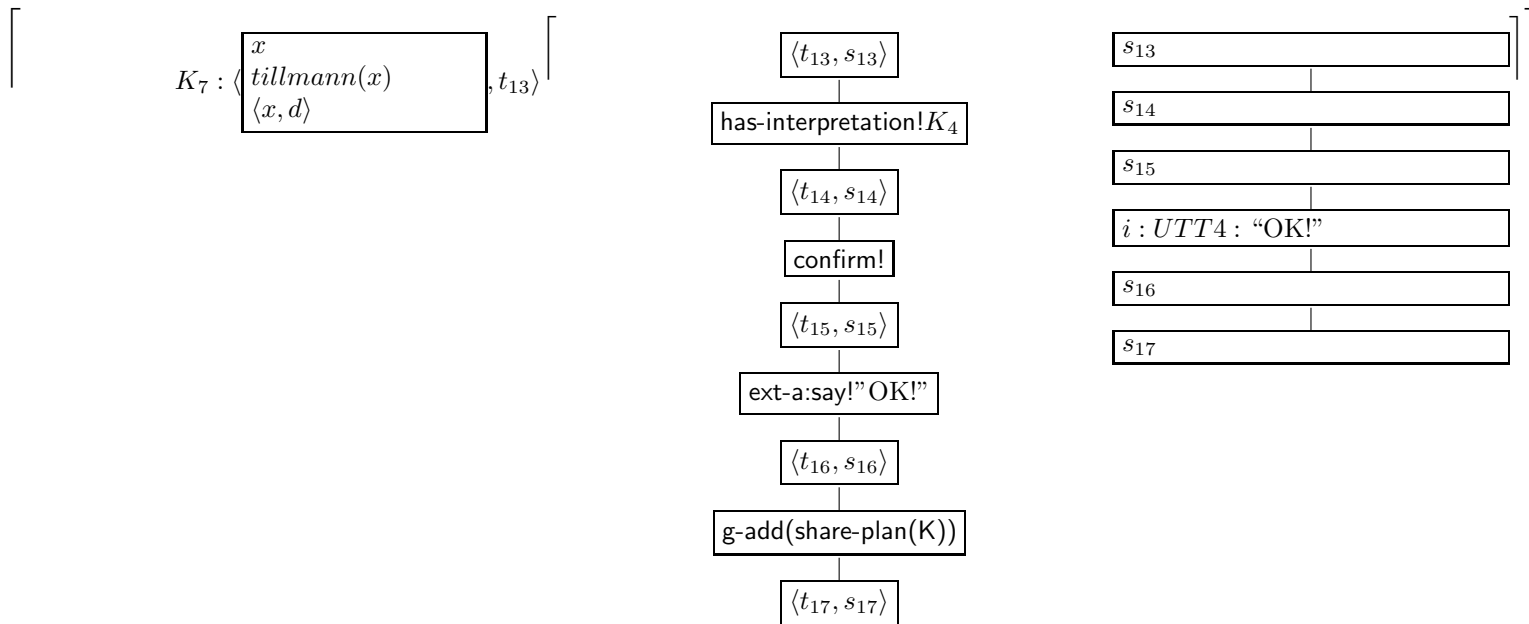


Figure 7.35: Discourse analysis of assistance mode, Part 3. The missing handle $?(x)$ is resolved via K_7 constructed from $UTT3$. The plan to share the main discourse goal as stated in IRS K_1 is activated.

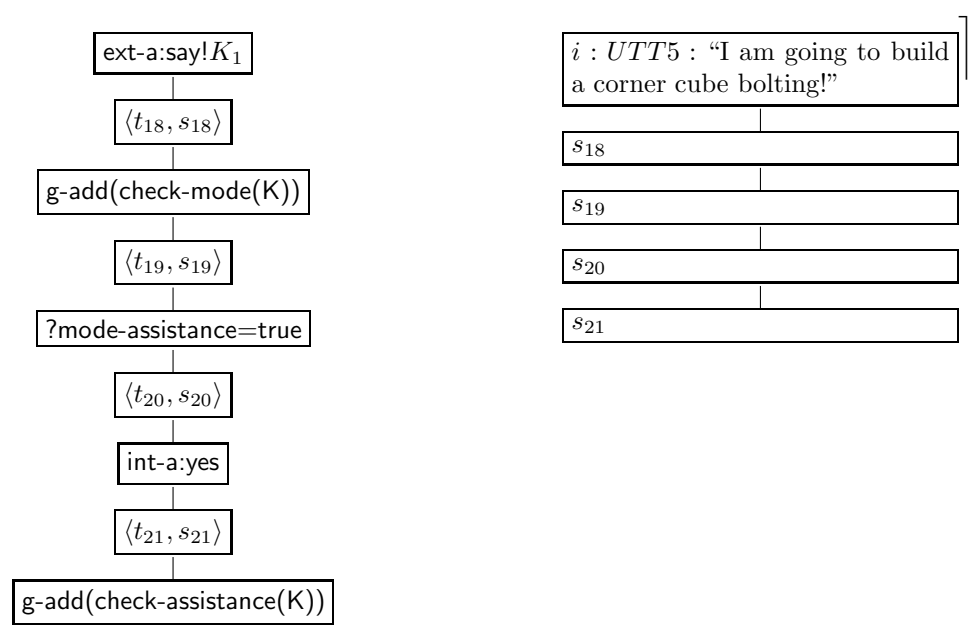


Figure 7.36: Discourse analysis of assistance mode, Part 4. Further execution of the plan for building a ccb. As mode-assistance is initially true, the plain-assistance branch of the check-mode plan triggered by the plan for building a ccb is chosen.

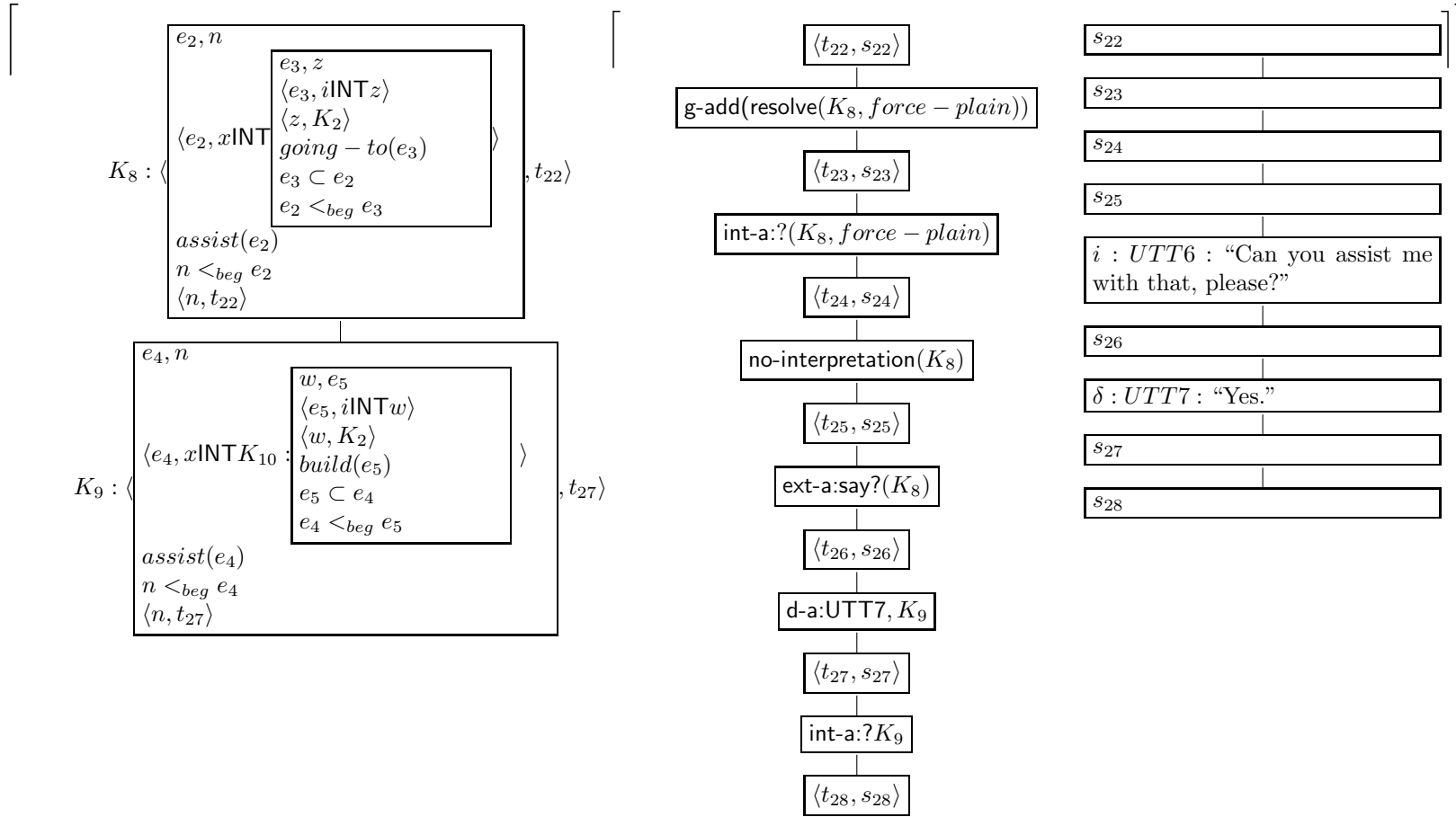


Figure 7.37: Discourse analysis of assistance mode, Part 5. Checking for the help of the human assistant via a verbal resolution of IRS K_8 .

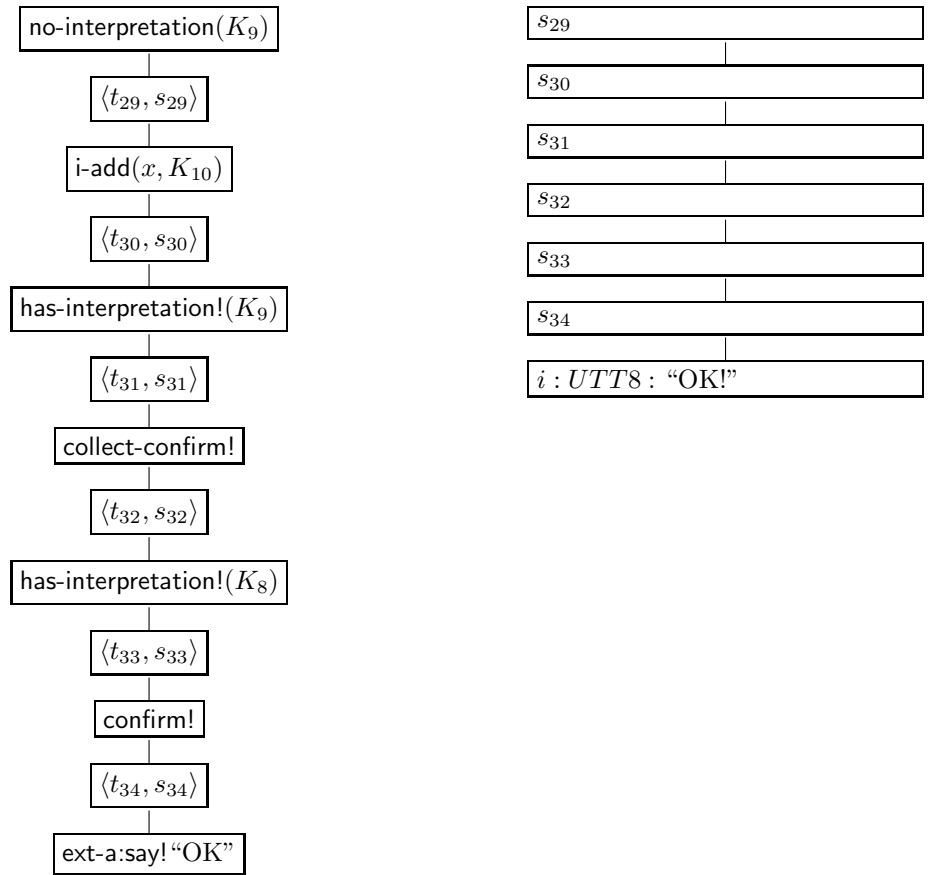


Figure 7.38: Discourse analysis of assistance mode, Part 6. Addition of x 's intention to assist Clara to x 's intention stack via the reactive interpretation of K_9 . Consequently, the pending IRSs K_9 and K_8 have an interpretation.

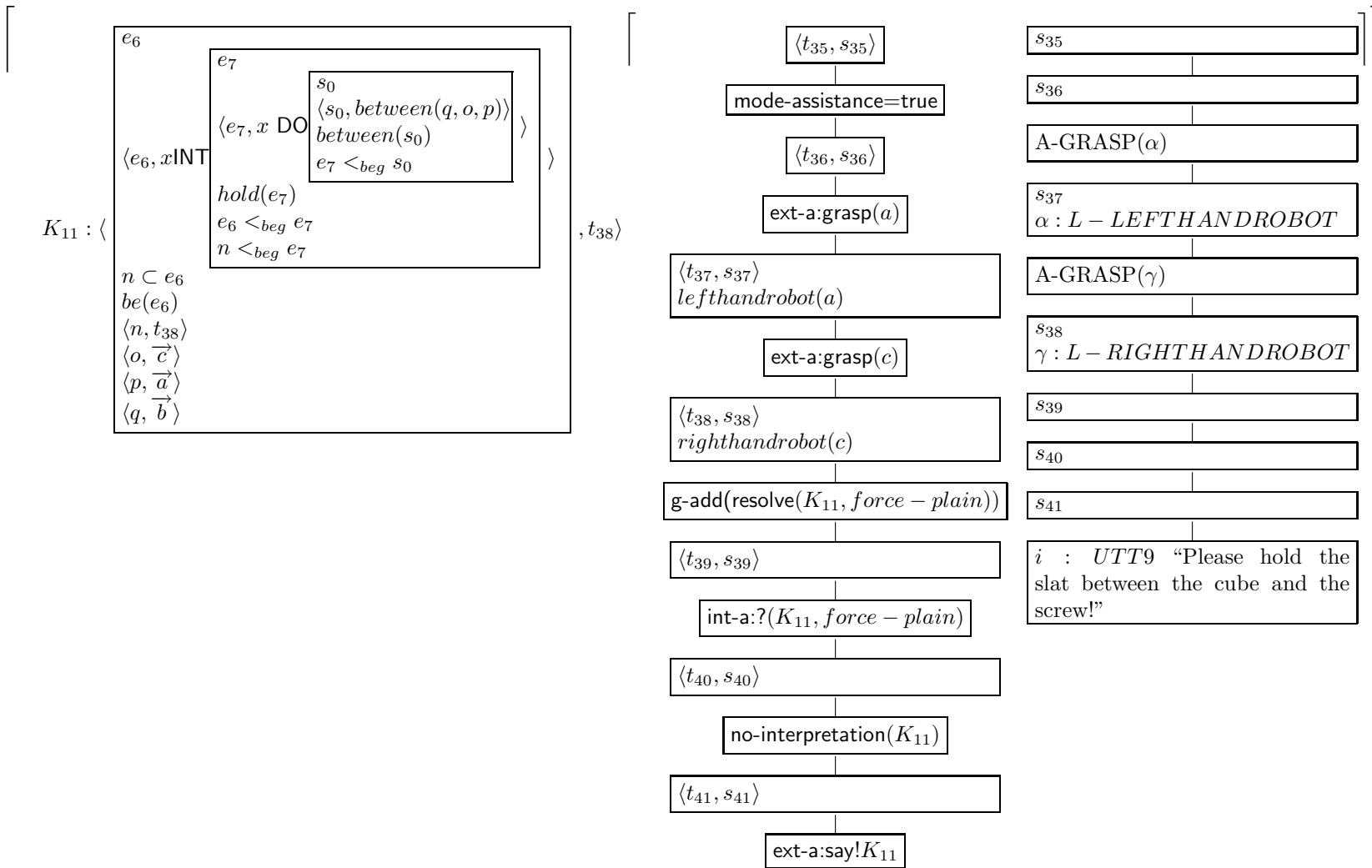


Figure 7.39: Discourse analysis of assistance mode, Part 7. The command to place the slat between the cube and the screw is uttered to the user. This is a precondition to the screwing of three objects together to the configuration of a corner cube bolting.

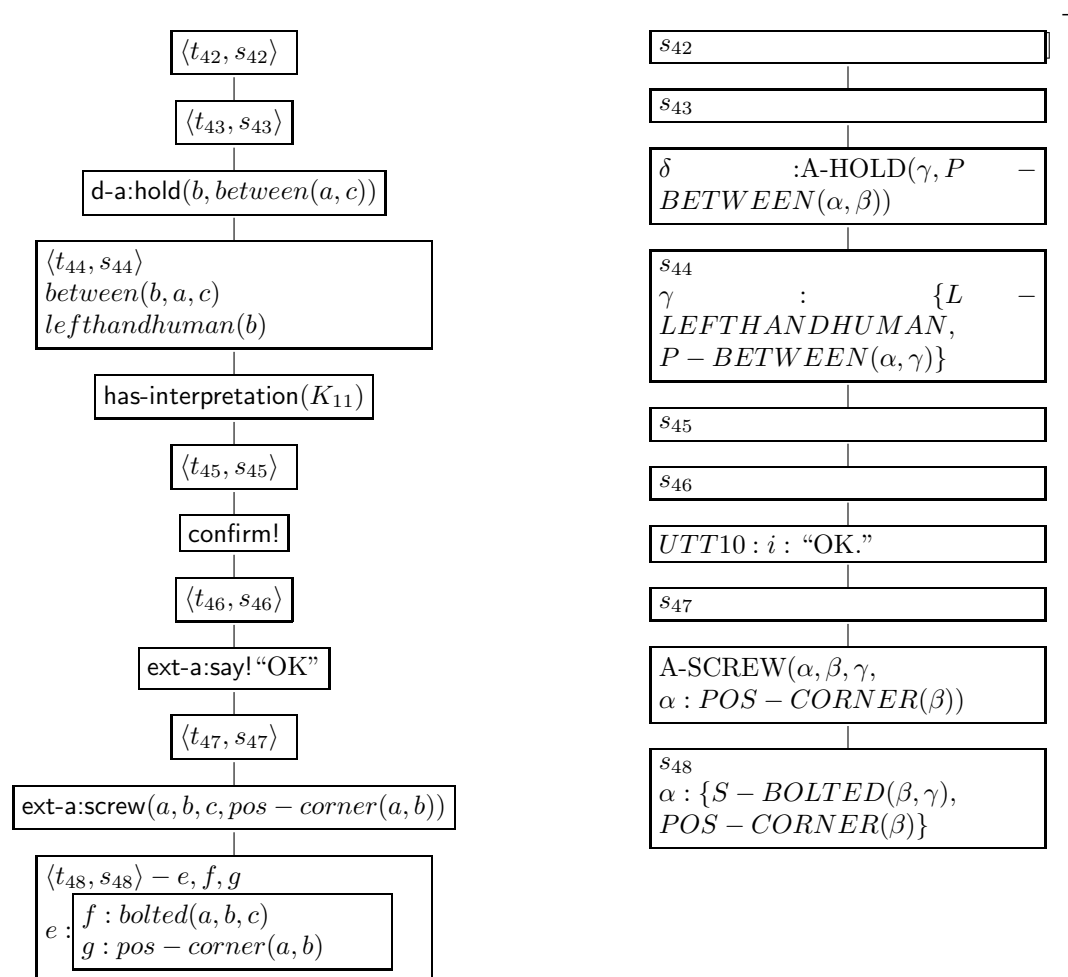


Figure 7.40: Discourse analysis of assistance mode, Part 8. The reaction of the user to *UTT9* renders possible the interpretation of IRS K_{11} . In turn, this launches the screwing together of the corner cube bolting.

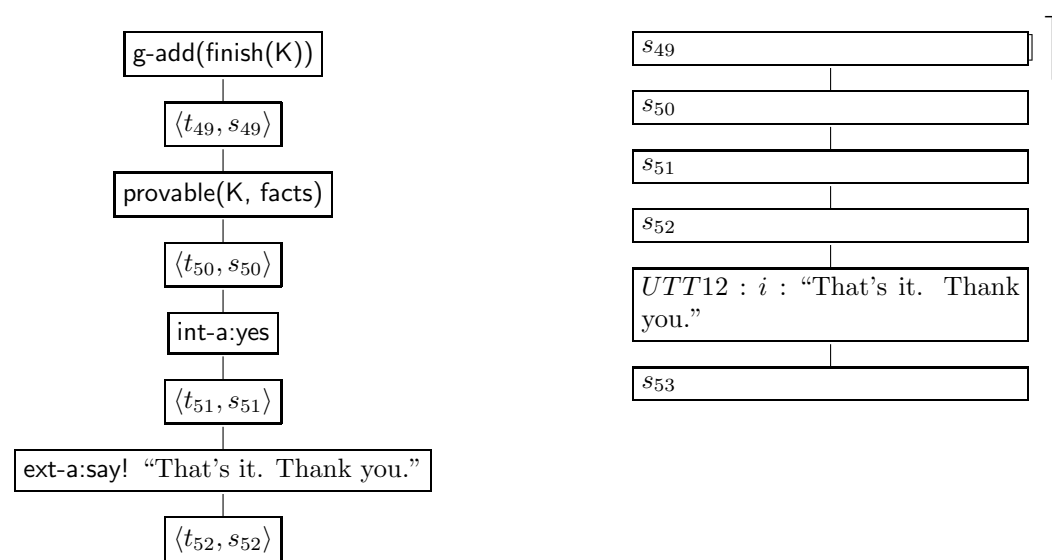


Figure 7.41: Discourse analysis of assistance mode, Part 9. The construction of the ccb renders the realization of the discourse goal true. Execution of the finishing sequence.

7.6 Collaboration mode

It distinguishes collaboration mode that plans and their executions are shared by the participants. That is, the work to be done must be distributed among the participants. Example 15 illustrates the sharing of work in collaboration mode, where the robot initiates the cooperation and distributes work based on the collaboration branch of the plan for building a corner cube bolting.

Example 15 *Example dialog for collaboration mode.*

The table holds a cube, a slat and a screw. The user approaches the table and stops in front of the robot. The robot opens her eyes and looks at the user. The robot does not know the user's name.

A: Hi, my name is Clara. What is your name?
B: My name is Tillmann.
A: I am going to build a corner cube bolting. Do you know how to do that?
B: Yes, I know how to do that.
A: OK. Then you take the slat and I screw it into the cube.
B grasps the slat.
A: OK.
A grasps the cube and the screw.
A: Please hold the slat between the cube and the screw.
B holds the slat between the cube and the screw.
A: OK.
A screws the cube into the slat.
A: OK. That's it. Thank you.

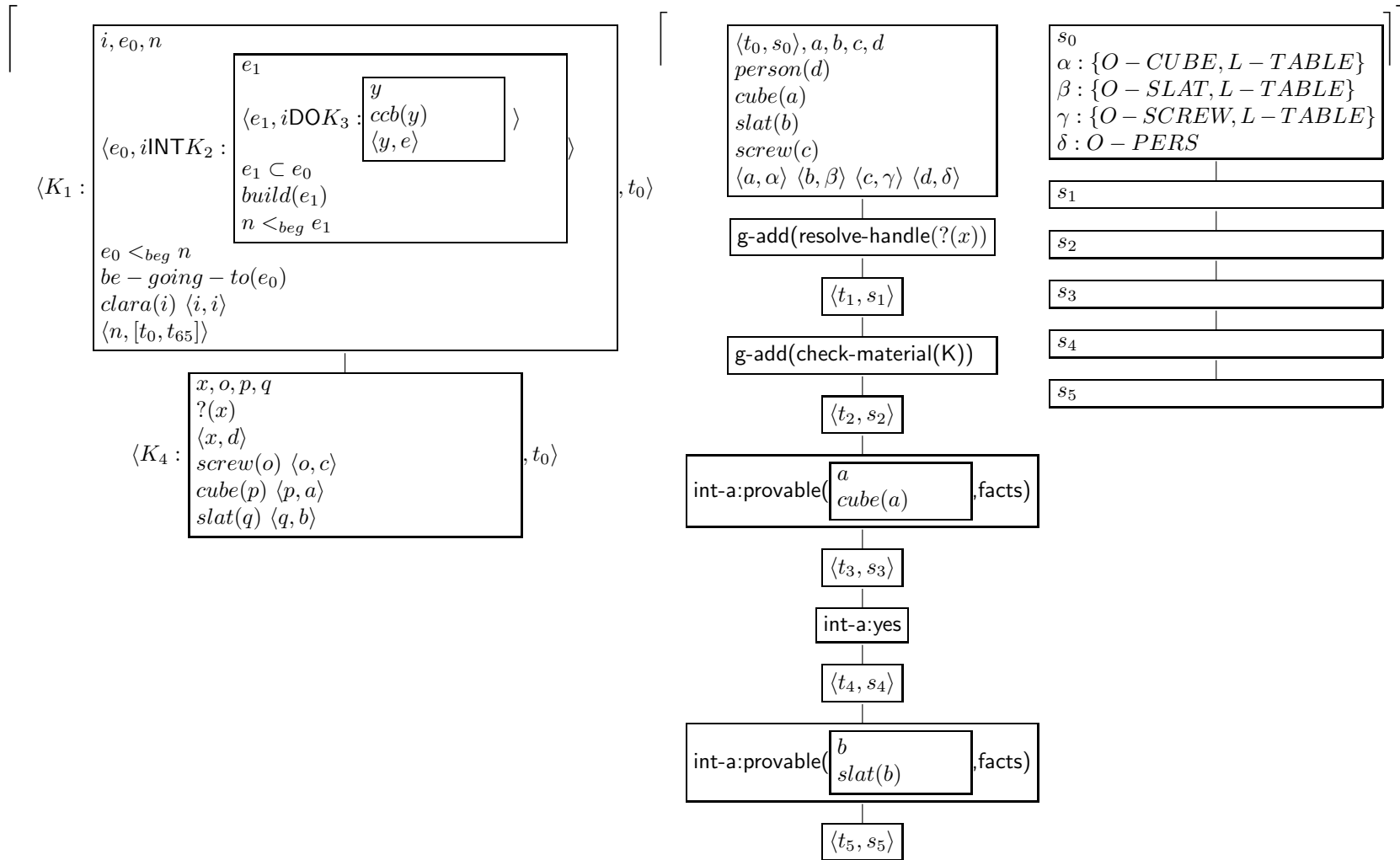


Figure 7.42: Discourse analysis of collaboration mode, Part 1. The plan for resolving the missing handle $?(x)$ is delayed until the init-flag is set to true by the plan for the initialization of a discourse.

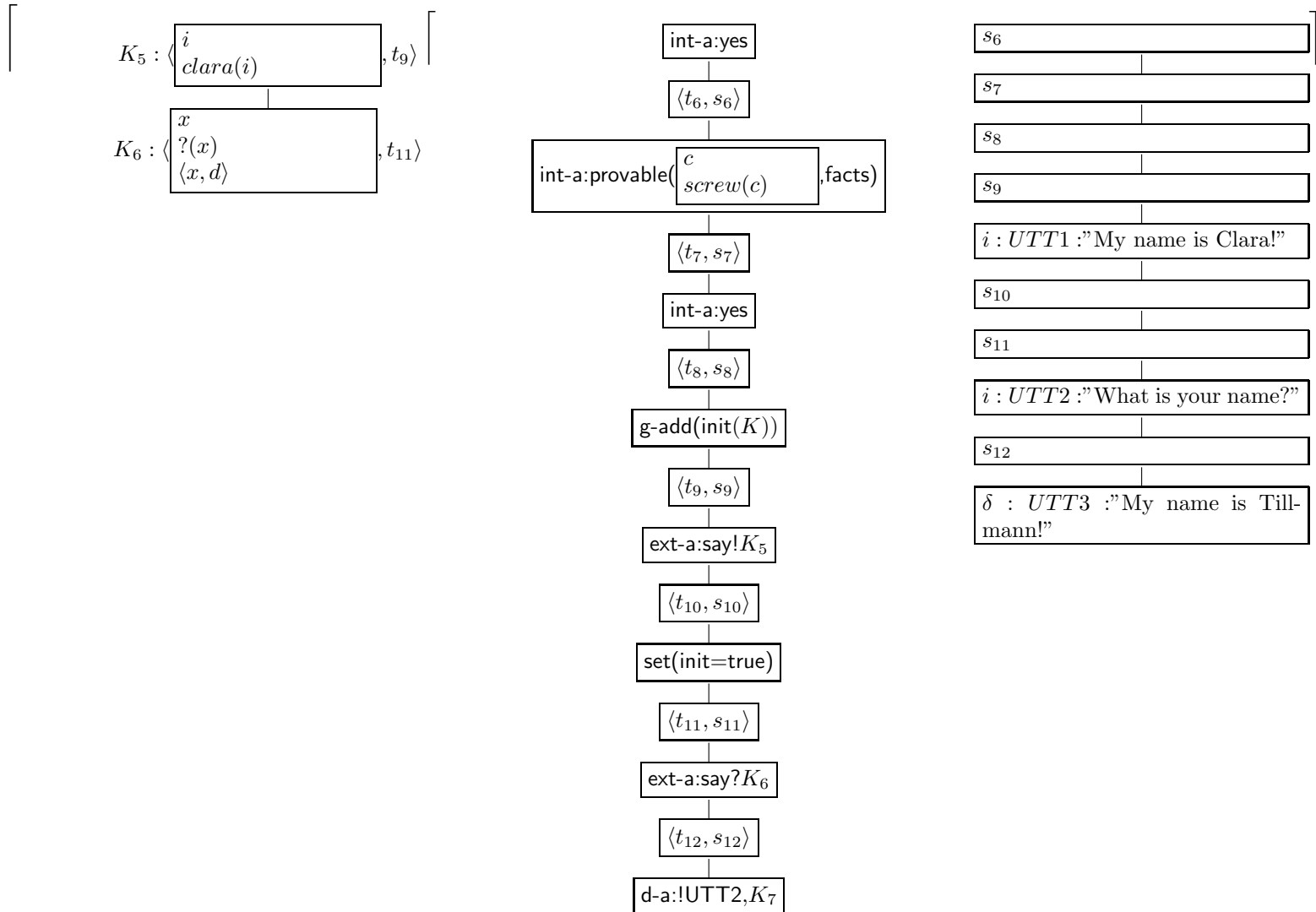


Figure 7.43: Discourse analysis of collaboration mode, Part 2. Initialization of the discourse. $UTT3$ resolves the missing handle $?(x)$.

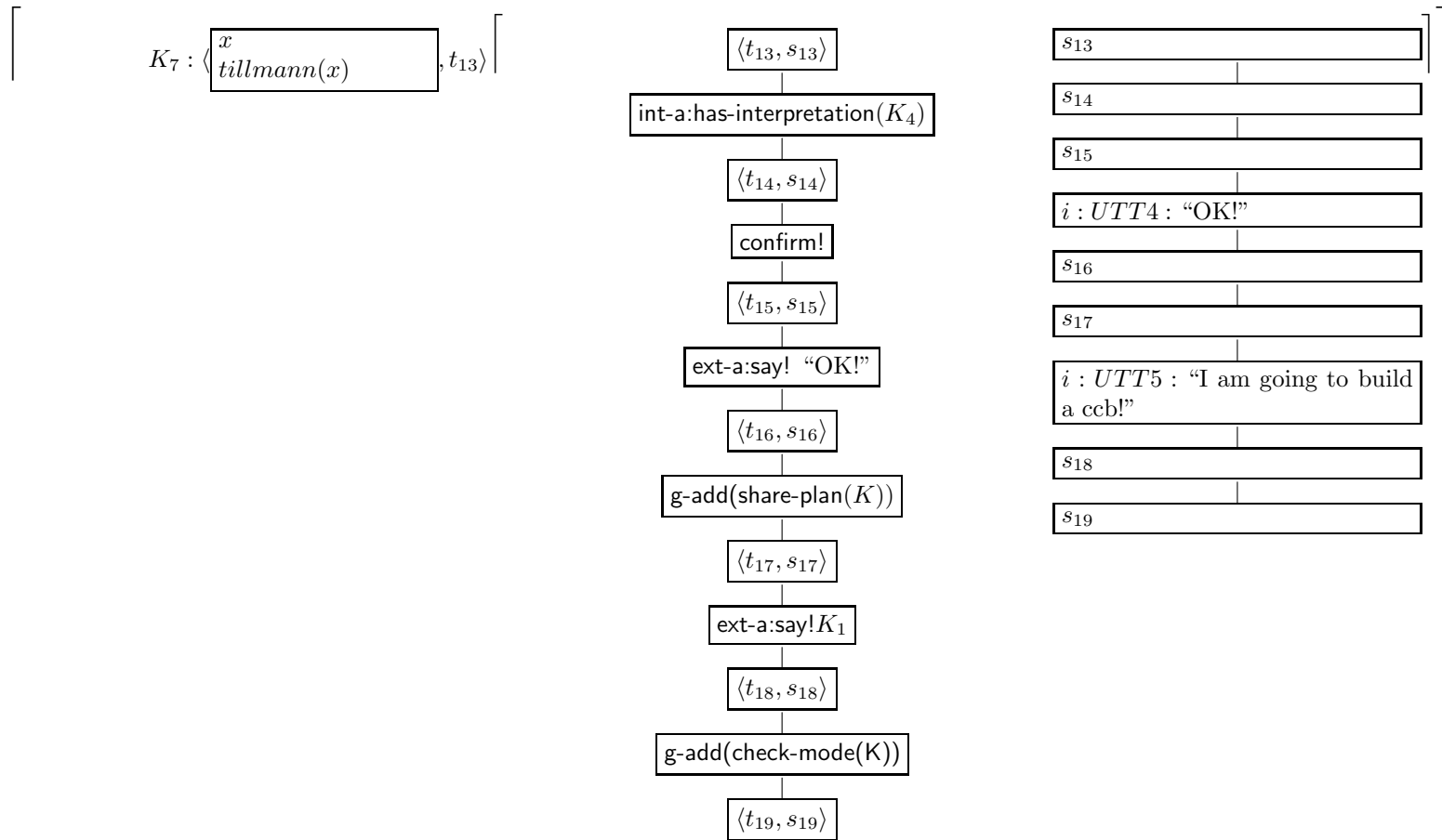


Figure 7.44: Discourse analysis of collaboration mode, Part 3. Confirmation of the successful interpretation of K_4 . Sharing of the initial representation K_1 of the discourse goal.

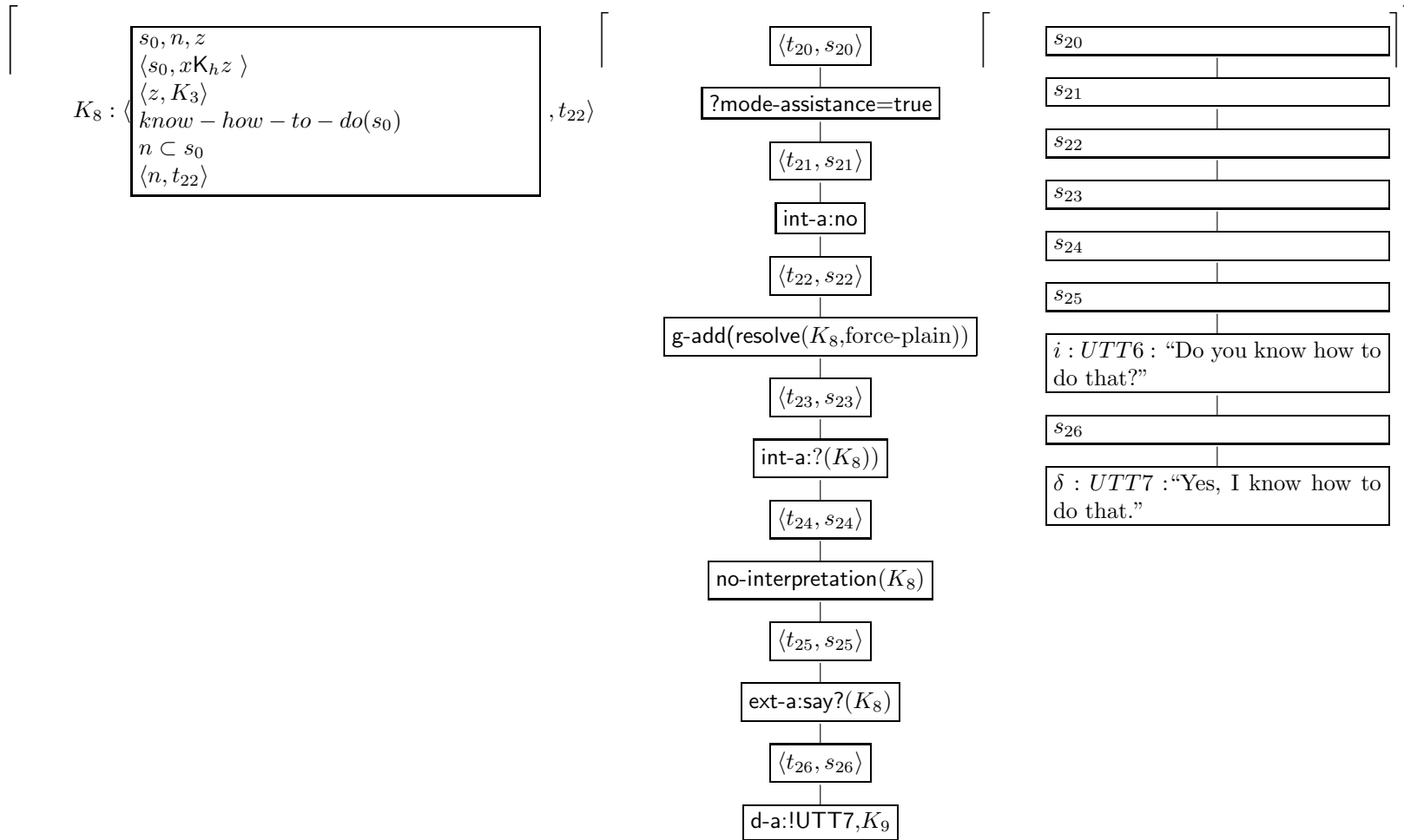


Figure 7.45: Discourse analysis of collaboration mode, Part 4. The mode of interaction is determined. Mode-assistance is not preset by initialize-state and the collaborator has know-how, thus the collaboration branch of the plan for building a ccb is executed.

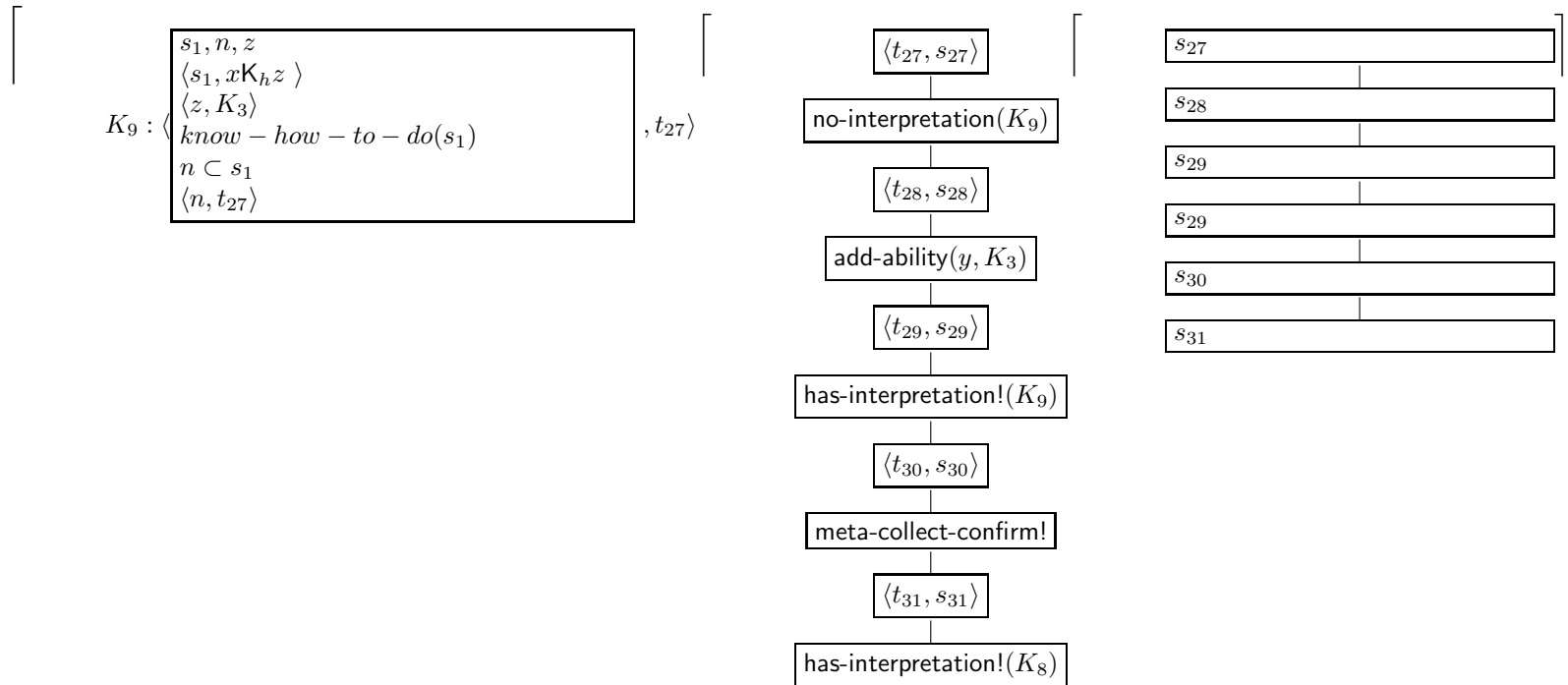


Figure 7.46: Discourse analysis of collaboration mode, Part 5. Resolution of K_9 and K_8 via the reactive interpretation of the answer $UTT7$ of the human in response to $UTT6$.

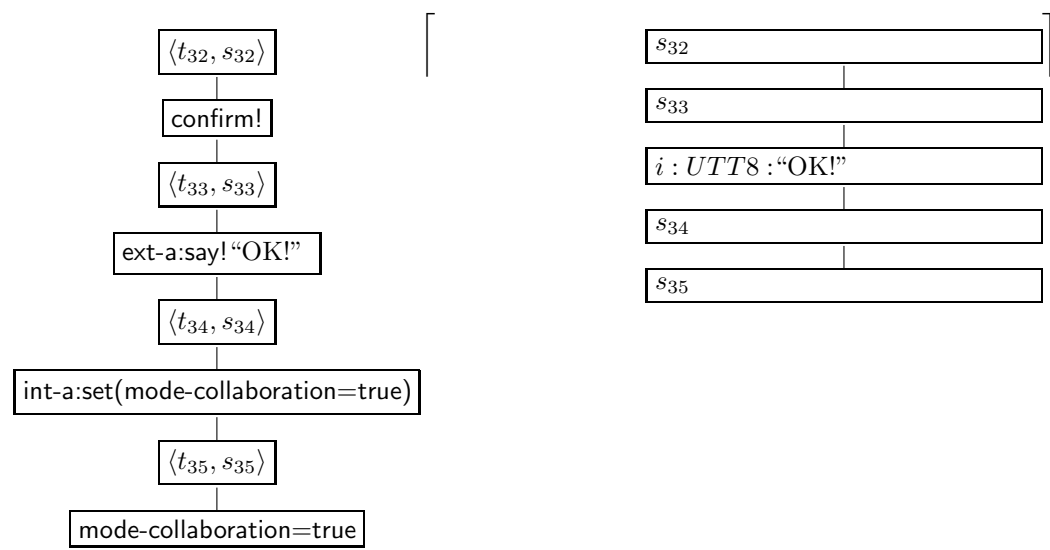


Figure 7.47: Discourse analysis of collaboration mode, Part 6. The collaboration branch of the plan for building is executed.

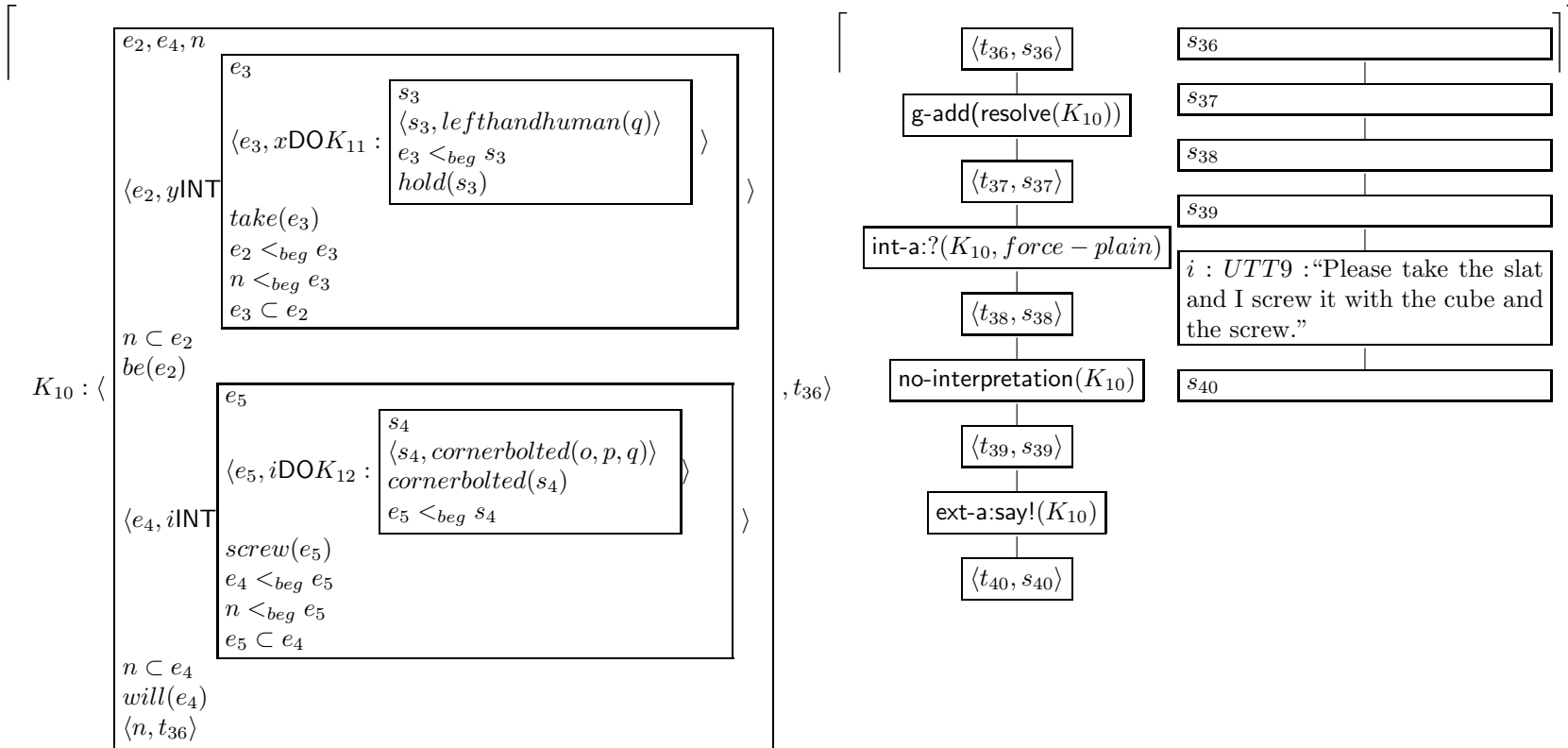


Figure 7.48: Discourse analysis of collaboration mode, Part 7. Distribution of work between the human and Clara. Resolution of the IRS K_{10} .

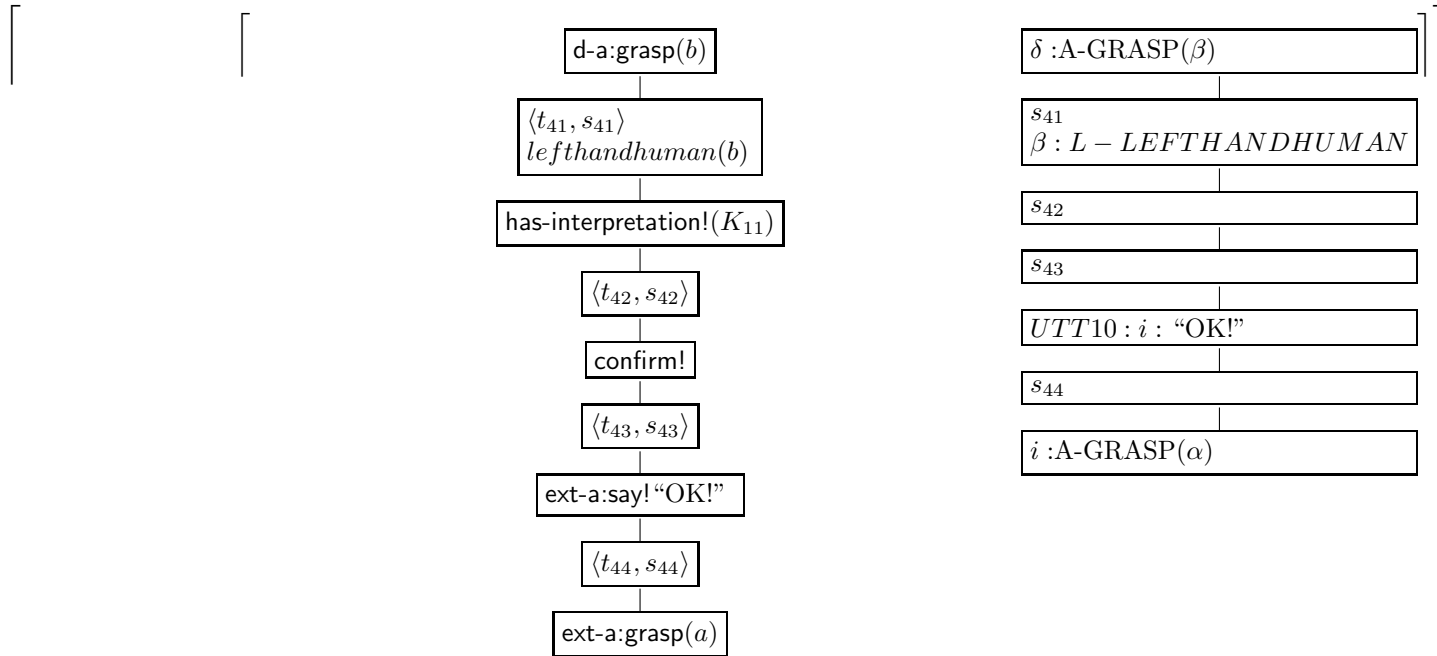


Figure 7.49: Discourse analysis of collaboration mode, Part 8. Confirmation of the partial resolution of K_{10} . Further execution of the plan for building a ccb.

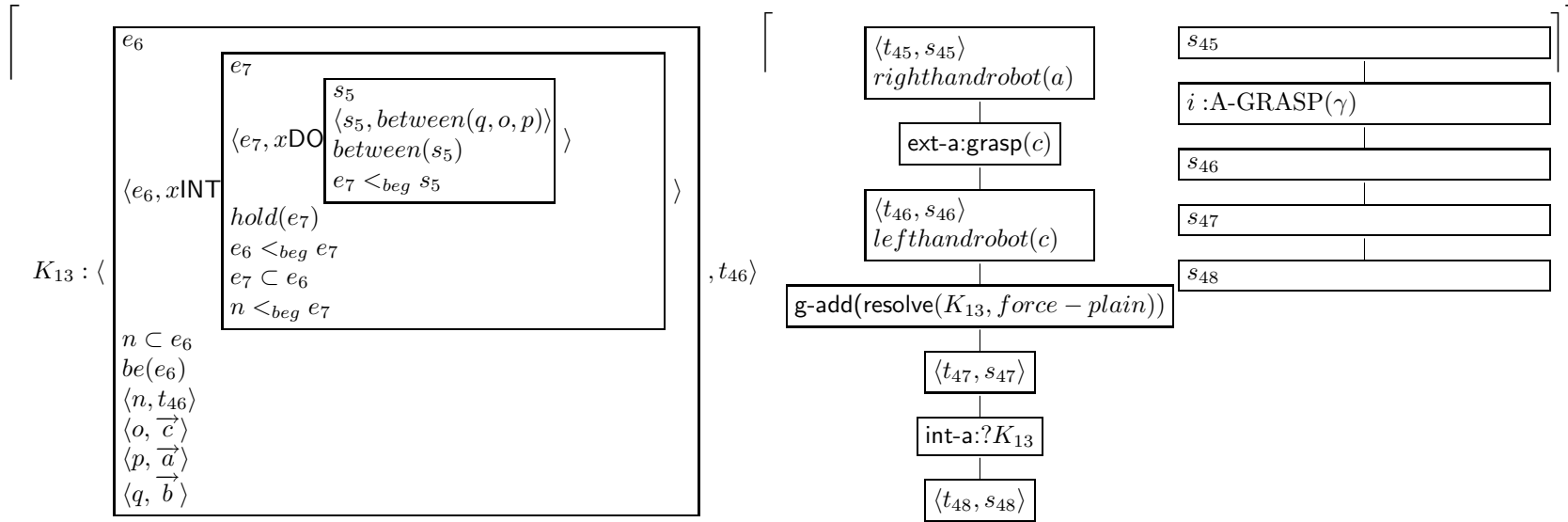
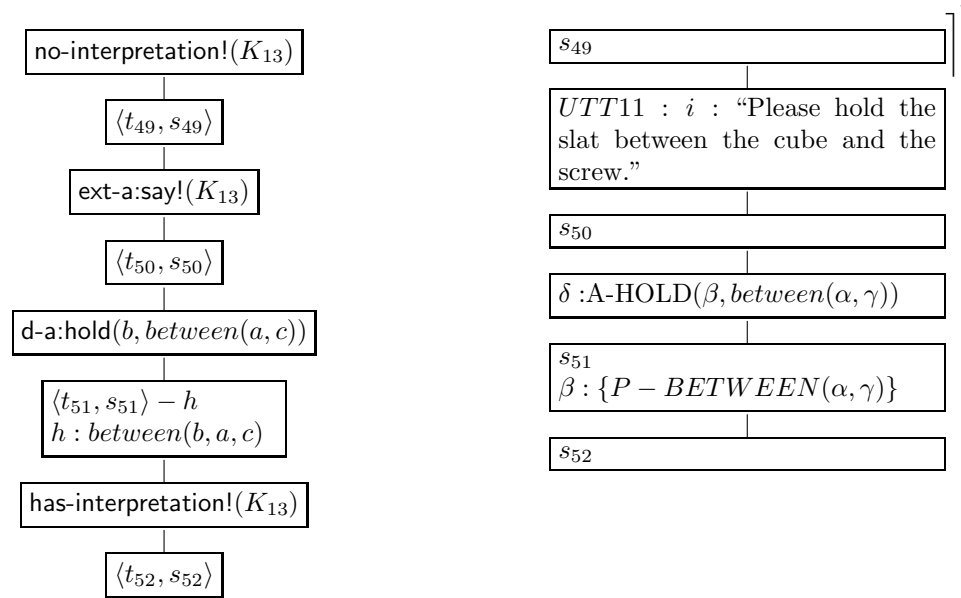


Figure 7.50: Discourse analysis of collaboration mode, Part 9. Resolution of the IRS K_{13} which specifies the preconditions for the plan for constructing a ccb.

Figure 7.51: Discourse analysis of collaboration mode, Part 10. Verbal resolution of K_{13} .

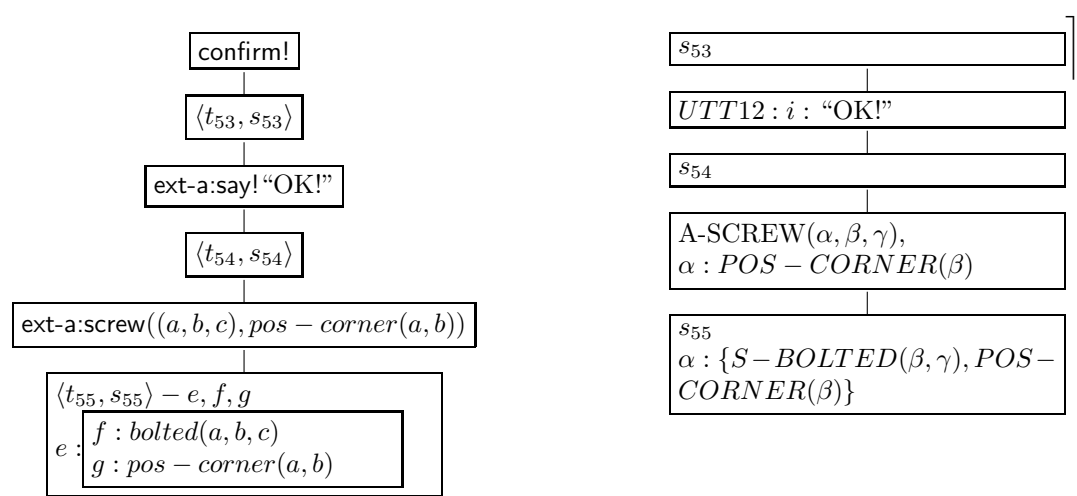


Figure 7.52: Discourse analysis of collaboration mode, Part 11. By constructing the corner cube bolting, K_{13} and K_{11} have an interpretation.

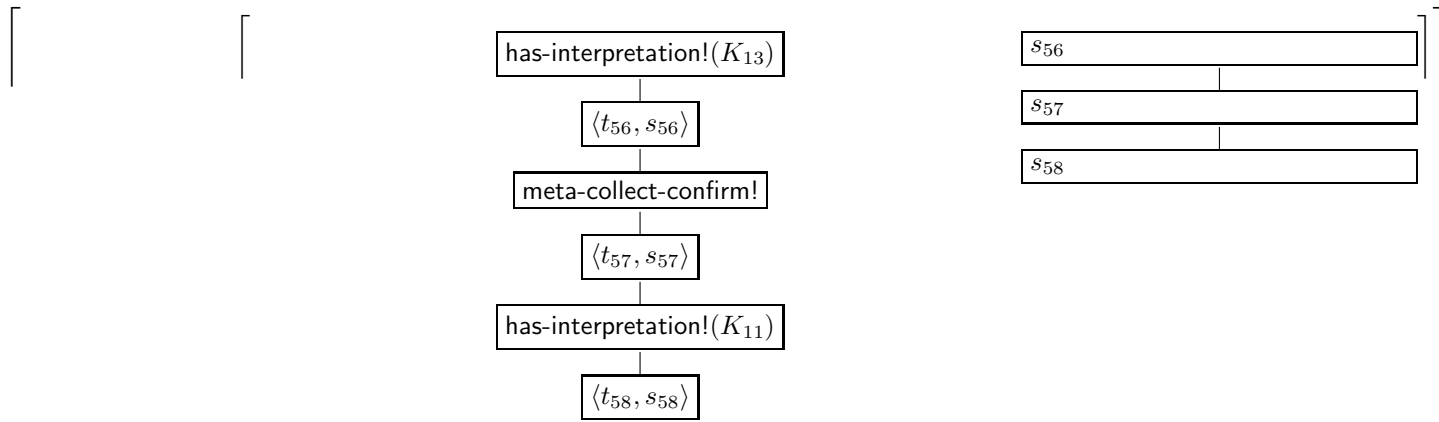


Figure 7.53: Discourse analysis of collaboration mode, Part 12. The confirmation of the successful interpretation of K_{13} and K_{11} is collected by the plan meta-collect-confirm!.

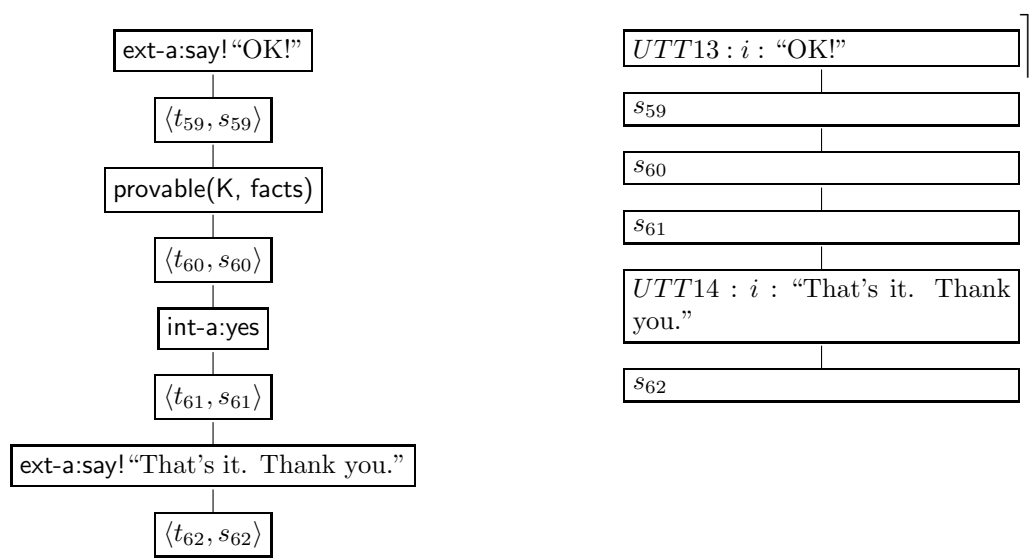


Figure 7.54: Discourse analysis of collaboration mode, Part 13. Finishing sequence.

7.7 Teaching Mode

Teaching mode differs from assistance and collaboration mode in that the human collaborator does not know how to build a corner cube bolting. Consequently, the robot should explain her goals and actions to the user. The analysis of the example interaction 16 executes the collaboration branch of the plan for building a corner cube bolting with the additional goal of teaching. The plan for teaching (figure 7.55) executes an elaboration of the goal of the discourse, an elaboration of the steps that lead to the realization of the goal and a final check whether the teaching was successful (pictured by the plans in figures 7.57,7.58,7.56).

Example 16 *Example dialog for teaching mode.*

	<i>The table holds a cube, a slat and a screw. The user approaches the table and stops in front of the robot. The robot looks at the user. The robot does not know the user's name.</i>
A:	Hi, my name is Clara. What is your name?
B:	My name is Tillmann.
A:	I am going to build a corner cube bolting. Do you know how to do that?
B:	No, I don't know how to do that.
A:	I will explain it to you.
A:	A corner cube bolting is a cube <i>A points to the cube</i>
A:	screwed into the corner hole of a slat <i>A points to the slat.</i>
A:	with a screw. <i>A points to the screw</i> <i>A grasps the screw.</i> <i>A grasps the cube</i>
A:	Please hold the slat between the screw and the cube. <i>B grasps the slat and holds it between the cube and the screw. The robot screws the cube into the slat.</i>
B:	OK.
A:	Do you know how to build a corner cube bolting now?
B:	Yes.
A:	OK. That's it. Thank you.

Type:	elaborate(K)
Invocation:	g-add(elaborate(K))
Context:	-
Feedback:	-
Body:	<div style="text-align: center;"> <pre> graph TD t0[t0] --> update[update(IRS,K)] update --> t1[t1] t1 --> loop["ext-a:say!IRS and for each <math>\langle x, a \rangle \in \text{IRS}</math> loop"] loop --> t2[t2] t2 --> point["ext-a:point(a)"] point --> t3[t3] t3 --> endloop[end loop] </pre> </div>

Figure 7.56: Plan for elaborating K. This plan updates the IRS with the given topic K (if it is an EPS) and verbalizes the updated IRS, where each occurrence of a thing is accompanied by a gesture of pointing to the respective thing.

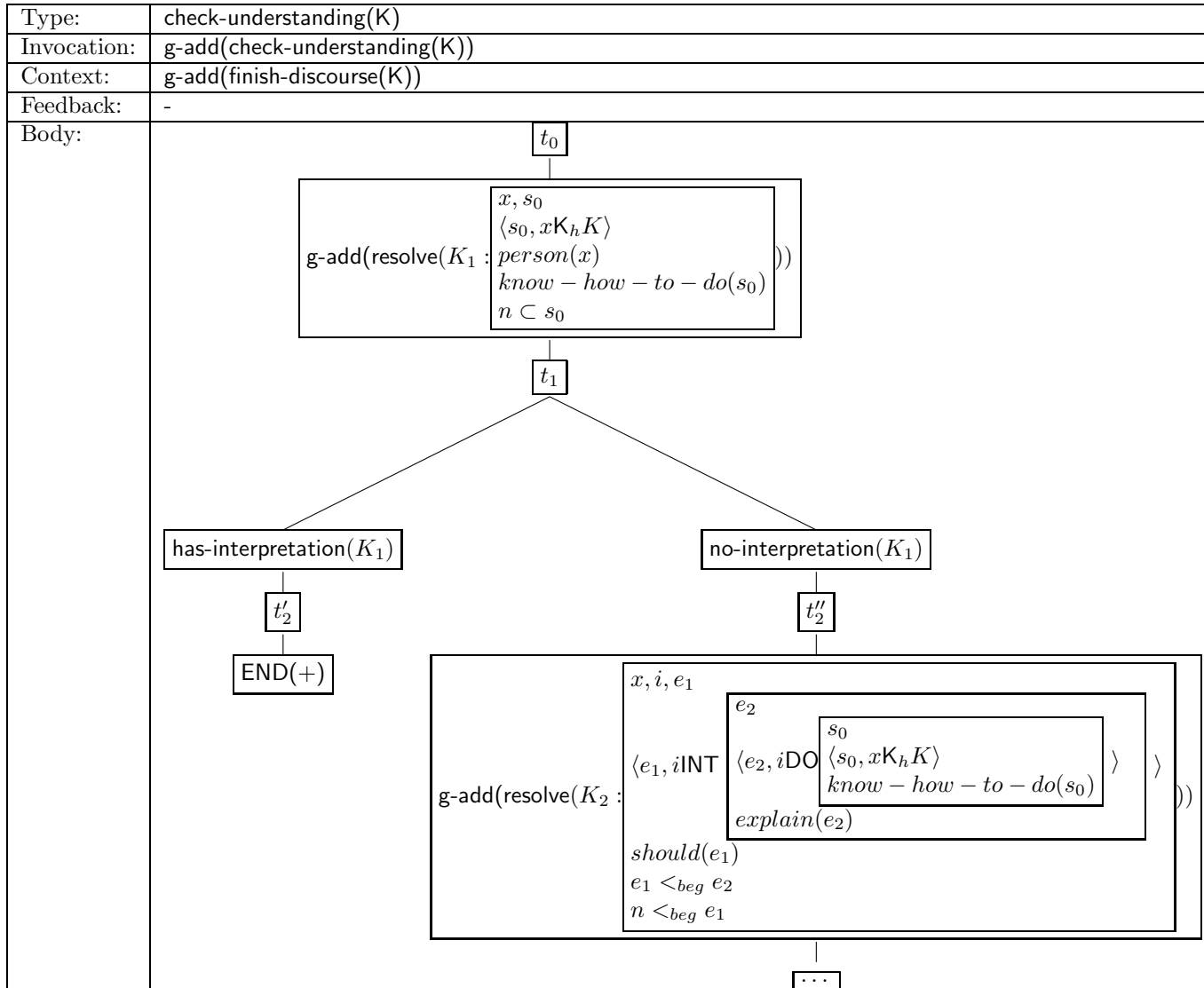


Figure 7.57: Plan for checking the result of teaching mode part 1. If the collaborator has grasped the taught plan, no further action is required. Otherwise it is checked whether the collaborator wants to have another run of explanation.

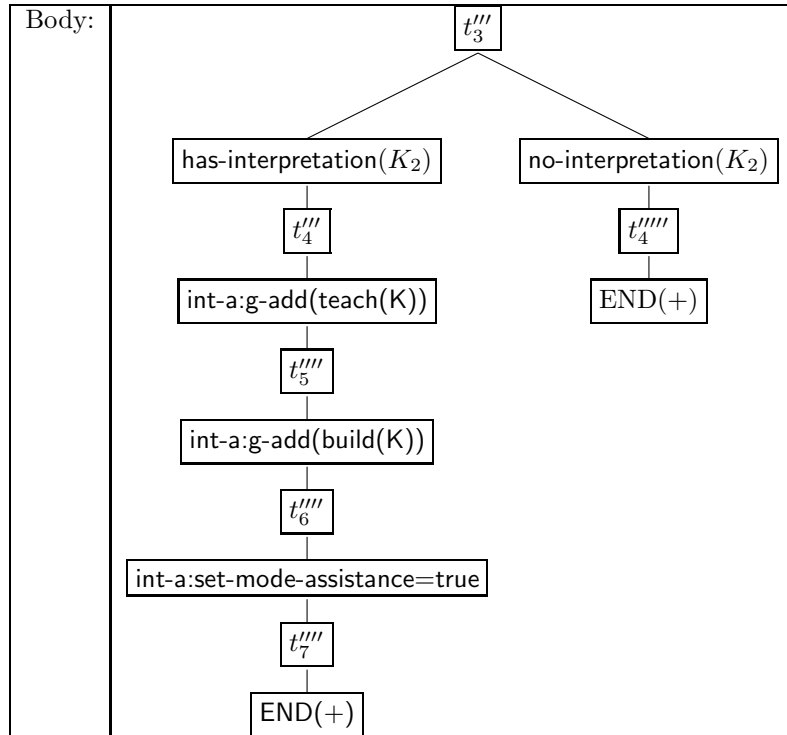


Figure 7.58: Plan for checking the result of teaching mode part 2.

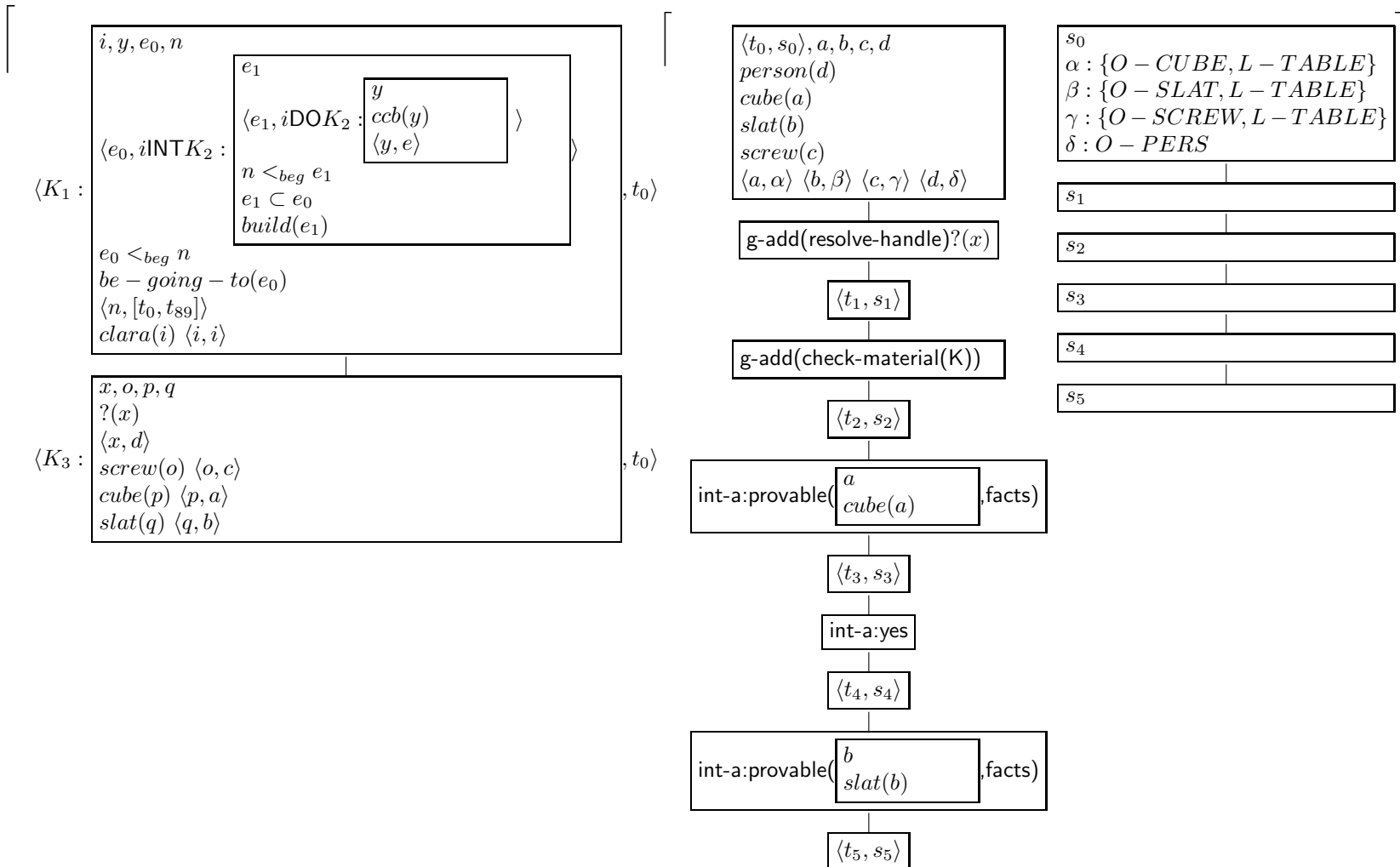
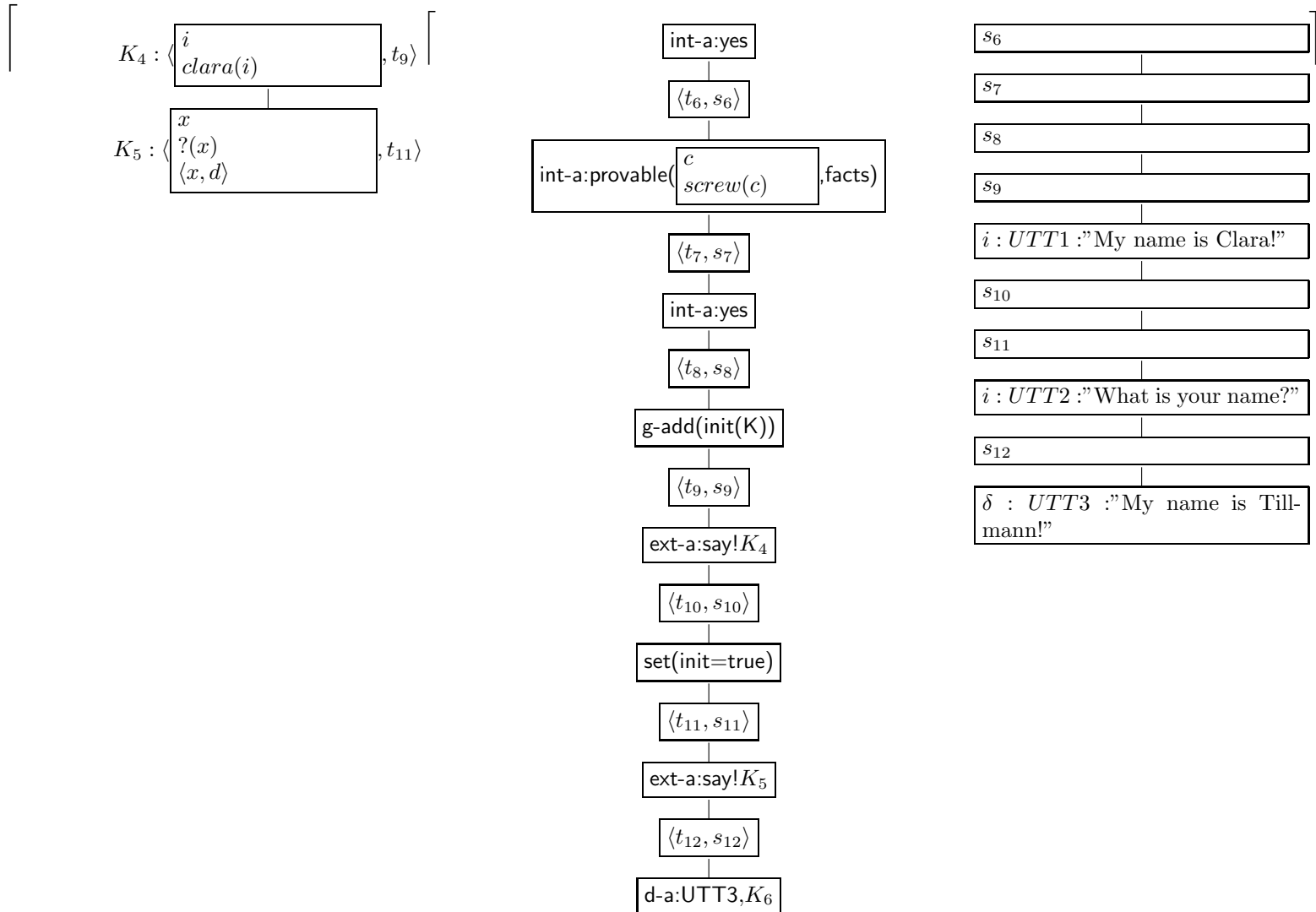


Figure 7.59: Discourse analysis of teaching mode, Part 1. Checking the availability of the materials necessary for the construction of a *ccb*.

Figure 7.60: Discourse analysis of teaching mode, Part 2. Resolution of the missing handle $?(x)$.

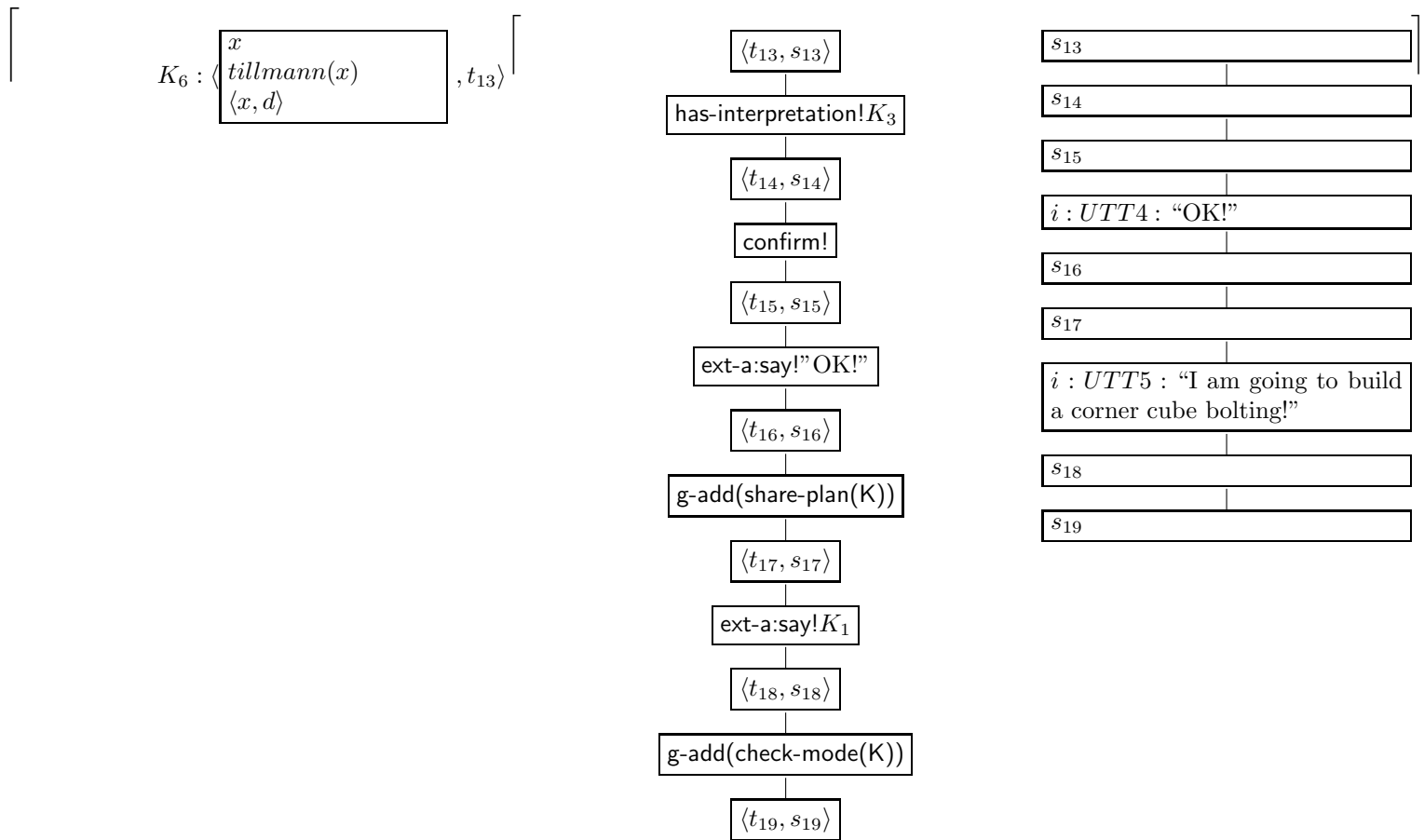


Figure 7.61: Discourse analysis of teaching mode, Part 3. Confirmation of the resolution of the missing name. Sharing of the main discourse plan and goal. Determination of the mode of interaction.

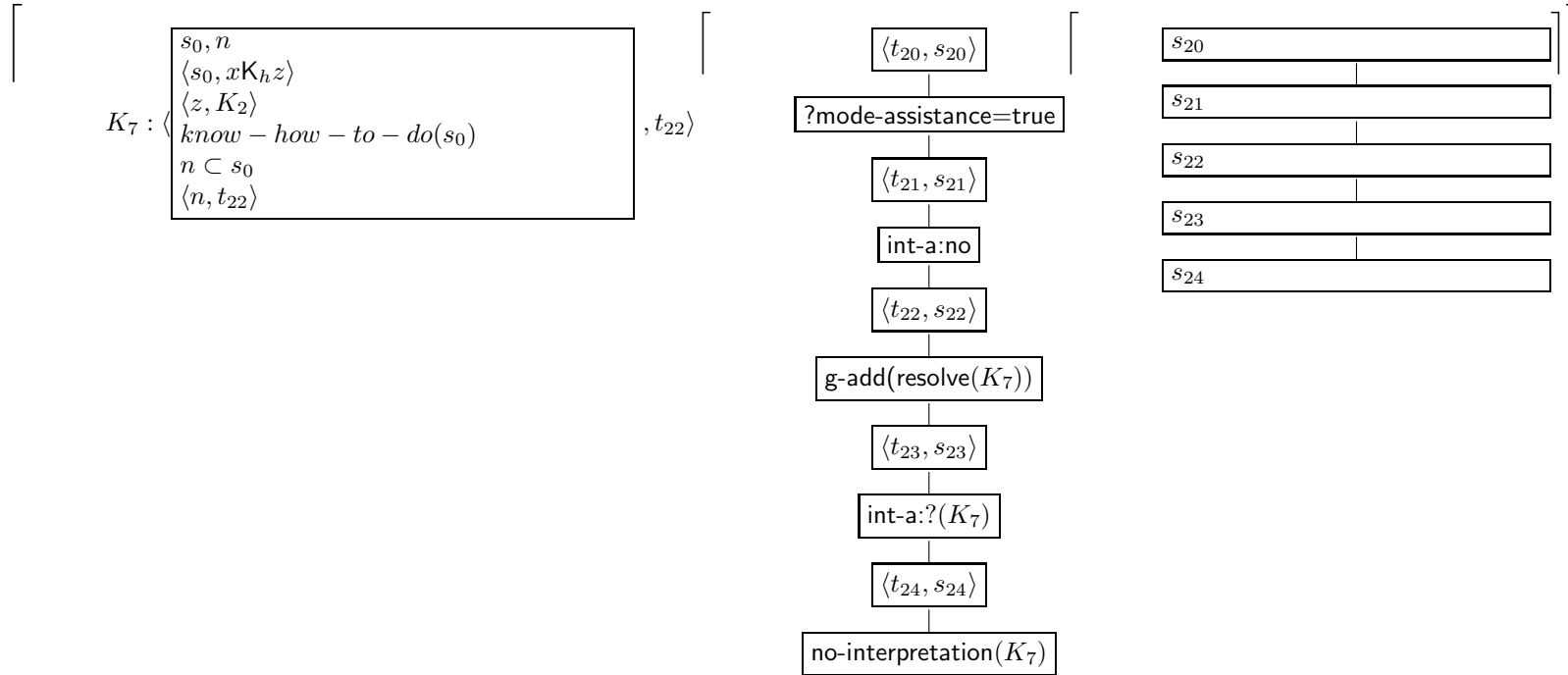


Figure 7.62: Discourse analysis of teaching mode, Part 4. As mode-assistance is not set by the initialization-procedure, the collaboration branch of the plan for building is executed.

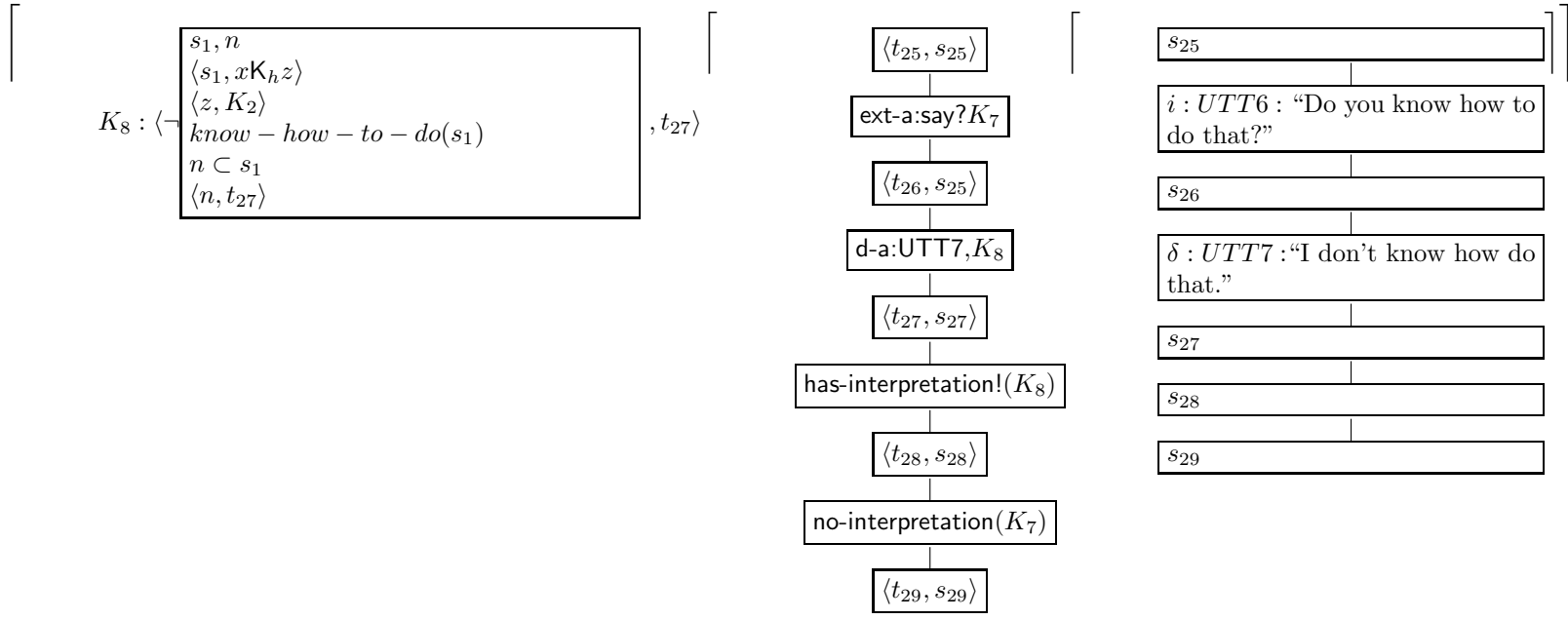


Figure 7.63: Discourse analysis of teaching mode, Part 5. Checking the know-how of the human user. The user has no know-how.

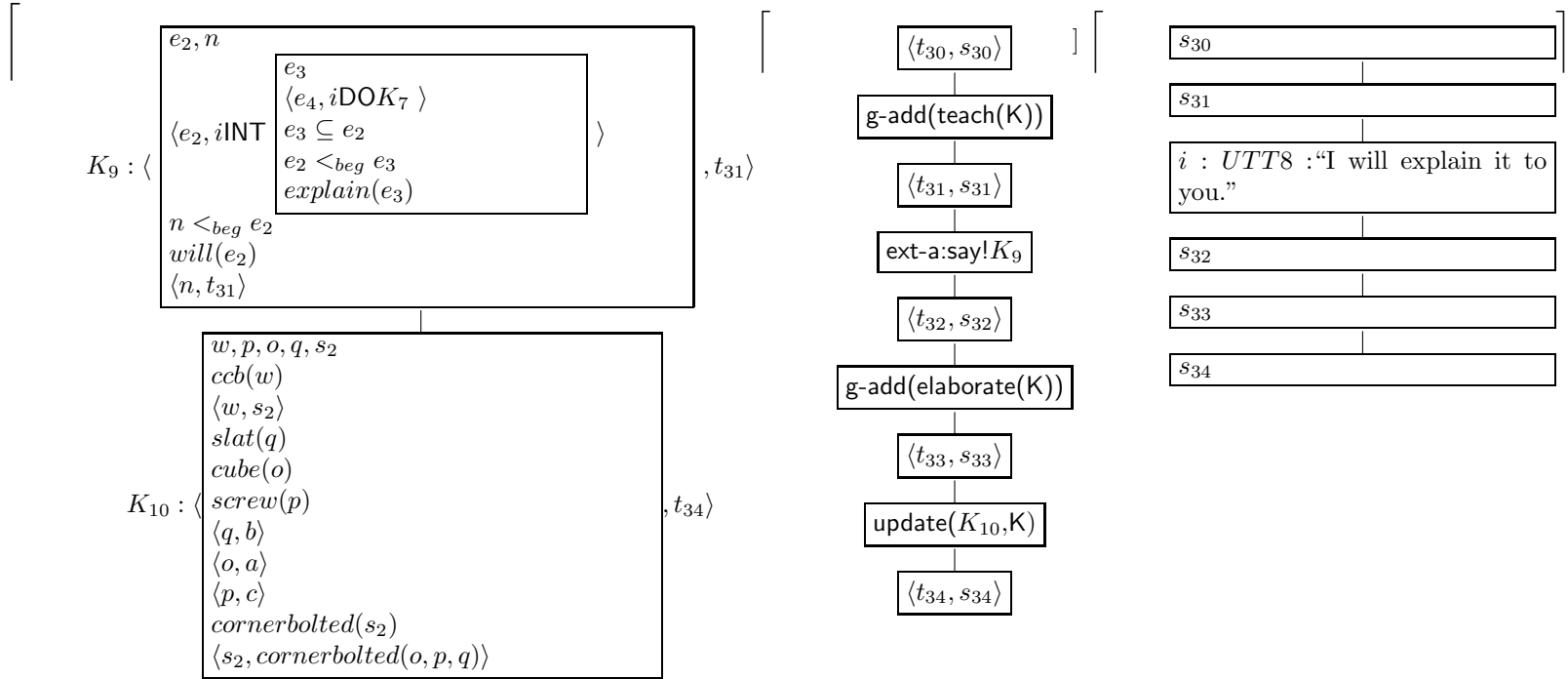


Figure 7.64: Discourse analysis of teaching mode, Part 6. As the user has no know-how, the plan for teaching is activated.

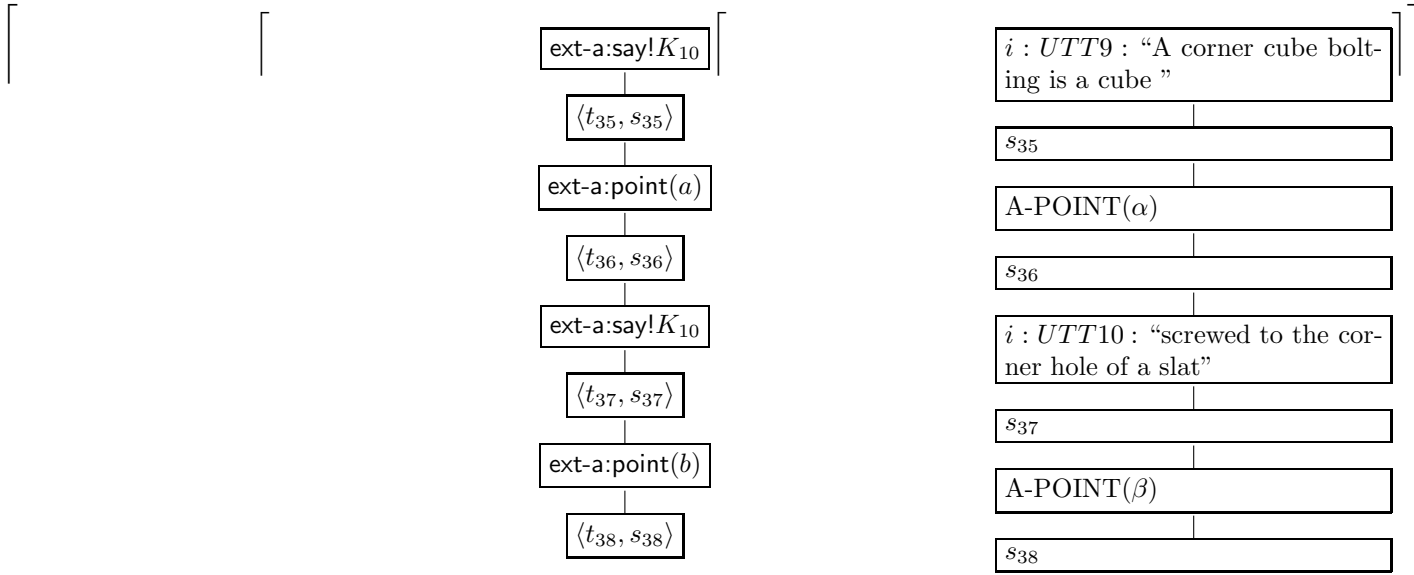


Figure 7.65: Discourse analysis of teaching mode, Part 7. Elaboration of the object corner cube bolting according to the plan elaborate(K).

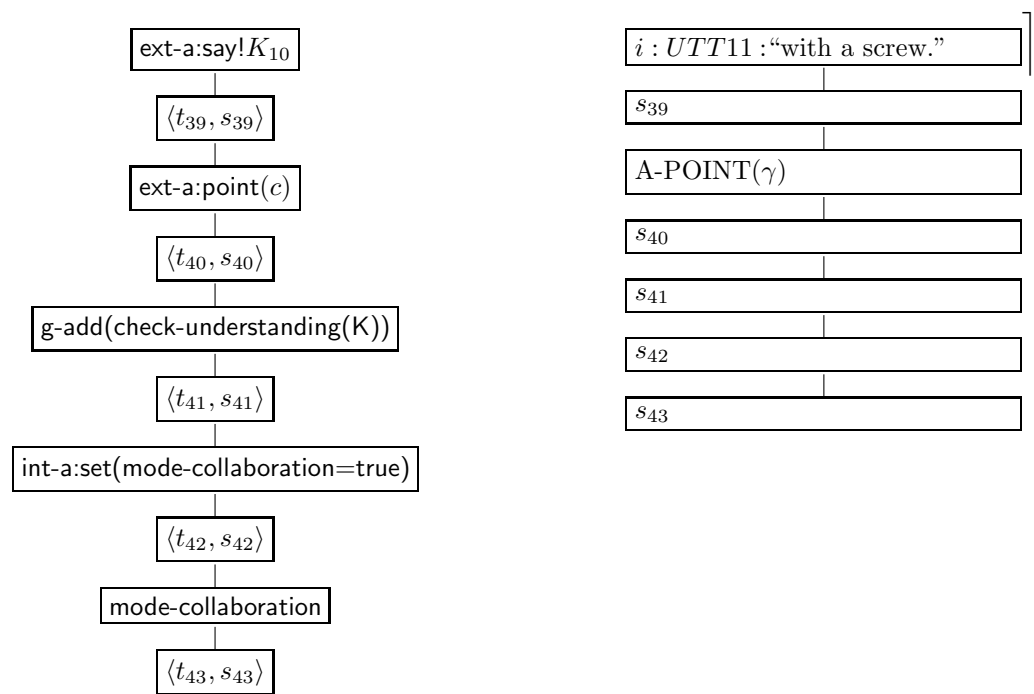


Figure 7.66: Discourse analysis of teaching mode, Part 8. Entering the collaboration branch of the plan for building a ccb.

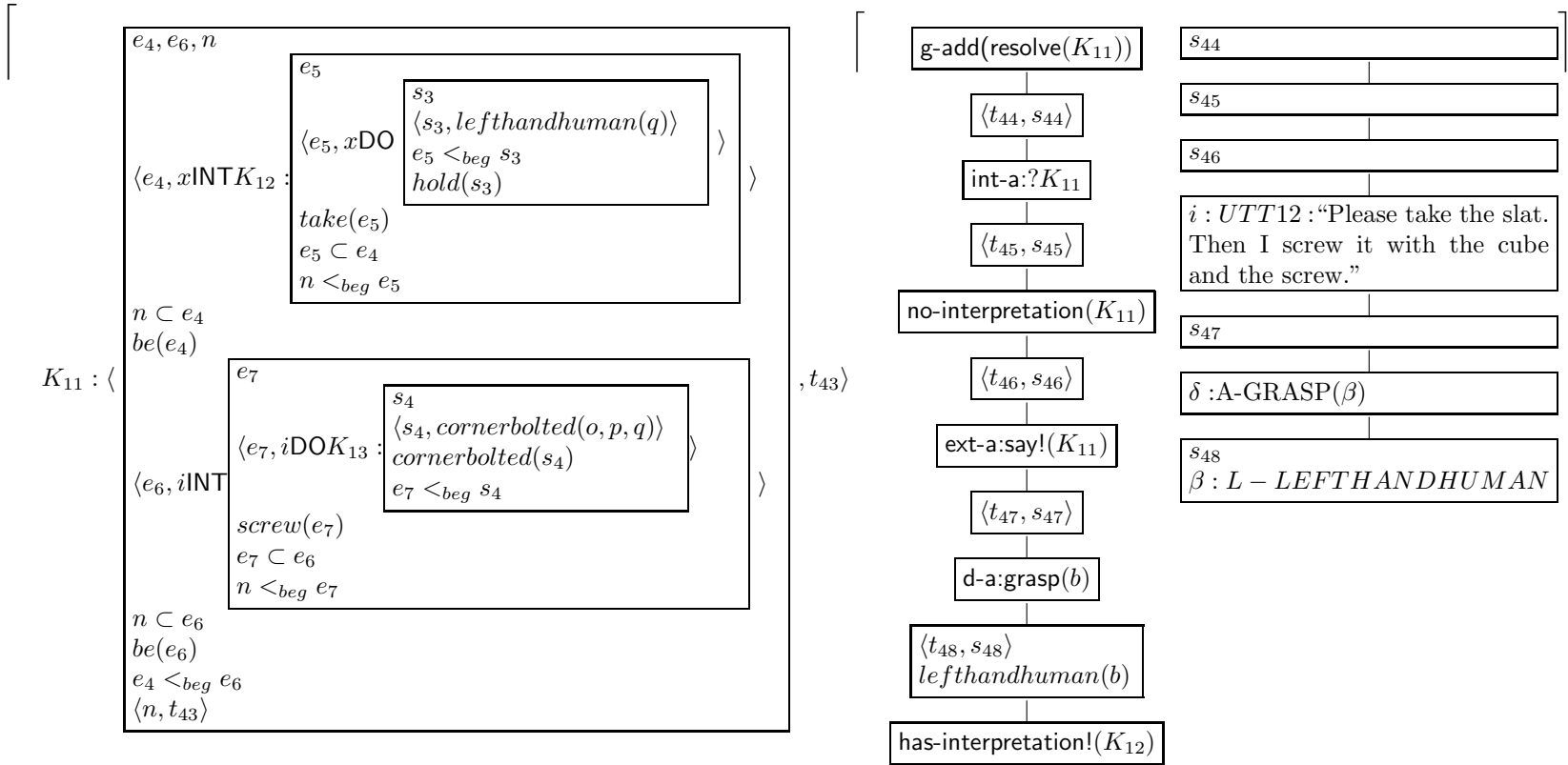


Figure 7.67: Discourse analysis of teaching mode, Part 9. The work to be done is distributed between human and robot.

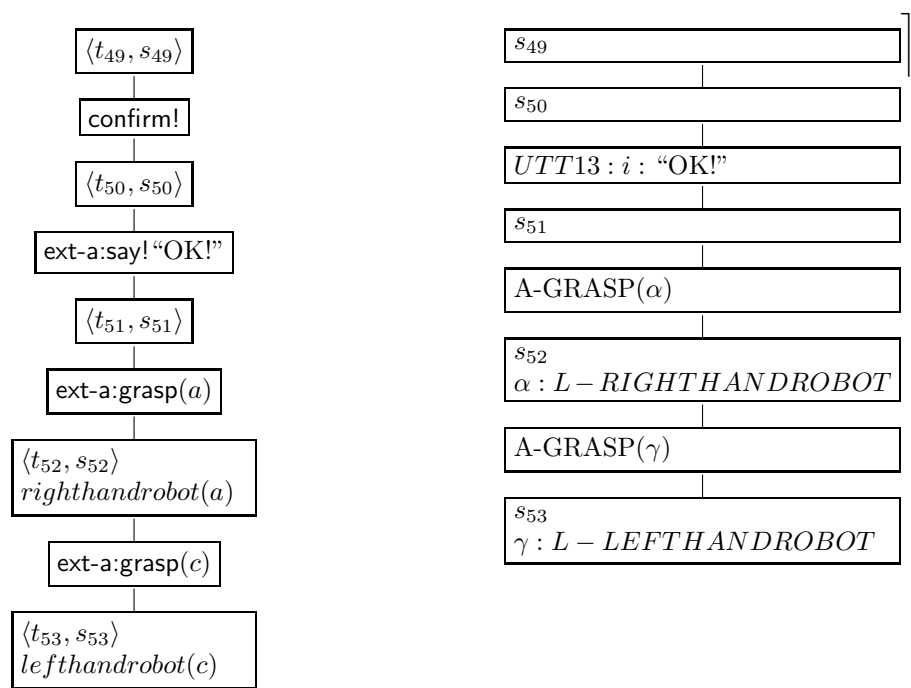


Figure 7.68: Discourse analysis of teaching mode, Part 10.

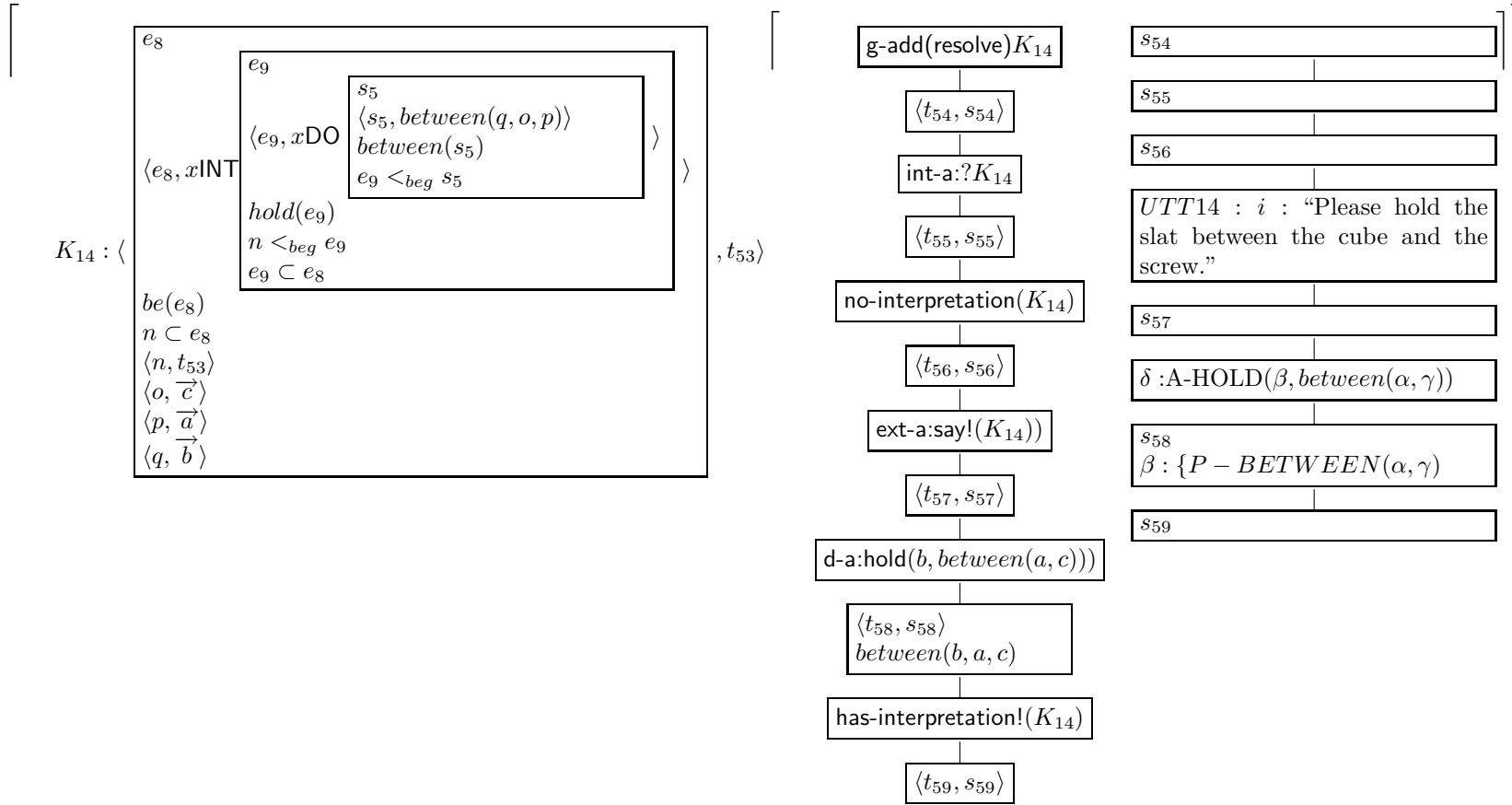


Figure 7.69: Discourse analysis of teaching mode, Part 11. Resolution of K_{14} as a prerequisite to the construction of the corner cube bolting.

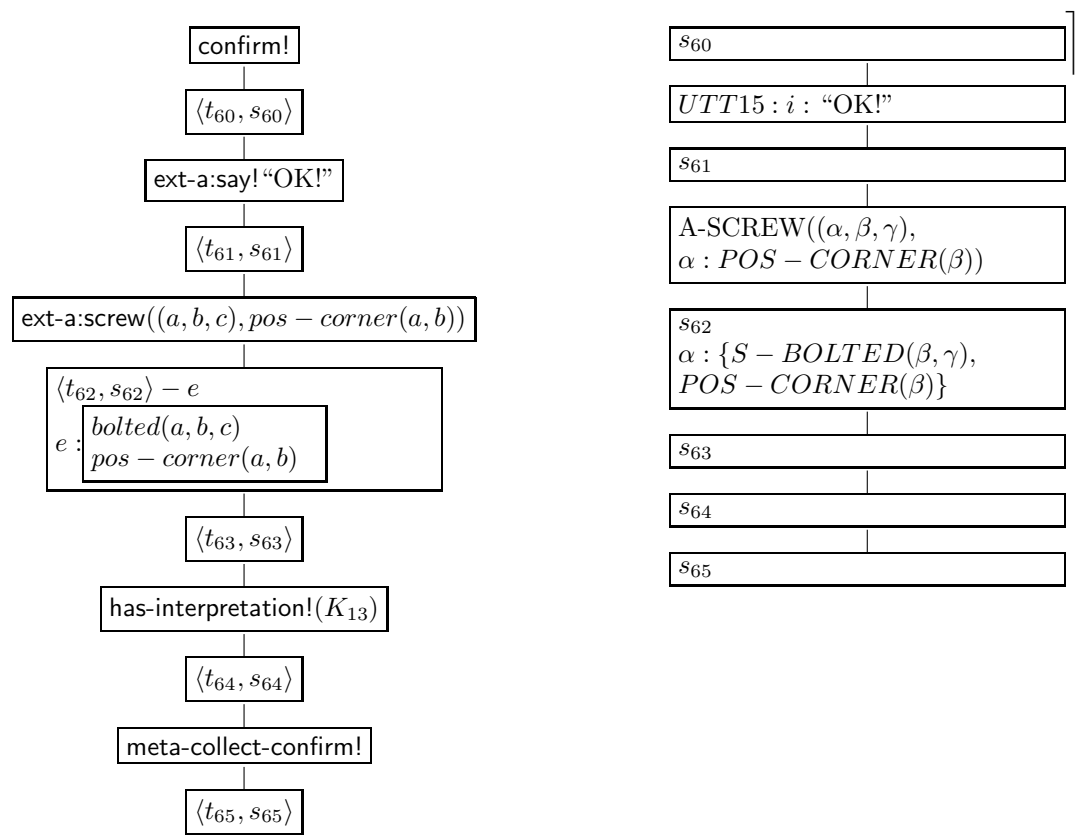


Figure 7.70: Discourse analysis of teaching mode, Part 12. Construction of the corner cube bolting. The resulting confirmations of successful interpretations of pending IRSs are collected.

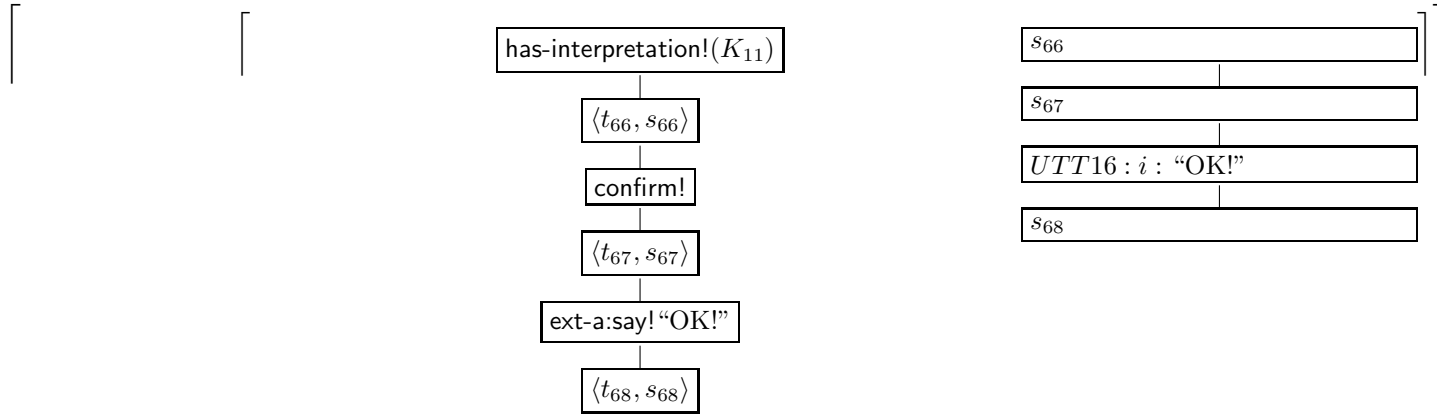


Figure 7.71: Discourse analysis of teaching mode, Part 13. Confirmation of the successful interpretation of pending IRSs.

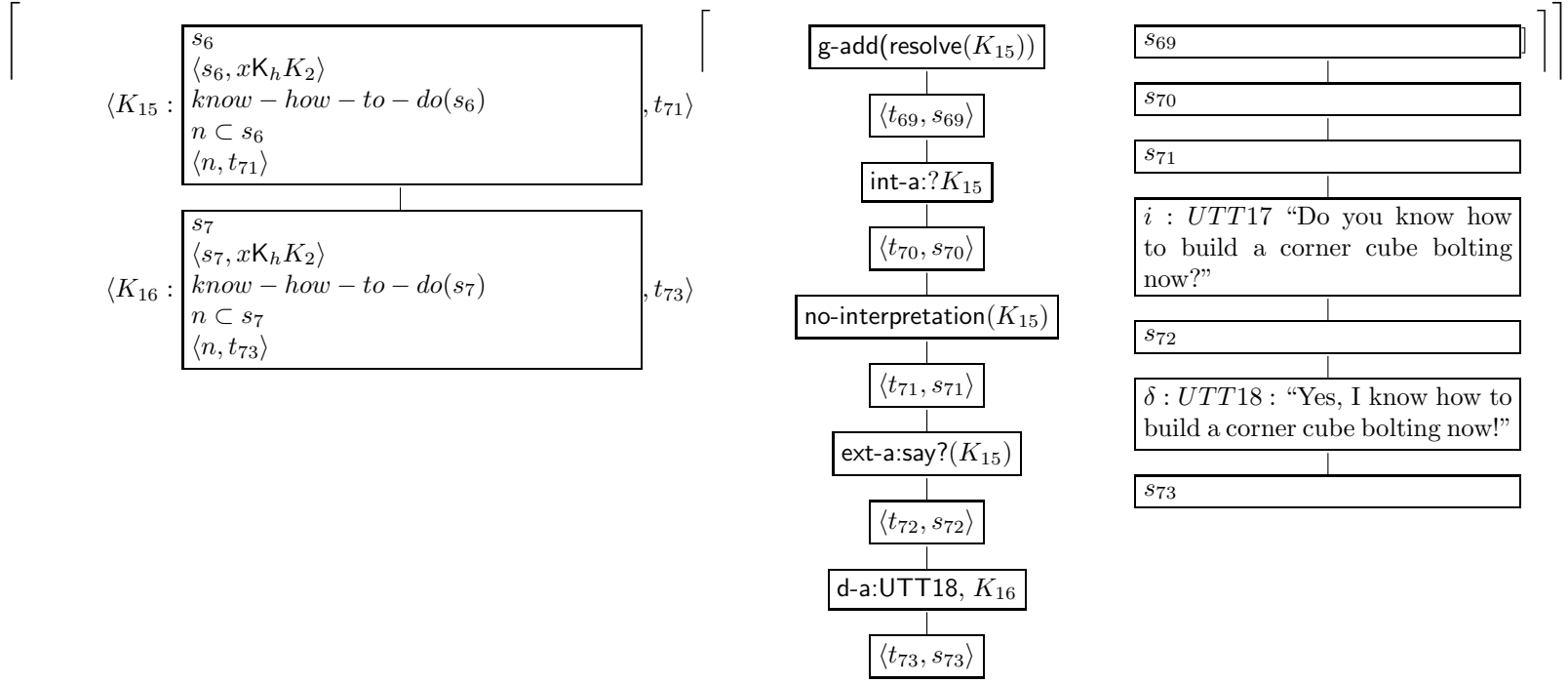


Figure 7.72: Discourse analysis of teaching mode, Part 14. Checking the result of teaching.

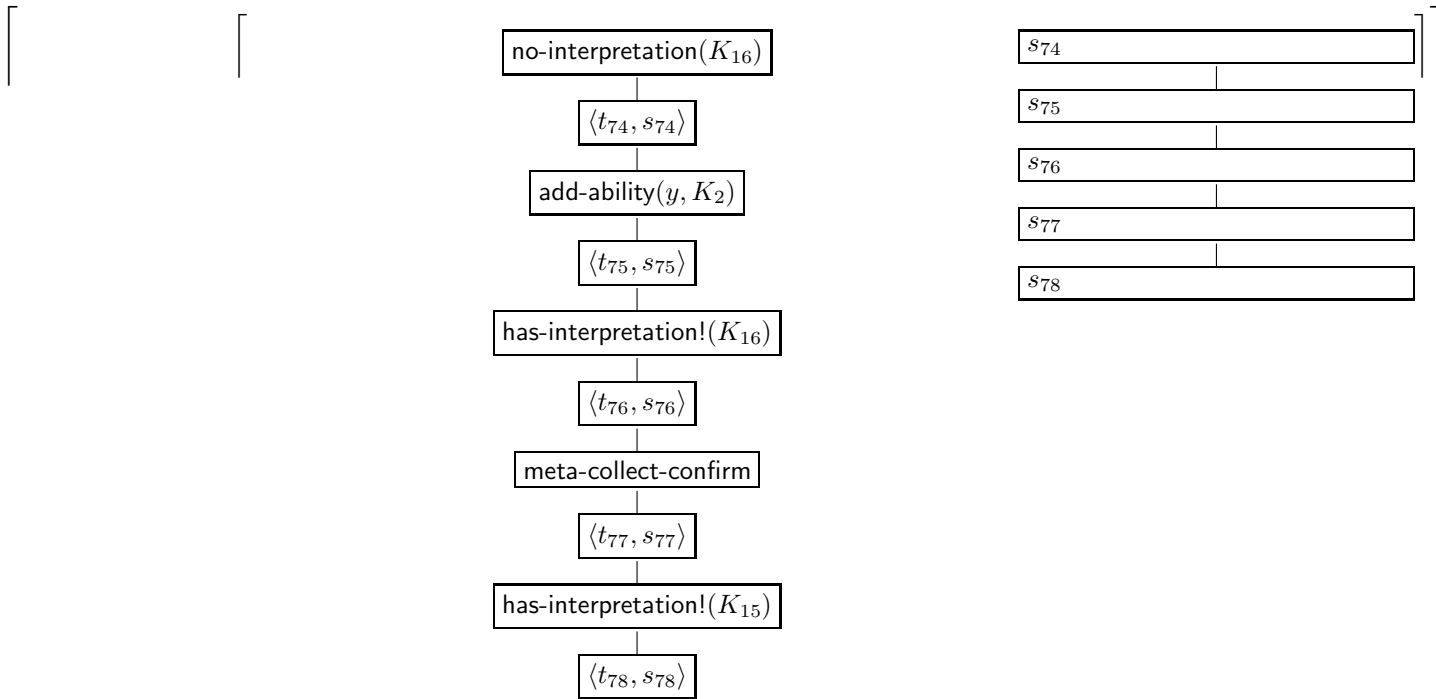


Figure 7.73: Discourse analysis of teaching mode, Part 15. The successful execution of teaching is confirmed.

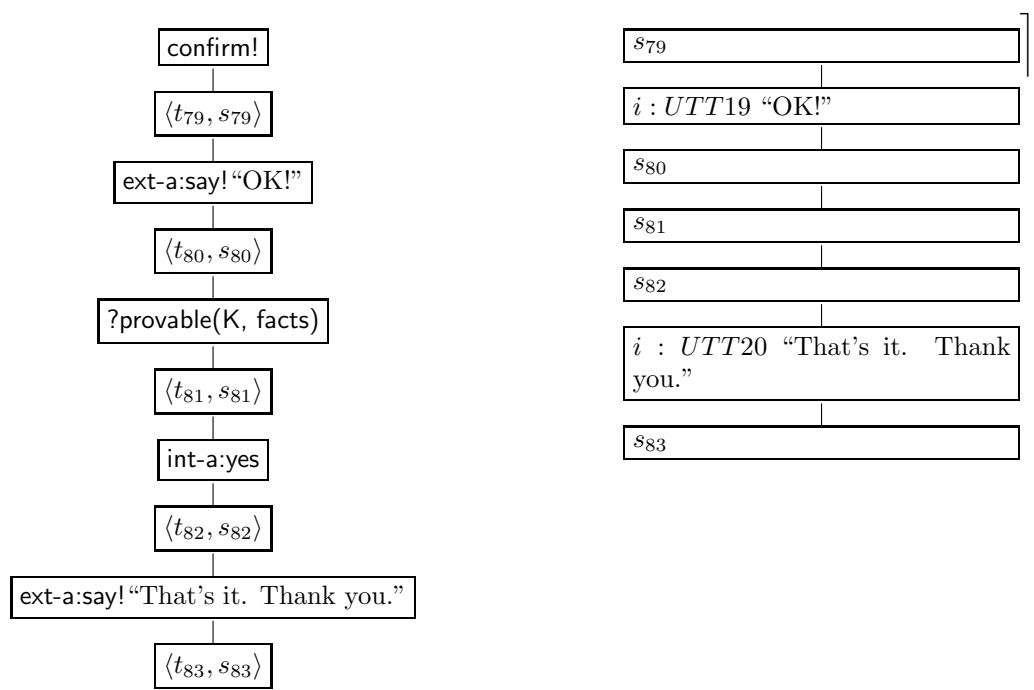


Figure 7.74: Discourse analysis of teaching mode, Part 16. Finishing sequence.

7.8 Free interaction

The final example illustrates the processing of free interaction. In the previous examples, it was the robot who controlled the interaction by executing the plan for building a corner cube bolting. In this example, the human initiates the interaction, whereas the robot must contribute to the discourse via the interpretation of the user's utterances.

Example 17 *Example dialogue for free interaction.*

	<i>The table holds a cube, a blue and a red slat and a screw. The user approaches the table and stops in front of the robot. The robot opens her eyes and looks at the user. Tillmann does not know the robot's name.</i>
A:	Hi, my name is Tillmann. What is your name?
B:	OK.
A:	What is your name?
B:	My name is Clara.
A:	I am going to build a corner cube bolting.
B:	OK.
A:	Do you know how to do that?
B:	Yes.
A:	Please take the slat and I screw it with the cube and the screw.
B:	Which slat?
A:	This one. <i>A points to the red slat.</i>
B:	OK. <i>B grasps the red slat.</i>
B:	OK. <i>A grasps the cube and the screw.</i>
A:	Please hold the slat between the cube and the screw.
B:	OK. <i>B holds the slat between the cube and the screw.</i>
	<i>A screws the cube into the red slat.</i>
B:	OK.
A:	That's it. Thank you.

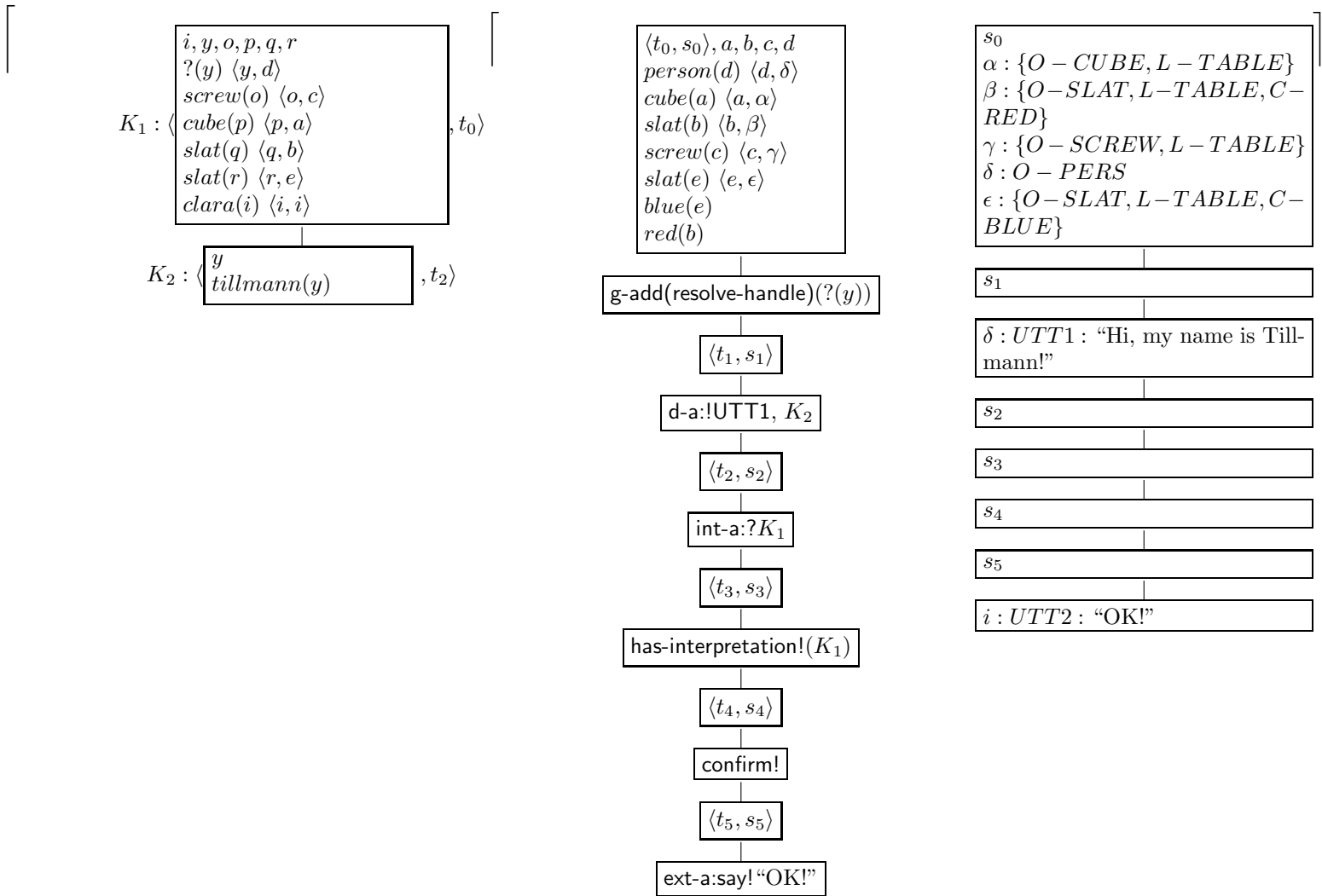


Figure 7.75: Discourse analysis of free interaction, Part 1. $?(y)$ is resolved by utterance $UTT1$ before further steps can be undertaken.

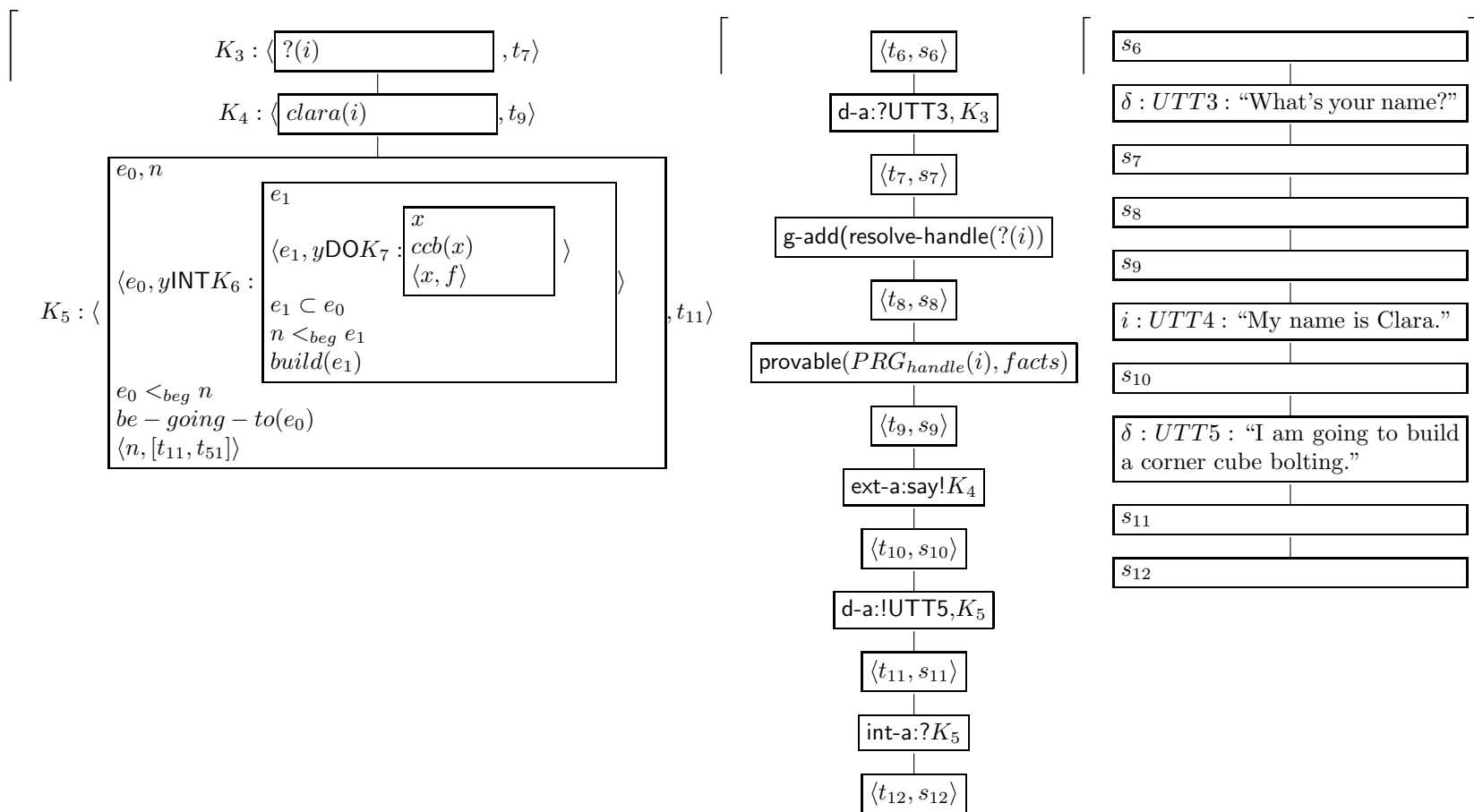


Figure 7.76: Discourse analysis of free interaction, Part 2. Resolution of $?(i)$ in $UTT3$. Tillmann shares his intention represented with K_5 .

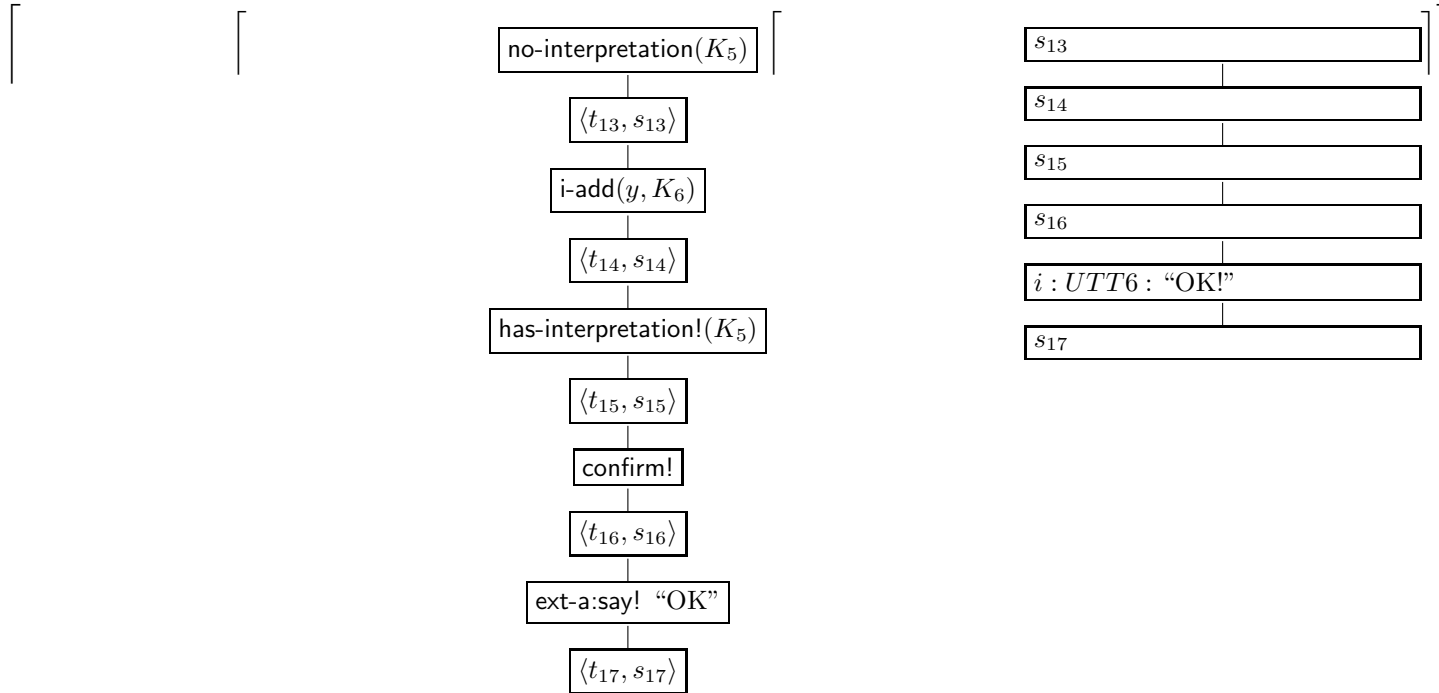


Figure 7.77: Discourse analysis of free interaction, Part 3. Interpretation of IRS K_5 as constructed from $UTT5$. Reactive interpretation triggers the addition of the intention K_6 to Tillmann’s instance of the BDI-interpreter. Consequently, K_5 has an interpretation.

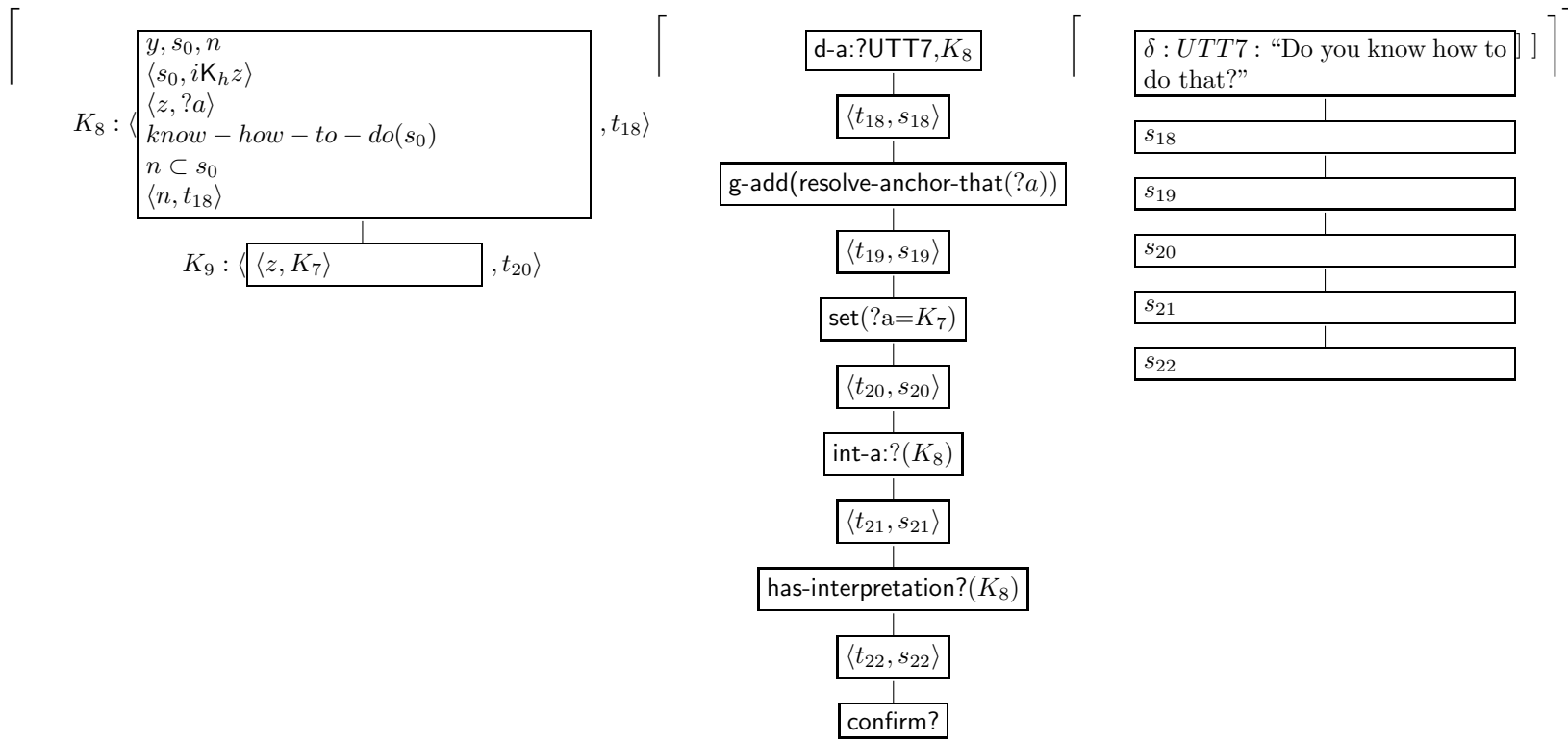


Figure 7.78: Discourse analysis of free interaction, Part 4. Construction of IRS K_8 from $UTT7$. Resolution of the anchor source for 'that' to K_5 . Interpretation of K_8 with forced plain interpretation.

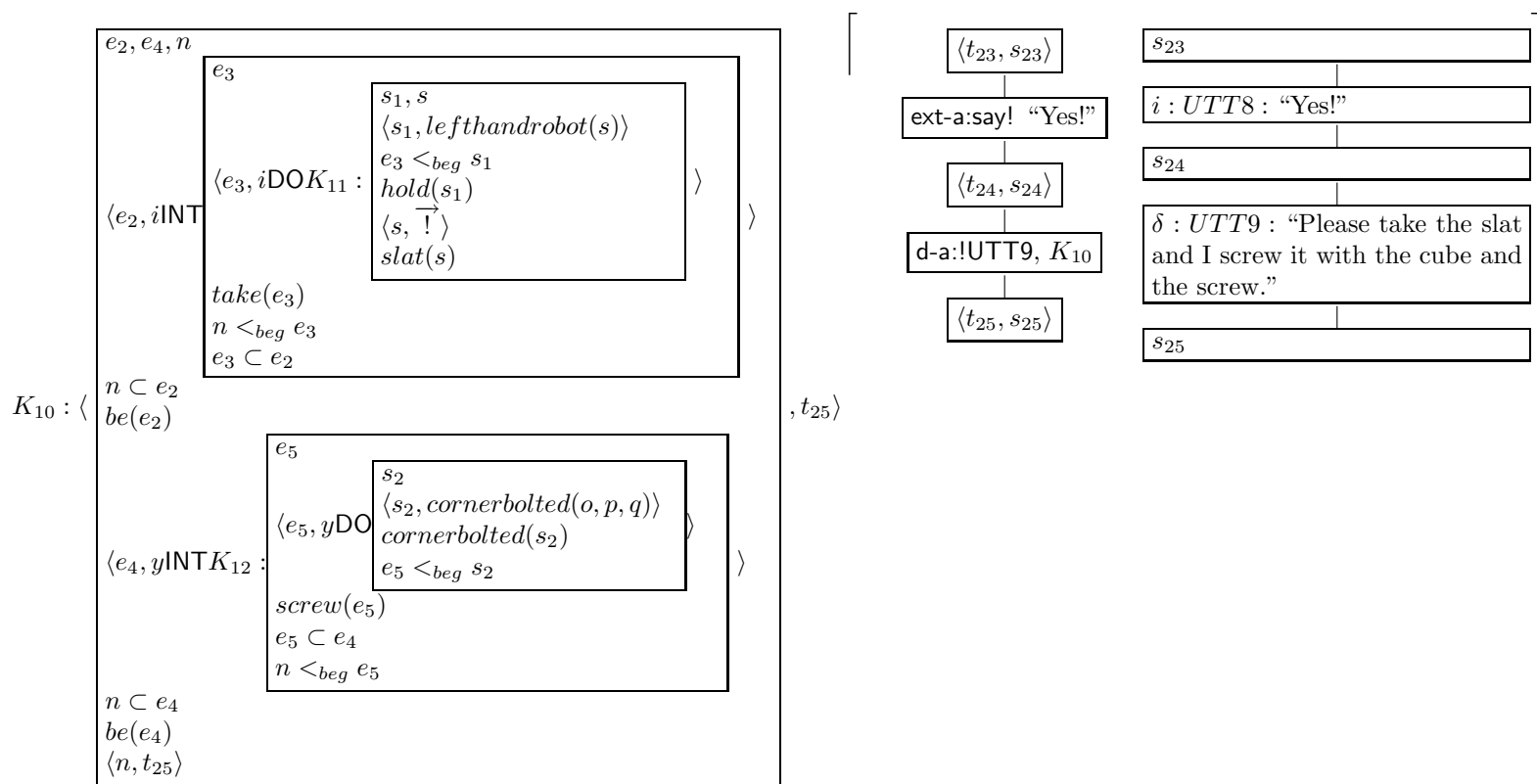


Figure 7.79: Discourse analysis of free interaction, Part 5. Successful plain interpretation of K_8 constructed from $UTT5$ and consequent confirmation. Interpretation of K_9 constructed from $UTT9$, where anchor sources for the screw and the cube can be uniquely determined, but not for the slat.

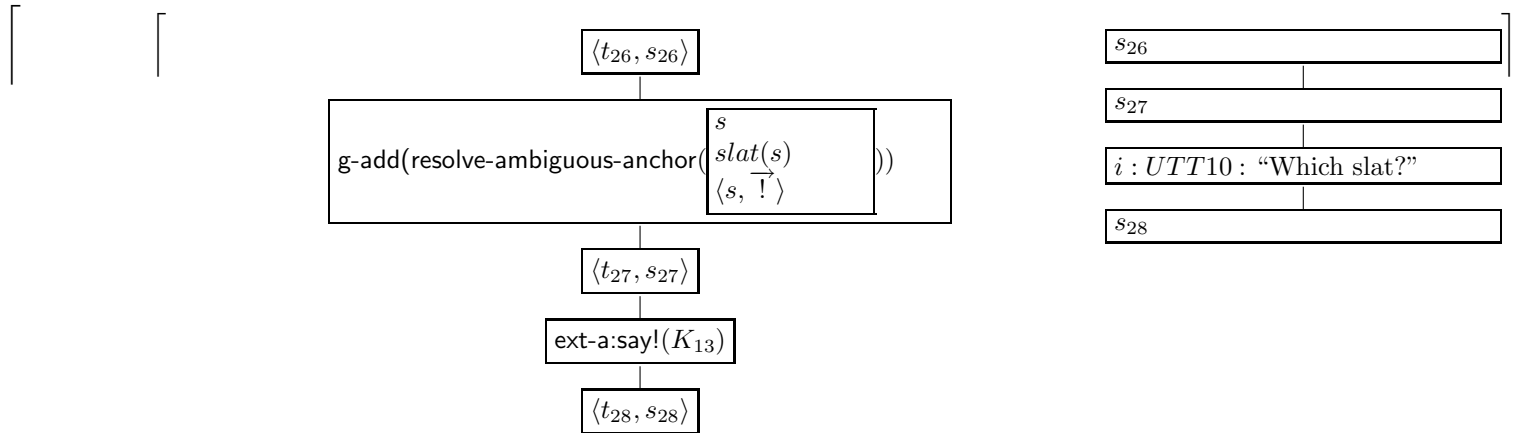


Figure 7.80: Discourse analysis of free interaction, Part 6. Resolution of the unique anchor source for the slat via an invocation of the plan `resolve-ambiguous-anchor` to ask for the right anchor source.

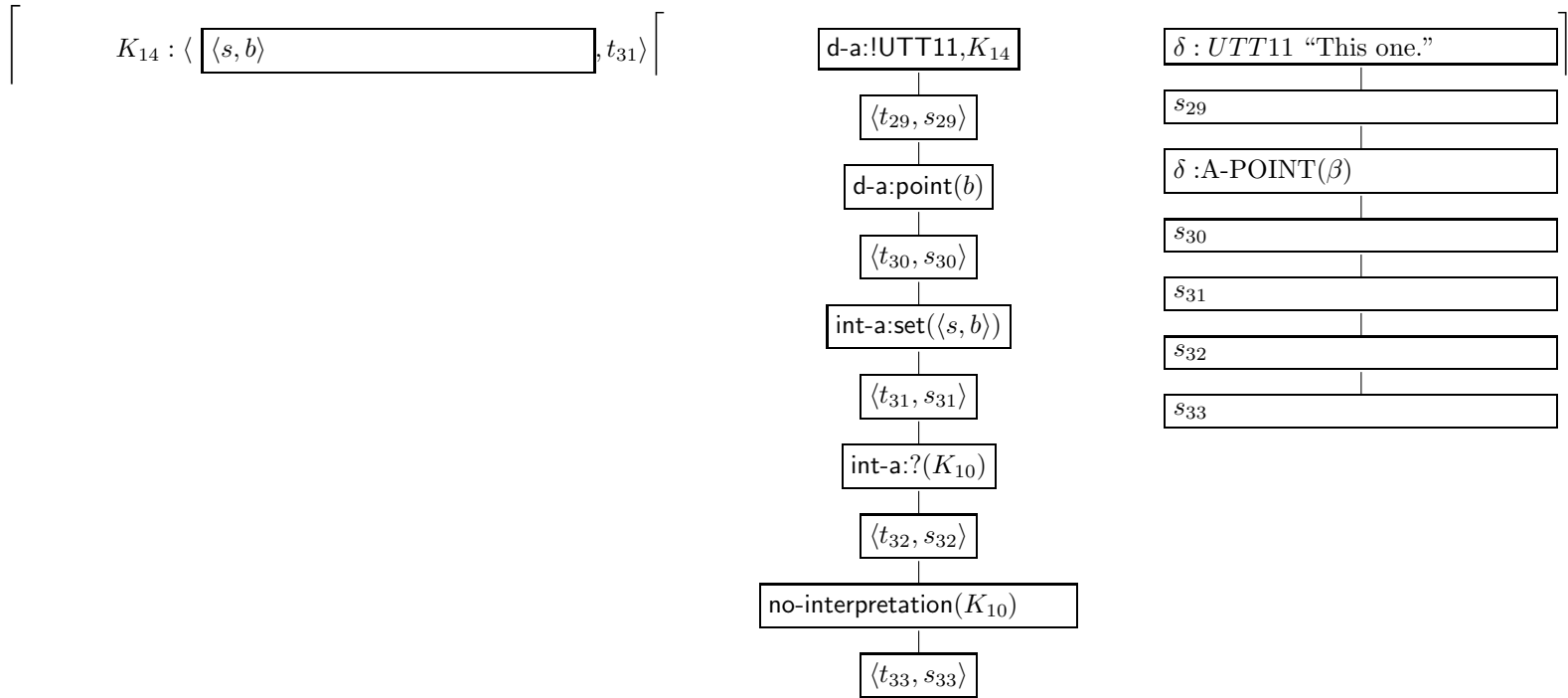


Figure 7.81: Discourse analysis of free interaction, Part 7. Resolution of the anchor source of the slat via the gesture of Tillmann. Interpretation of K_{10} with resolved anchor sources.

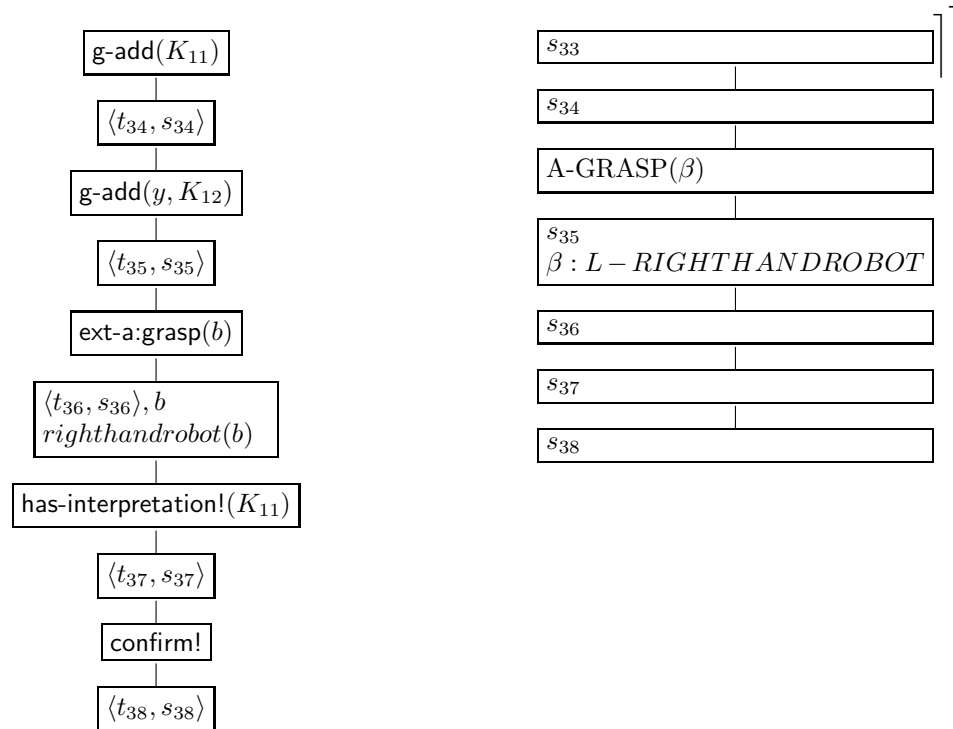


Figure 7.82: Discourse analysis of free interaction, Part 8. Execution of the plan invoked by the interpretation of the Tillmann's goal K_{11} . Consequent confirmation.

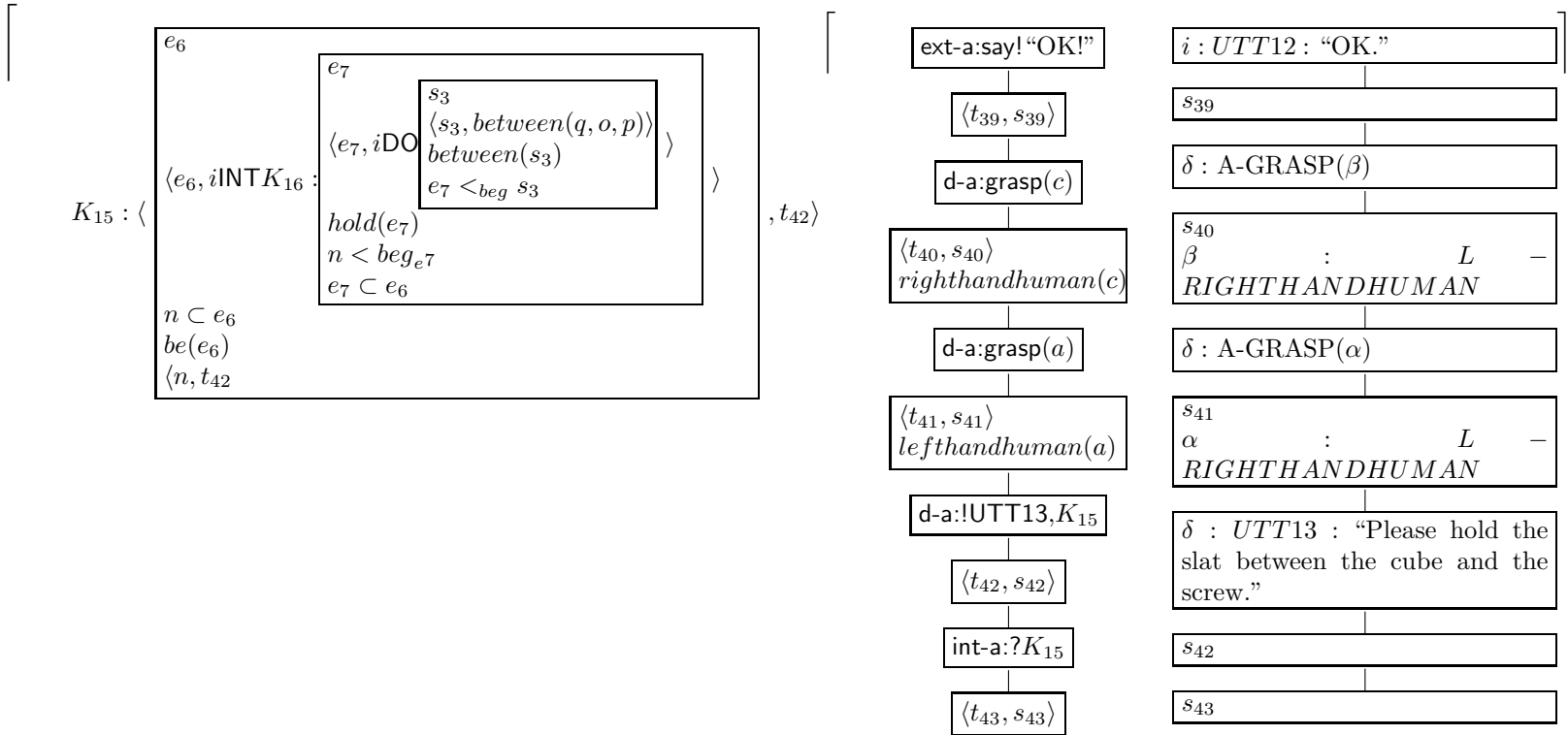


Figure 7.83: Discourse analysis of free interaction, Part 9. Interpretation of K_{15} constructed from $UTT13$.

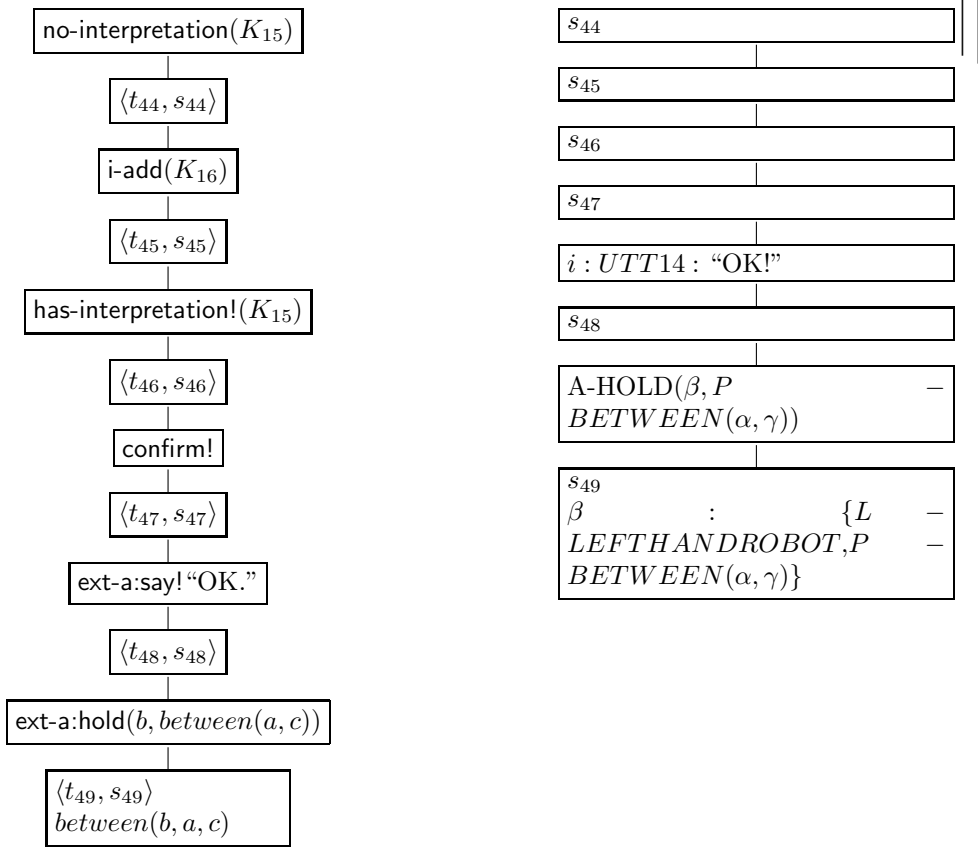


Figure 7.84: Discourse analysis of free interaction, Part 10. K_{15} has no interpretation, thus a reactive addition of the goal K_{16} and consequent execution of the corresponding plan is triggered. Then, K_{15} has an interpretation.

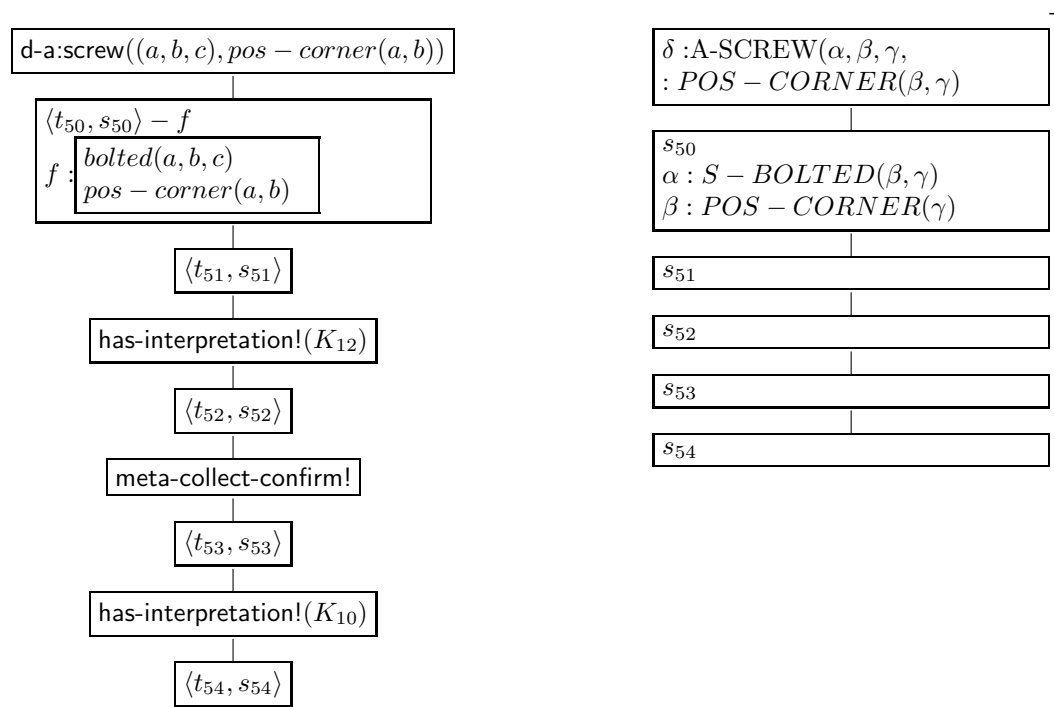


Figure 7.85: Discourse analysis of free interaction, Part 11. The state of affairs at t_{50} results in a series of successful interpretations of K_{12} , K_{10} which are collected by the plan *meta-collect-confirm!*.

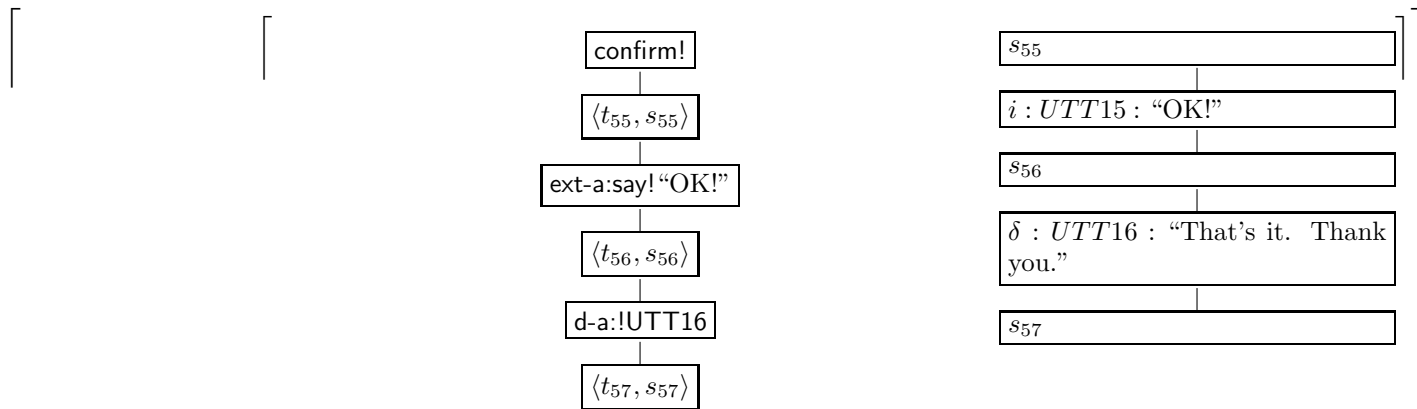


Figure 7.86: Discourse analysis of free interaction, Part 12. Finishing sequence. All open goals and intentions have been realized.

Part IV

Conclusion

Chapter 8

Conclusion

This section reviews the findings of this thesis and gives an outlook on consequent future research associated with the proposed formalism of Grounded Discourse Representation Theory.

8.1 Summary

8.1.1 Grounded Discourse Representation Theory

This study introduced Grounded Discourse Representation Theory (GDRT), a formalism for the semantics-pragmatics interface of a robot in the framework of goal-oriented human-machine collaboration.

As promised in the introduction (chapter 1), I hope that I have convinced the reader of the central importance of pragmatics-based methods for the proper computational processing of discursive interaction. Also, it should have become clear that the use of action-theory based methods for the semantic analysis of discursive interaction opens up new options to treat some of the tough problems of human-machine discourse. Within the limitations of this thesis, I have illustrated only some of the possible applications of GDRT by means of example, e.g. the connection between tense and planning, the connection between reference and explanation or the use of GDRT in the processing of discursive interaction. However, some of the applications of GDRT that are possible within the framework specified in chapter 6 have not been illustrated with examples - this concerns e.g. the use of variable external anchors or the analysis of quantification. Making use of the full range of possibilities GDRT offers for the analysis of discursive interaction could be considered as part of a larger research program in 'computational pragmatics' for which this thesis spelled out the basic ingredients. I say more on future research topics related to GDRT in paragraph 8.3. Before I do so, I give an overview of the argumentation pursued in this thesis and discuss some of the remaining open questions.

8.1.2 Overview of the argumentation

- Chapter 1 defined the goals of this thesis as well as a general cognitive architecture within which these goals were tackled. I also discussed the methods and limitations of this thesis.
- Chapter 2 outlined the general motivation for the use of the basic terms *reference*, *model* and *meaning* in the context of this thesis, as these notions constitute the theoretical backbone of

the subsequent argumentation. After sketching common attempts to the analysis of those terms, I discussed the issues crucial to the development of an artificial agent who is able to make meaningful and successful use of language were discussed.

- Chapter 3 prepared the ground for the formalization of the concepts related to reference and models introduced in chapter 2 by linking the proposed theories of reference and models to the overall architecture of GDRT proposed in chapter 1.
- Chapter 4 introduced the real-time control architecture of GDRT, the agent layer.
- Chapter 5 considered how internal states of an agent provide a background against which she realizes her goals and intentions and in turn how other agents in her environment can make sense of her behavior. The interpretation of external manifestations of an agent's internal states was identified as the core concept behind the processing of goal-directed interactions.
- Chapter 6 stated the formal specifications of the formalism developed so far. This includes the syntactic and semantic definition of the EPS, SMS and IRS layer.
- Chapter 7 applied the developed mechanisms to the analysis of examples of discursive interaction.

8.2 Some notes on open questions

8.2.1 The relation between DRT and GDRT

A question that may have occurred to the reader who is familiar with DRT is how GDRT relates to DRT. The answer to this question is a bit tricky, as on the one hand GDRT is developed as an extension of the core formalism of DRT which is intended to preserve the ideas that motivated the development of DRT with respect to the semantic analysis of natural language beyond the level of single sentences. But on the other hand, the scope of application for which GDRT is designed falls outside the scope of Standard DRT. GDRT renders possible to treat some of the problems for which Standard DRT has not yet provided fully specified solutions. This concerns in particular the systematic use of anchors - for discourse reference markers that represent thing individuals as well as for markers that represent time individuals. While it is possible to reduce the representations of time individuals in the IRS to n-place relations between agents and patients as in Standard DRT, I argued in extenso that the central role of temporal anchoring for the pragmatically meaningful interpretation of utterances and thoughts can hardly be handled within a relational approach to time individuals but only within a functional framework.

The consequent use of anchors allows to combine *normative* (pragmatic) and *descriptive* (semantic) approaches to discourse processing. I call GDRT normative in the sense that its central goal is to derive appropriate *future* options of (re-action) that serve the realization of discourse goals, that it says what should be done. Theories such as DRT are descriptive in the they *describe* the processes which are supposed to take place in the minds of the discourse participants when they try to make sense of a given discourse. The combination of descriptive semantic and normative pragmatic meaning via the concept of dynamic interpretation probably constitutes the main technical innovation of GDRT with respect to DRT. GDRT renders possible the explicit modeling of the pragmatic and semantic processes involved in the proper participation in a discursive interaction. From this perspective, GDRT takes

one step back behind the scenes of discourse processing in that the agent layer allows to explicitly state the processes underlying the use and interpretation of mental representations in the framework of goal-directed interaction.

With respect to the initial question after the relation between DRT and GDRT, the above considerations result in at least two possible answers. First, GDRT can be considered as an *extension* of DRT with respect to the semantic evaluation of representations. Second, GDRT can also play the role of a *metaphor-malism* for DRT that specifies the otherwise implicitly assumed operations that are to be performed to construct and evaluate DRSs. Consequently, GDRT can be considered an attempt to specify the construction and maintenance of a general conception of context, which closely relates GDRT to recent attempts in DRT to incorporate discourse- *and* utterance-related contexts into the framework of DRT [Kamp, 2008].

8.2.2 GDRT and Axiomatization

It should have been noticed that the formalism of GDRT as spelled out in this thesis comes without explicit inference mechanisms in the sense of a syntactic system of axioms. In fact, this is intended. In the following I want to specifically address this point, as it is crucial to a proper understanding of the intention behind GDRT. Roughly speaking, the duties which are usually taken over by explicit systems of axioms (e.g. as in [Singh and Asher, 1993, Asher and Lascarides, 2003]) are implicitly built into GDRT via the BDI-interpreter and the related system of plans and invocation conditions. This is primarily motivated by the fact that planning is much more flexible than static systems of axioms. This concerns in particular the uncertainty attached to (representations of) temporal variation. While systems of axioms must usually be weakened by the introduction of non-monotonic implication or additional constraints to capture temporal side-effects, branching-time planning considers the influence of metaplans, the context of reality and utterances as well as concurrent plans and intentions and thus allows for a more flexible picture of human (and consequently robotic) commerce with temporal variation.

A counter-example to the non-axiomatic approach I propose in this thesis is that of logic programming as spelled out in [van Lambalgen and Hamm, 2004] and in connection with DRT in [Hamm et al., 2006]. While I share their basic assumptions with respect to the relation between planning and tense or the segmentation of temporal variation, the Lambalgen-Hamm approach differs from GDRT in several important aspects. First of all, it only allows for causal explanations (thus it is reductionistic in the sense spelled out in section 2.1.2). Second, it assumes time to be equal to the real numbers (against which good arguments exist from both a linguistic [Fernando, 2006] and physiological point of view [Fingelkurts et al., 2007]). Third, it does not use a cognitively acceptable notion of planning with respect to intentions (section 4.1.2). Finally, as the Lambalgen-Hamm approach is a syntactic approach to temporal variation, it has to face the objections against non-grounded approaches to temporal variation as raised in section 5.3. However, the approach of logic programming can help in determining the temporal profiles of given temporal entities if the interpretation of a given time-individual is considered as a proof from the interpreter's cognitive state.

If necessary, an explicit syntactic axiomatization of the proposed formalism of GDRT can be constructed along the lines of established logics of multi-agent-systems [Singh, 1994, Wooldridge, 2000, e.g.] - the output of the BDI-interpreter is in accordance with the usual axioms for multi-agent-systems as these axioms are constructed with respect to the output of BDI-interpreters.

8.2.3 GDRT and other approaches to human-machine interaction

While there exist numerous *semantic* approaches to human-machine interaction, the in-depth discussion of the *semantics-pragmatics interface* in the framework of GDRT is new to the literature. Consequently, it is difficult to compare GDRT and other approaches in which the semantics-pragmatics interface (in the sense it is proposed in this thesis) is discussed only informally and assigned a marginal role. This holds for both approaches from linguistics and robotics. To name only some examples from the literature, existing approaches to the semantics-pragmatics interface focus either on the cognitive abilities of a robot who is able to make sense of her environment (e.g. [Ahn, 2000]) or on the linguistic ability of instruction understanding (e.g. [Lauria et al., 2002]). Linguistically motivated approaches to the semantics-pragmatics interface without a direct connection to human-machine interaction (e.g. [Asher and Lascarides, 2003, Poesio and Traum, 1998]) blind out the central role of sensomotoric grounding in the relation between semantics and pragmatics and can consequently not directly be employed for the processing of human-robot interaction. Approaches that explicitly seek to integrate methods from linguistics and robotics (e.g. [Christensen et al., 2009] or [Rickert et al., 2007]) go a way toward the goals of this thesis in a technical sense. However, these approaches lack the tight integration of truth-based semantics and action-based pragmatics this thesis has argued for.

8.3 Outlook

8.3.1 Future research

This thesis focused on a particular view of the semantics-pragmatics interface and consequently presented only a partial picture of the processes necessarily involved in an application of the theory developed here to human-robot interaction. Hence, the fragment of GDRT I presented in this thesis should be extended in several directions of which I will briefly mention some:

- *Syntactic Integration* It is necessary to integrate GDRT into a syntactic framework of parsing and generation and to precisely formulate the mappings between the surface of utterances, their semantic representation and pragmatic interpretation. This thesis has illustrated only the very basic mechanisms for doing so and must be extended to a more fine-grained linguistic analysis (e.g. along the lines of [Kamp and Rossdeutscher, 1994]).
- *Integration of research in formal semantics* Another promising line of research associated with GDRT concerns the integration of more findings in e.g. Standard DRT into the framework of GDRT, e.g. with respect to quantification, temporality, underspecification, presupposition and nominalizations. I think that the core architecture of GDRT introduced in this thesis can substantially contribute to the analysis of each of these topics.
- *Serious planning*: A serious account to discourse processing must consult on more elaborated theories of planning (e.g. along the lines of [Inverno et al., 2004]) than the rudimentary 'toy'-plans employed in this thesis did. It also stands to reason to integrate conceptions of group intentions and work distribution (e.g. along the lines of [Grosz and Kraus, 1996]).

- *Learning*: I have not discussed in detail the learning of new concepts. Further investigations into this direction constitute a prime example for the application of stochastic methods in machine learning to the symbol-based framework of GDRT.
- *Multi-modality* I have only made sparse use of multi-modality in the analysis of examples. Nevertheless, the theory of GDRT proposed here offers the possibility to integrate multi-modal interactions without further efforts.

8.3.2 Closing words

The approach of GDRT presented in this thesis can be considered a first step toward a theory of 'computational pragmatics' in the sense of an action-theory based processing of discursive interaction. However, the research area of computational pragmatics has to face problems quite similar to those in the established field of computational semantics. In both areas of research, the development of a large-scale lexicon (of words resp. actions) and grammar (of sentences resp. plans) is of central importance. While substantial efforts have been made in the last years in the area of computational semantics, computational pragmatics (in the sense proposed in this thesis) has not yet been tackled in a similar way.

Combining the elaborated algorithms of computational semantics with the more hardware-based research in robotics under consideration of insights from philosophy and psychology is an area of research that promises new insights into the nature of thinking and communication. I hope that this thesis can be considered a plea for the fruitfulness of interdisciplinary research methods.

Part V

References

Bibliography

- R. P. Abelson. Psychological status of the script concept. *American Psychologist*, 36:715 – 729, 1981.
- R. Ahn. *Agents, Objects and Events. A computational approach to knowledge, observation and communication*. Library Technische Universiteit Eindhoven, Eindhoven, 2000.
- E. Anscombe. *Intention*. Basil Blackwell, Oxford, 1957.
- N. Asher. Belief in discourse representations theory. *Journal of Philosophical Logic*, 15:127 – 189, 1986.
- N. Asher. *Reference to Abstract Objects in Discourse*. Kluwer, Dordrecht, 1993.
- N. Asher and A. Lascarides. *Logics of Conversation*. Cambridge University Press, Cambridge, 2003.
- J. L. Austin. *How to do Things with Words*. Oxford University Press, New York, 1962.
- R. G. Barker and H. F. Wright. *Midwest and its children: The psychological ecology of an American town*. Row, Peterson and Company, Evanston, 1954.
- J. Barwise. Scenes and other situations. *The Journal of Philosophy*, 77:369 – 397, 1981.
- J. Barwise and J. Perry. *Situations and Attitudes*. MIT Press, Cambridge, 1983.
- A. Benz, G. Jaeger, and van Rooij R. *Game Theory and Pragmatics*. Palgrave Macmillan., (2005).
- D. Bonevac and H. Kamp. Quantifiers defined by parametric extensions. Technical report, Center for Cognitive Science GRG 220 The University of Texas at Austin, Austin, Texas 78712, 1987.
- J. Bos and P. Blackburn. Working with discourse representation theory an advanced course in computational semantics. Draft, 2010. URL www.comsem.org.
- G. Bower. Plans and goals in understanding episodes. In A. Flammer and W. Kintsch, editors, *Discourse Processing*. North-Holland Publishing Company, Amsterdam, 1982.
- M. E. Bratman. *Intention, Plans, and Practical Reason*. Harvard University Press, Cambridge, 1987.
- M. E. Bratman. Plans and resource-bounded practical reasoning. *Cocomputational Intelligence*, 4(4):349 – 355, 1988.
- J. Carletta, C. Nicol, T. Taylor, R. Hill, J. P. de Ruiters, and E. G. Bard. Eyetracking for two-person tasks with manipulation of a virtual world (under revision). *Behavior Research Methods, Instruments, and Computers*., Under Revision.

-
- R. Carnap. *Meaning and Necessity: A Study in Semantics and Modal Logic*. University of Chicago Press, Chicago, 1947.
- H.I. Christensen, A. Sloman, G.-J. Kruijff, and J. Wyatt. Cognitive systems. Technical report, 2009. URL <http://www.cognitivesystems.org/cosybook/index.asp>.
- P. R. Cohen and J. H. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42:213–261, 1991.
- P. R. Cohen and C. R. Perrault. Elements of a plan-based theory of speech acts. pages 423–440, 1986.
- K. J. W. Craik. *The Nature of Explanation*. Cambridge University Press, Cambridge, 1967.
- D. Davidson. Actions, reasons and causes. *Journal of Philosophy*, 60:695 – 700, 1963.
- D. Davidson. The logical form of action sentences. In N. Rescher, editor, *The Logic and Decision of Action*, pages 81 – 95. The University of Pittsburgh Press, Pittsburgh, 1967.
- D. Davidson. *Essays on Actions and Events*. Clarendon Press, Oxford, 2001a.
- D. Davidson. Truth and meaning. In *Inquiries into Truth and Interpretation.*, pages 17 – 42. Clarendon Press, Oxford, second edition, 2001b.
- D. Dennett. *The intentional stance*. MIT Press, Cambridge, 1989.
- D. Dennett. Three kinds of intentional psychology. In R. Boyd, P. Gasper, and J.D. Trout, editors, *The Philosophy of Science*, pages 631–650, Cambridge, 1991a. MIT Press.
- D. Dennett. *Consciousness explained*. Little, Brown and Co., Boston, New York, 1991b.
- D. R. Dowty. *Word Meaning und Montague Grammar*. Springer, New York, 1979.
- F. Dretske. *Seeing and Knowing*. Routledge & Keagan Paul, London, 1969.
- F. Dretske. *Explaining Behavior. Reasons in a World of Causes*. MIT Press, 1988.
- E. A. Emerson. Temporal and modal logic. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science Vol. B*, Amsterdam, 1990. North-Holland Publishing Company.
- V. Evans. *The Structure of Time. Language, meaning and temporal cognition*. John Benjamins Publishing Company, Amsterdam, 2003.
- R. Fagin and J. Y. Halpern. Belief, awareness, and limited reasoning. *Artificial Intelligence*, 34(1):39–76, 1987.
- T. Fernando. Representing events and discourse; comments on Hamm, Kamp and van Lambalgen. *Theoretical Linguistics*, 32(1):57–64, 2006.
- A. A. Fingelkurts and A. A. Fingelkurts. Timing in cognition and EEG brain dynamics: discreteness versus continuity. *Cognitive Processing*, 7:135 – 162, 2006.

- A. A. Fingelkurts, Fingelkurts A. A., and C. M. Krause. Composition of brain oscillations and their functions in the maintenance of auditory, visual and audio-visual speech percepts: an exploratory study. *Cognitive Processing*, 8:183 – 199, 2007.
- G. Frege. *Translations from the Philosophical Writings of Gottlob Frege*. Blackwell, Oxford, 1960.
- G. Frege. On sense and reference. In A. W. Moore, editor, *Meaning and Reference*, Oxford Readings in Philosophy. Oxford University Press, Oxford, 1993.
- D. M. Gabbay. The declarative past and imperative future: Executable temporal logic for interactive systems. In *Temporal Logic in Specification*, pages 409–448, London, 1987. Springer.
- M. P. Georgeff and F. F. Ingrand. Decision-making in an embedded reasoning system. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, Detroit, August 1989.
- M. P. Georgeff and A. L. Lansky. Reactive reasoning and planning. In *Proceedings of the Sixth National Conference on Artificial Intelligence*, pages 677–682, 1987.
- H. P. Grice. Meaning. *The Philosophical Review*, 66(3):377 – 388, July 1957.
- H. P. Grice. Logic and conversation. In P. Cole and J. Morgan, editors, *Syntax and semantics*, volume 3. Academic Press, New York, 1975.
- J. Groenendijk and M. Stokhof. Dynamic predicate logic. *Linguistics and Philosophy*, 14:39 – 100, 1991.
- J. Groenendijk and M. Stokhof. Meaning in motion. In K. von Heusinger and U. Egli, editors, *Reference and Anaphoric Relations*, pages 47 – 76. Kluwer, Dordrecht, 1999.
- B. J. Grosz and S. Kraus. Collaborative plans for complex group action. Technical Report TR-20-95, Center for Research in Computing Technology, Harvard University, Cambridge, 1996.
- B. J. Grosz and C. L. Sidner. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204, 1986.
- P. Gärdenfors. Belief revision: An introduction. In P. Gärdenfors, editor, *Belief Revision*, pages 1 – 20. Cambridge University Press, 1992.
- F. Hamm, H. Kamp, and M. van Lambalgen. There is no opposition between formal and cognitive semantics. *Theoretical Linguistics*, 32:1–40, 2006.
- D. Hartmann and P. Janich. *Methodischer Kulturalismus*. Suhrkamp, Frankfurt a. M., 1991.
- D. Hartmann and P. Janich. *Die Kulturalistische Wende. Zur Orientierung des philosophischen Selbstverständnisses*. Suhrkamp, Frankfurt a. M., 1998.
- P. J. Hayes. Naive Physics I: Ontology for liquids. In J. R. Hobbs and R. C. Moore, editors, *Formal Theories of the Commonsense World*, Ablex Series in Artificial Intelligence, pages 71 – 108. Ablex, Norwood, 1985.
- J. Hintikka. *Knowledge and Belief*. Cornell University Press, Ithaca, 1962.

- J. Hintikka. Impossible possible worlds vindicated. *Journal of Philosophical Logic*, 4(3):475 – 484, 1975.
- C. von Hofstein. Planning and perceiving what is going to happen next. In M. M. Haith, B. B. Janette, R. J. Jr. Roberts, and B. F. Pennington, editors, *The Development of Future-Oriented Processes*, chapter 3, pages 63 – 86. The University of Chicago Press, Chicago, 1994.
- C. Hubig. *Mittel*. transcript, Bielefeld, 2002.
- M. N. Huhns and M. P. Singh, editors. *Readings in Agents*. Morgan Kaufmann, San Francisco, 1998.
- F. F. Ingrand and M. P. Georgeff. Managing deliberation and reasoning in real-time ai systems. In *Proceedings of the 1990 DARPA Workshop on Innovative Approaches to Planning*, San Diego, 1990.
- F. F. Ingrand, R. Chatila, R. Alami, and F. Robert. PRS: A high level supervision and control language for autonomous mobile robots. In *Proceedings of the 1996 IEEE International Conference on Robotics and Automation*, Minneapolis, 1996.
- M. D' Inverno, M. Luck, M. P. Georgeff, D. Kinny, and M. J. Wooldridge. The dMARS architecture: A specification of the distributed multi-agent reasoning system. *Autonomous Agents and Multi-Agent Systems*, 9:5–53, 2004.
- P. N. Johnson-Laird. *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Cambridge University Press, Cambridge, 1983.
- P. N. Johnson-Laird. Mental models and thought. In K. J. Holyoak and R. G. Morrison, editors, *The Cambridge Handbook of Thinking and Reasoning*, chapter 9, pages 186 – 208. Cambridge University Press, Cambridge, 2005.
- P. N. Johnson-Laird. The history of mental models. In K. Manktelow and M. C. Chung, editors, *Psychology of Reasoning: Theoretical and Historical Perspectives.*, chapter 8, pages 179 – 212. Psychology Press, New York, 2004.
- H. Kamp. A theory of truth and semantic representation. In J. Groenendijk, T. M. V. Janssen, and M. Stokhof, editors, *Truth, Interpretation and Information: Selected Papers from the Third Amsterdam Colloquium*, pages 1–41. Foris Publications, Dordrecht, 1984.
- H. Kamp. Prolegomena to a structural account of belief and other attitudes. In J. Anderson, C. Owens, editor, *Propositional Attitudes. The Role of Content in Logic, Language and Mind.*, volume 20, pages 27 – 90. CSLI Lecture Notes, Stanford, 1990.
- H. Kamp. Einstellungszustände und Einstellungszuschreibungen in der Diskursrepräsentationstheorie. In Ulrike Haas-Spohn, editor, *Intentionalität zwischen Subjektivität und Weltbezug*. Mentis, 2003.
- H. Kamp. Intentions, plans and their execution: Turning objects of thought into entities of the external world. Unpublished Manuscript, IMS University of Stuttgart, 2007.
- H. Kamp. Discourse structure and the structure of context. Unpublished Manuscript, IMS University of Stuttgart, 2008.

- H. Kamp and A. Bende-Farkas. Verbs of creation. Unpublished Manuscript, IMS University of Stuttgart, 2005.
- H. Kamp and U. Reyle. *From Discourse to Logic*. Kluwer, Dordrecht, 1993.
- H. Kamp and A. Rossdeutscher. Remarks in lexical structure and DRS construction. *Theoretical Linguistics*, 20:98–164, 1994.
- H. Kamp, J. van Genabith, and U. Reyle. Discourse Representation Theory. In D. Gabbay and F. Guenther, editors, *Handbook of Philosophical Logic*, Dordrecht, 2007. Kluwer.
- S. C. Kleene. *Introduction to Metamathematics*. North-Holland Publishing Company, 1952.
- K. Konolige. *A Deduction Model of Belief*. Morgan Kaufmann Publishers, San Francisco, 1986.
- M. Krifka. *Nominalreferenz und Zeitkonstitution. Zur Semantik von Massentermen, Individualtermen, Aspektklassen*. PhD thesis, University of Munich, 1986.
- M. Krifka. The origins of telicity. In S. Rothstein, editor, *Events and Grammar*, pages 197 – 235. Kluwer, Dordrecht, 1998.
- S. A. Kripke. *Naming and Necessity*. Harvard University Press, Cambridge, 1980.
- S. Lauria, T. Kyriacou, G. Bugmann, J. Bos, and E. Klein. Converting natural language route instructions into robot-executable procedures. In *Proceedings of the 2002 IEEE Int. Workshop on Robot and Human Interactive Communication (Roman'02)*, pages 223–228, Berlin, Germany, 2002.
- D. Lewis. *Convention*. Harvard University Press, Cambridge, 1969.
- D. Lewis. Languages and language. In D. Lewis, editor, *Philosophical Papers*, volume One, pages 163 – 188. Oxford University Press, 1983.
- K. Lochbaum. Plan recognition in collaborative discourse. Technical Report TR-14-91, Center for Research in Computing Technology, Division of Applied Sciences, Harvard University, 1991.
- D. Marr. *Vision*. W.H. Freeman and Company, New York, 1982.
- M. Minsky. A framework for representing knowledge. In P. H. Winston, editor, *The psychology of computer vision*. McGraw-Hill, New York, 1972.
- M. Moens and M. Steedman. Temporal ontology and temporal reference. *Computational Linguistics*, 14: 15–28, 1988.
- R. Montague. Universal Grammar. In R.H. Thomason, editor, *Formal Philosophy*, chapter 7, pages 222–246. Yale University Press, New Haven and London, 1979.
- Y. N. Moschovakis. Sense and Denotation as Algorithm and Value. In J. Oikkonen and J. Vaananen, editors, *Lecture Notes in Logic*, number 2, pages 210–249. Springer, Berlin and New York, 1994.
- M. Poesio and D. Traum. Towards an axiomatization of dialogue acts. In J. Hulstijn and A. Nijholt, editors, *Proceedings of the Twente Workshop on the Formal Semantics and Pragmatics of Dialogues Enschede*, pages 207–222, 1998.

- M. E. Pollack. Plans as complex mental attitudes. In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication*, pages 77–103. MIT Press, Cambridge, 1990.
- M. E. Pollack. The uses of plans. *Artificial Intelligence*, 57(1):43 – 68, 1992.
- M. Rickert, M.E. Foster, M. Giuliani, T. By, G. Panin, and A. Knoll. Integrating language, vision and action for human robot dialog systems. In C. Stephanidis, editor, *Proceedings of the 4th International Conference on Universal Access in Human-Computer Interaction, HCI International.*, volume 4555 of *Lecture Notes in Computer Science*, pages 987 – 995. Springer, 2007.
- Fadiga L. Gallese V. Rizzolatti, G. and L. Fogassi. Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3:131 – 141, 1996.
- B. Rogoff, J. Baker, Sennett, and E. Matusov. Considering the concept of planning. In M. M. Haith, B. B. Janette, R. J. Jr. Roberts, and B. F. Pennington, editors, *The Development of Future-Oriented Processes*, chapter 12, pages 353 – 374. The University of Chicago Press, Chicago, 1994.
- D. E. Rumelhart. Notes on a schema for stories. In D. G. Bobrow and A. Collins, editors, *Representation and Understanding: Studies in Cognitive Science.*, pages 211 – 236. Academic Press, London, 1975.
- D. E. Rumelhart. Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce, and Brewer W. F, editors, *Theoretical issues in reading comprehension: Perspectives from cognitive psychology, linguistics, artificial intelligence, and education.*, pages 33 – 58. L. Erlbaum Associates, Hillsdale, 1980.
- R. C. Schank and R. P. Abelson. *Scripts, plans, goals and understanding: an inquiry into human knowledge structures*. L. Erlbaum Associates, Hillsdale, 1977.
- J. R. Searle. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press, Cambridge, 1969.
- J. R. Searle. Minds, brains, and programs. *Behavioral and Brain Sciences*, 3:417 – 424, 1980.
- J. R. Searle. *Intentionality: An essay in the philosophy of mind*. Cambridge University Press, Cambridge, 1983.
- M. P. Singh. *Multiagent Systems. A theoretical framework for Intentions, Know-How and Communications*. Springer, New York, 1994.
- M. P. Singh. A semantics for speech acts. In M. N. Huhns and M. P. Singh, editors, *Readings in Agents*, pages 458 – 470. Morgan Kaufman, San Francisco, 1998.
- M. P. Singh and N. Asher. A logic of intentions and beliefs. *Journal of Philosophical Logic*, 22:513 – 544, 1993.
- M. P. Singh, A. S. Rao, and M. P. Georgeff. Formal methods in DAI: Logic-based representation and reasoning. In G. Weiss, editor, *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*, chapter 8, pages 331 – 376. MIT Press, Cambridge, MA, 1999.
- J. J. C. Smart. The River of Time. *Mind*, LVIII:483–494, 1949.

- D. Sperber and D. Wilson. *Relevance: Communication and Cognition*. Blackwell, Oxford, 1986.
- D. Sperber and D. Wilson. Linguistic form and relevance. *Lingua*, 90(2):1 – 25, 1993.
- P. F. Strawson. On referring. *Mind*, 59:320 – 344, 1950.
- P. F. Strawson. *Individuals: An Essay in Descriptive Metaphysics*. Methuen, London, 1959.
- A. Tarski. The Concept of Truth in Formalized Languages. In *Logic, Semantics, Metamathematics. Papers from 1923 to 1938 by Alfred Tarski.*, chapter 7, pages 152 – 278. Clarendon Press, Oxford, 1956.
- T. A. van Dijk and W. Kintsch. *Strategies of Discourse Comprehension*. Academic Press, Orlando, 1983.
- B. van Fraassen. Singular terms, truth-value gaps and free logic. *Journal of Philosophy*, 63(17):481 – 495, 1966.
- M. van Lambalgen and F. Hamm. *The Proper Treatment of Events*. Blackwell, Oxford, 2004.
- Z. Vendler. Verbs and times. *The Philosophical Review*, 66(2):143 – 160, April 1957.
- G. H. von Wright. *Explanation and Understanding*. Cornell University Press, Ithaca, 1971.
- L. Wittgenstein. *Tractatus Logico-Philosophicus*. Routledge & Keagan Paul, London, 1922.
- L. Wittgenstein. *Philosophical Investigations*. Blackwell, Oxford, 1953.
- M. Wooldridge. *Reasoning about Rational Agents*. MIT Press, Cambridge, 2000.
- J. M. Zacks and K. M. Swallow. Event segmentation. *Current Directions in Psychological Science*, 16(2):80–84, 2007.
- J. M. Zacks and B. Tversky. Event structure in perception and conception. *Psychological Bulletin*, 127:3 – 21, 2001.
- J. M. Zacks, B. Tversky, and G. Iyer. Perceiving, remembering and communicating structure in events. *Journal of Experimental Psychology: General*, 130:29 – 58, 2001.
- J. M. Zacks, Swallow K. M., J. M. Vettel, and McAvoy M. P. Visual motion and the neural correlates of event perception. *Brain Research*, 1076(1):150 – 162, March 2006.

Teil VI

Zusammenfassung

Zusammenfassung

Die vorliegende Dissertation entwickelt einen Formalismus für das Semantik-Pragmatik-Interface eines Roboters zur Verarbeitung zielorientierter Mensch-Maschine-Interaktion.

Motivation

Die Akzeptanz von Robotern im Alltag hängt entscheidend von der Möglichkeit natürlicher und intuitiver Mensch-Maschine-Interaktion ab. Dies betrifft im speziellen die Kommunikation durch gesprochene Sprache, Gesten und Gesichtsausdrücke. Die vorliegende Arbeit formuliert einen theoretischen Ansatz für die Verarbeitung zielgerichteter Interaktionen zwischen Menschen und Robotern an der Schnittstelle zwischen formaler Semantik und Pragmatik. Die Entwicklung eines solchen Formalismus, der es einem Roboter erlaubt in natürlicher Weise an gemeinsamen Aufgabenstellungen zu partizipieren, involviert zentrale Problemstellungen aus verschiedenen Wissenschaftsdisziplinen: Informatik, Linguistik, Robotik, Logik, Psychologie und Philosophie. So involviert z.B. die adäquate Behandlung des Problems der Bezugnahme bzw. Referenz Aspekte aus allen diesen Einzelwissenschaften. Die in dieser Arbeit entwickelte 'Grounded Discourse Representation Theory' (GDRT) zielt darauf ab, in integrativer Weise Forschungsergebnisse dieser unterschiedlichen Wissenschaftsdisziplinen in einer Theorie zu vereinen.

Ausgangspunkt

Diese Arbeit entwickelt keine vollkommen neue Theorie, vielmehr basiert sie auf existierenden de-facto Standardtheorien. Ihr wesentlicher Beitrag besteht darin, diese unterschiedlichen Theorien zu kombinieren und widerspruchsfrei zu vereinen. Im Fokus steht dabei die Verankerung von expliziten semantischen Repräsentationen im Sinne der Diskursrepräsentationstheorie (DRT, [Kamp et al., 2007, Kamp and Reyle, 1993]) in einer formalen Modelltheorie, die ihrerseits in Perzeptionen und Planungsstrukturen eines Roboters verankert ist. Es ist diese Form der Rückführung der Referenzstrukturen von Repräsentationen auf reale Zustände und zukünftige Pläne, die die Interpretation von Repräsentationen fundiert (daher 'Grounded' Discourse Representation Theory). Andersherum kann die Arbeit als Versuch verstanden werden, die sensomotorischen Fähigkeiten eines Roboters um den Umgang mit komplexen semantischen Repräsentationen zu erweitern und damit an die Forschungsrichtung der Computerlinguistik anzuschliessen. Prinzipiell stellt die GDRT damit einen Versuch dar, eine Brücke zwischen Robotik und Linguistik zu schlagen, indem sie dynamische Modellstrukturen als Mittler zwischen der hardwarenahen Seite der Objekterkennung und Motorkontrolle und der kognitionstheoretisch motivierten Seite des

Gebrauchs von Repräsentationsstrukturen installiert. Der entwickelte Formalismus erweitert dabei die Standardtheorie der DRT um eine dynamische Modelltheorie, die dem Vorwurf, die Dynamik der DRT beschränke sich auf eine Dynamik der Repräsentation [Groenendijk and Stokhof, 1999] insofern erwidert, als die GDRT eine dynamische Theorie der Interpretation von Repräsentationsstrukturen bereitstellt, die mit den zum Standard avancierten Verfahren der DRT in Bezug auf u.a. Anaphora, Quantifikation und Tempus kompatibel ist.

Die GDRT wie sie in dieser Arbeit entwickelt wird, bemüht sich dabei insbesondere die auf [Asher, 1986, Kamp, 1990] zurückgehende Ankertheorie der Referenz im Rahmen der DRT durch einen detailliert ausformulierten Interpretationsmechanismus auf Basis einer mengentheoretischen Semantik zu untermauern. Grundidee ist hierbei die Annahme, dass die Identifikation von Sachverhalten und involvierter Entitäten, auf die sich eine bestimmte Repräsentation bezieht, durch Erklärungen solcher Sachverhalte und Entitäten ermöglicht wird. Eine Erklärung wird dabei so verstanden, dass sie Informationen liefert, wie ein gegebener Sachverhalt von seinem Kontext unterschieden werden kann, d.h. welche spezifische Qualität diesen Sachverhalt auszeichnet. Für die Analyse von Temporalreferenz wird dabei auf den psychologisch fundierten Vorschlag von [Zacks et al., 2001] zurückgegriffen, in dem die Segmentierung kontinuierlicher temporaler Prozesse als auf Kausal- und Planungsstrukturen basierende Erklärung verstanden wird. D.h. ein beobachteter Prozess kann durch die Annahme einer zugrunde liegenden Temporalstruktur (Kausal oder Plan) als eine temporale Entität identifiziert werden, auf die dann mit einer Repräsentation referiert werden kann. Formal wird die Beziehung von Temporalstruktur und Repräsentation mithilfe von temporalen Ankerpunkten abgebildet, die einem gegebenen Temporalreferenten ein temporales 'Profil' zuordnen.

Die Arbeit stellt auch einen neuen Ansatz zur Behandlung epistemischer und doxastischer Einstellungen vor, deren Interpretation auf der spezifischen Art und Weise der Verankerung involvierter Sachverhalte beruht.

Die GDRT kann auch als Versuch aufgefasst werden die von [Asher, 1986] formulierte Idee, die DRT als eine 'Sprache der Gedanken' aufzufassen, in die Tat umzusetzen und damit den Anwendungsbereich der DRT auf nichtsprachliche Domänen zu erweitern, ohne die vielfältigen Möglichkeiten der DRT zur Analyse Phänomene natürlicher Sprache einzuschränken.

Übersicht über die GDRT

Der technische Hintergrund der GDRT

Aus technischer Perspektive stellt sich die Aufgabe dieser Arbeit als Einbettung der Grundannahmen der DRT in eine Kombination aus verzweigender Zeitlogik (basierend auf der Modelltheorie der Computational Tree Logic, CTL [Emerson, 1990])) und Multiagenten- und Planungstheorie [Singh, 1994, Singh et al., 1999, Inverno et al., 2004] dar. Philosophisch und psychologisch wird diese Einbettung durch die Belief-Desire-Intention-Handlungstheorie ([Bratman, 1987]) und Erklärungsmechanismen zeitlicher Veränderung ([Dretske, 1988, Dennett, 1989, Hartmann and Janich, 1991, Zacks et al., 2001]) fundiert. Für die GDRT wird eine dreigeteilte Architektur vorgeschlagen, die aus einer Repräsentationsschicht (IRS, interne Repräsentationsstruktur), einer Präsentationsstruktur (EPS, externe Präsentationsstruktur) und den Daten der Sensomotorik (SMS, Sensomotorische Struktur) besteht. Diese drei

Strukturen werden durch eine agentenbasierte Kontrollstruktur verbunden und in Beziehung gesetzt. Die Architektur der GDRT stellt sich damit folgendermassen dar:

- **IRS, Mentale Repräsentation**

- Repräsentation von Gedanken, Äußerungen, innerer und äußerer Zustände
- Umgang mit Objekt- und Temporalindividuen

- **EPS, Mentales Modell**

- Laufzeitumgebung
- Koordination und Planung von Interaktionen
- Koordination von Realität und IRS
- Präsentation perzeptueller Daten
- Derivation zukünftiger Möglichkeiten

- **SMS, Hardwareschicht**

- Sprach- und Objekterkennung sowie Sprachgenerierung
- Sensomotorische Kontrolle

Eingebettet sind diese drei Komponenten in eine agentenbasierte Kontrollstruktur, die zum einen den Informationsfluss zwischen den einzelnen Komponenten regelt und zum anderen das korrekte Verhalten des Roboters in einer Mensch-Maschine-Interaktion steuert.

Graphisch lässt sich die Architektur wie folgt darstellen:

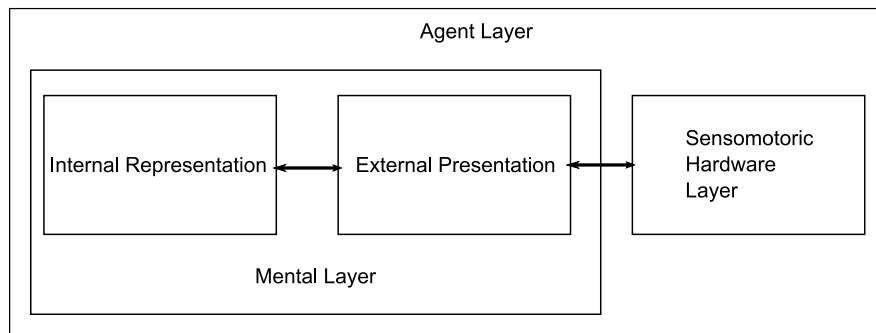


Abbildung 8.1: Schematische Übersicht über die Architektur der GDRT.

Die EPS

Die EPS ist der Dreh- und Angelpunkt des entwickelten Formalismus. Sie spielt dabei eine duale Rolle. Zum einen fungiert die EPS als formales, mengentheoretisches Modell für die Interpretation von IRSen, da sie als formale modale Kripke-Struktur definiert ist. Zum anderen spielt die EPS die Rolle eines dynamischen mentalen Modells, das fortwährend der aktuellen Entwicklung der Realität als auch der inneren Zustände des Agenten in Bezug auf Ziele, Pläne und Intentionen angepasst wird. Insbesondere wird die EPS durch die Objekterkennung und die Planungsmaschinerie des Agenten generiert, d.h. die EPS wird "automatisch" und selbständig generiert. Damit erlaubt es die EPS, die Beschränkungen der klassischen Modelltheorie zu umgehen, deren Modellstrukturen und Verbindung zur Metasprache durch Interpretationsfunktionen im Voraus festgelegt sein müssen. Die vorgeschlagene Definition der semantischen Relation von IRS und EPS erlaubt es, IRSen dynamisch, d.h. inkrementell und in Abhängigkeit von ablaufenden Prozessen zu interpretieren. Damit einher geht eine Ersetzung der klassischen Auffassung von Wahrheitswertsemantik durch eine Interpretationspragmatik, in der IRSen nicht falsch sein können, sondern ein fehlgeschlagener Interpretationsversuch als Aufforderung aufgefasst wird, eine erfolgreiche Interpretation durch Manipulation interner und externer Zustände herbeizuführen.

Die IRS

Die IRS kann als Pendant zur Diskursrepräsentationsstruktur (DRS) der DRT verstanden werden. Sie übernimmt die Aufgabe einer mentalen Repräsentationsstruktur, die in abstrakter Art und Weise den Informationsgehalt von Perzeptionen, Vorstellungen und Äußerungen abbildet und durch Interpretationsmechanismen semantisch mit der EPS verbunden ist. Die GDRT definiert dabei Mechanismen für die Konstruktion von IRSen in Bezug auf Temporal- und Objektindividuen, die die Rolle von Diskursreferenten übernehmen, als auch für die gegenläufige Richtung, der Interpretation von IRSen.

Der Interpretationsalgorithmus der GDRT

Die der IRS zugrundeliegende Theorie zielt konsequent auf die Verankerung von Diskursreferenten. Dies hat unter anderem zur Folge, dass die Einbettungsfunktion der DRT, die normalerweise die Zuweisung von Referenten an modelltheoretische Entitäten vornimmt, durch die Menge der Anker einer IRS ersetzt wird. Dies wiederum ermöglicht die Formulierung eines dynamischen Interpretationsalgorithmus, der die schrittweise Auflösung von Ankern vornimmt, um eine erfolgreiche Interpretation (deren notwendige Bedingung eine erfolgreiche Verankerung ist) herbeizuführen.

Charakteristika der GDRT

Die GDRT ist so konstruiert, dass sie als Theorie zielgerichteter diskursiver Interaktion verstanden werden kann; zwischen Agent und Umwelt oder zwischen mehreren Agenten und ihrer Umwelt. Der zweite Teil dieser Arbeit wendet die theoretischen Ergebnisse der Arbeit auf die Analyse praktischer Beispiele diskursiver Interaktion an. Dabei werden verschiedene Konstellationen wie Lehrer- und Zusammenarbeitsmodus oder freie Interaktion diskutiert.

Die zentrale Neuerung der GDRT gegenüber herkömmlichen Analyseverfahren ist die Verwendung eines 'normativen' Interpretationsmechanismus für Äußerungen. Normativ ist dabei so zu verstehen, dass nicht

nur eine Transformation der syntaktischen Oberfläche der Äußerung in eine semantische Repräsentation stattfindet, sondern die Interpretation einer Äußerung zu einer angemessenen Reaktion führt. Dies hat unter anderem zur Folge, dass das klassische Konzept der Wahrheitswertsemantik in der GDRT durch den Begriff einer erfolgreichen Interpretation, die wiederum auf dem Ausführen von Handlungen beruhen kann, ersetzt wird. Im obigen Beispiel ist es also nicht von Bedeutung, ob es wahr ist, dass Clara eine Würfelverschraubung bauen soll, sondern dass sie ein Verfahren besitzt, mit dem sie die nötigen Operationen bestimmen kann um angemessen auf diesen Satz zu reagieren. Die theoretische Argumentation für eine solche Auffassung von pragmatischer Interpretation von Bedeutung ist das Hauptanliegen des ersten Teils der vorliegenden Arbeit. Der zweite Teil illustriert anhand detaillierter Analysen von Beispielen diskursiver zielorientierter Mensch-Maschine-Interaktion die Anwendung der Ergebnisse des ersten Teils.

Übersicht über den Aufbau der Arbeit

Die Arbeit ist folgendermassen aufgebaut.

- Kapitel 1 führt in Aufgabenstellung, Motivation, Ziele und Architektur der Arbeit ein.
- Kapitel 2 diskutiert die grundlegenden Begriffe von Referenz, Modell und Bedeutung wie sie im Kontext dieser Arbeit eingesetzt werden.
- Kapitel 3 verbindet die entwickelten Konzepte von Referenz, Modell und Bedeutung mit der in Kapitel 1 eingeführten Architektur.
- Kapitel 4 führt die Kontrollstruktur des Roboters ein.
- Kapitel 5 untersucht, wie interne Zustände eines Agenten den Hintergrund für die Realisierung ihrer Ziele und Absichten darstellen und wie andere Agenten aus der Annahme eines solchen Hintergrundes Sinn aus dem Verhalten des Agenten ziehen können.
- Kapitel 6 gibt eine formale Spezifikation der GDRT.
- Kapitel 7 wendet den entwickelten Formalismus der GDRT auf Beispiele diskursiver Interaktion zwischen Mensch und Roboter an.

Danksagung und Eigenständigkeitserklärung

Ohne die Unterstützung, Geduld und Diskussionsbereitschaft von Prof. Hans Kamp, Lehrstuhl for formale Logik und Sprachphilosophie am IMS Stuttgart, wäre diese Arbeit nicht möglich gewesen. Hans Kamp ist im Sinne des Wortes der Vater dieser Doktorarbeit. Ihm gilt mein spezieller Dank.

Diese Arbeit wurde von der Studienstiftung des deutschen Volkes mit einem Promotionsstipendium gefördert. Ich möchte mich hierfür bei der Studienstiftung und den zuständigen Gutachtern bedanken.

Danken möchte ich auch den Teilnehmern der Doktorandenforen der Studienstiftung, des Workshops "Reference to abstract objects in natural language" und der Konferenz "Philosophy's relevance in Information Science" für Anregungen und Kommentare.

Weiterhin gilt mein Dank Prof. Alois Knoll, Manuel Giuliani und Tomas By vom Lehrstuhl für Robotik und Echtzeitsysteme an der TU München für anregende Diskussionen und die Gewährung der Möglichkeit, meine Arbeit auf ein reales Forschungsszenario der Robotik auszurichten.

Bedanken möchte ich mich auch bei Prof. Christoph Hubig, Institut für Philosophie, Universität Stuttgart, der - wenn auch nicht in so offensichtlicher Weise wie meine beiden Berichte - seinen Anteil an dieser Arbeit hat.

Ein ganz besonderes Dankeschön gilt Anja für ihr Verständnis und ihre Unterstützung und meinem Sohn Merlin. Ich widme diese Arbeit euch beiden.

Danke auch an meine Eltern, die mich immer unterstützt und ermutigt haben.

Hiermit erkläre ich, dass ich diese Arbeit abgesehen von den im Literaturverzeichnis genannten Hilfsmitteln selbstständig und nur mit Ratschlägen der oben genannten Personen verfasst habe.

Stuttgart, den 15.01.2010

Tillmann Pross