

# Completions, Coordination, and Alignment in Dialogue

Massimo Poesio, Università di Trento and University of Essex, poesio@essex.ac.uk  
Hannes Rieser, University of Bielefeld, hannes.rieser@uni-bielefeld.de

## *Abstract*

Collaborative completions are among the strongest evidence that dialogue requires *coordination* even at the sub-sentential level; the study of sentence completions may thus shed light on a number of central issues both at the ‘macro’ level of dialogue management and at the ‘micro’ level of the semantic interpretation of utterances. We propose a treatment of collaborative completions in PTT, a theory of interpretation in dialogue that provides some of the necessary ingredients for a formal account of completions at the ‘micro’ level, such a theory of incremental utterance interpretation and an account of grounding. We argue that an account of semantic interpretation in completions can be provided through relatively straightforward generalizations of existing theories of syntax such as Lexical Tree Adjoining Grammar (LTAG) and of semantics such as (Compositional) DRT and Situation Semantics. At the macro level, we provide an intentional account of completions, as well as a preliminary account within Pickering and Garrod’s alignment theory.

## 1 Introduction

Utterances such as 1.2, 1.3, and 2.2 in the following fragment of a transcript from the Bielefeld Toy Plan Corpus of task-oriented dialogues (Skuplik, 1999) are examples of the constructions that Clark (1996) called COLLABORATIVE COMPLETIONS.

(1.1)

1.1	Inst	So, jetzt nimmst Du [pause] <i>Well, now you take</i>
1.2	Cnst	eine Schraube <i>a screw</i>
1.3	Inst	eine <-> orangene mit einem Schlitz. <i>an &lt;-&gt; orange one with a slit</i>
1.4	Cnst	Ja <i>Yes</i>
2.1	Inst	Und steckst sie dadurch, also <i>And you put it through there, let’s see</i>
2.2	Cnst	von oben <i>from the top</i>
2.3	Inst	von oben, daß also die drei festgeschraubt werden dann <i>from the top, so that the three bars get fixed then</i>
2.4	Cnst	Ja <i>Yes</i>

That agents cooperate is a fundamental assumption in many theories of dialogue (see, e.g., the papers in Cohen et al, 1990, or Clark, 1996). That cooperation requires coordination is a central aspect particularly of Clark's and Garrod's work. Collaborative completions such as those in the example dialogue are among the strongest evidence for the argument that dialogue requires *coordination* even at the sub-sentential level (Clark, 1996; Garrod and Anderson, 1987; Pickering and Garrod, 2004). More in general, the study of sentence completions can shed light on a number of central issues for models of the semantics and pragmatics of dialogues, both at the 'macro' level of dialogue management and at the 'micro' level of the semantic interpretation of utterances. At the macro level, this type of data may be used to compare competing claims about coordination—i.e., whether it is best explained with an intentional model like Clark's and the models discussed in Cohen et al's book, or with a model based on simpler alignment mechanisms like Pickering and Garrod's. At the micro level, completions are clear evidence that intention recognition in dialogue proceeds incrementally, and may provide insights about semantic composition.

In this paper we propose a treatment of collaborative completions in PTT (Poesio and Traum, 1997, 1998; Poesio and Muskens, 1997; Matheson, Poesio and Traum, 2000) a theory of interpretation in dialogue incorporating ideas from (Compositional) DRT (Kamp and Reyle, 1993; Muskens, 1996), Situation Semantics (Barwise and Perry, 1983), and Lexical Tree Adjoining Grammar (LTAG) (Abeille and Rambow, 2000, Joshi 2004) that, we believe, provides some of the necessary ingredients for a formal account of completions, such as an account of interpretation with fragments and an account of grounding moves—the rest being provided either by a theory of intentions such as Tuomela's (Tuomela, 2000) or perhaps by a Pickering and Garrod-style account. In the time it took us to write this paper at least one additional account has appeared (Purver et al, 2006), but we believe that a number of central questions are still outstanding. One issue explored in some detail here is the debate between 'intention-based' and 'alignment-based' models of dialogue. In addition to a more traditional 'intentional' account, we also outline an alignment-based treatment of the phenomenon.

The structure of the paper is as follows. In Section 2 we briefly summarize Clark's theory of coordination in dialogue and our main empirical evidence about completions, data from the Bielefeld Toy Plane Corpus. In Section 3 we introduce PTT. In Section 4 we introduce a treatment of intentions and shared plans incorporating a number of recent ideas in the area. In Section 5, we use the notions introduced in the previous two sections to provide a preliminary account of sentence continuations. In Section 6, we provide a preliminary treatment of continuations in terms of the Pickering and Garrod model. In Section 7 we compare our analysis with related work, especially by Purver *et al.*

## 2 Coordination, Grounding, and Completions

### 2.1 Coordination and Grounding in Dialogue

The theory of interpretation in dialogue developed in this paper relies heavily on the views on dialogue developed by Herbert Clark and colleagues, and summarized, e.g., in (Clark, 1992; 1996). Clark views conversation as a form of **joint activity** just like playing football or playing in a string quartet (Clark 1996, ch. 3). Participants in a conversation engage in **joint projects** at all communicative levels, from uttering sounds to performing illocutionary acts. Clark points out that the traditional view of speech acts as developed primarily by Searle (e.g., 1969) and exposed in (Levinson 1983) ignores the role of listeners, and argues that speech acts are really joint projects between the speaker and the listener (Clark 1996, ch. 5).

In order for these joint communicative activities to be successful, agents need to **coordinate** both on matters of timing (e.g., to avoid speech overlap) and on matters of content, developing together what Clark calls **joint construal** (joint interpretation) of these actions, that need not be the interpretation the speaker originally intended. This coordination is only made possible by the **common ground** (Stalnaker, 1978) between the participants; yet Clark points out that the establishment of a common ground is not automatic (“the common ground isn’t just there, ready to be exploited” – Clark 1996, p. 116). Common ground has to be established with each agent with whom we interact; this is particularly the case for that part of the common ground that has to do with the present conversation. The participants to a joint action need to establish the mutual belief that they have succeeded well enough for current purposes (*Principle of joint closure*, Clark 1996, p. 226).

This establishment does not occur by default, but through a process called **grounding** in which **positive evidence** of understanding at different levels is required (Clark and Wilkes-Gibbs, 1986; Clark and Schaefer, 1989; Clark and Brennan, 1991). Such evidence may consist of signals of understanding like nods or “uh huh”, of presuppositions of understanding like taking up the relevant part of the joint project—e.g, answering the question just asked—etc. This process is structured around **contributions** to discourse—signals successfully understood. Contributions are organized into a **presentation** phase, in which conversant A presents the signal to B, and an **acceptance** phase, in which B gives evidence that she believes she understands what A meant (Clark & Schaefer, 1987, 1989). Presentations and acceptances organize dialogues according to a ‘collateral’ structure which is separate and independent from the types of structure traditionally studied in computational linguistics (Grosz and Sidner, 1986; Asher and Lascarides, 2003), which are focused on what Clark calls the ‘official business’ of the conversation; yet they are still (meta-communicative) acts (Allwood et al, 1993; Bunt, 1995; Clark 1996). Traum (1994) developed this view by providing a systematic theory of the grounding process articulated around **grounding acts** which is the basis of the account of grounding adopted in this paper. Traum’s account also provides a solution to the problem of infinite regress incurred by the Clark and Schaefer formulation of grounding: if every presentation needs an acceptance, how can dialogues ever end? We will see Traum’s solution in Section 3.

### 2.2 Coordination on Utterances and Completions

The typical conversational utterance is not the flawless delivery by the speaker of a complete sentence. Speakers generally do not plan utterances in their entirety, and as a result

often realize mid-way that their delivery is being less than ideal (Schegloff et al, 1977; Clark, 1992, 1996; Alwood et al, 1993; Traum, 1994; Ginzburg, 2008, *inter alia*). In addition, speakers need to coordinate with listeners and ensure the signal is understood. As a result, the typical utterance has a non-linear structure with interruptions, and restarts, and with collateral utterances whose objective is to ensure grounding.

When a speaker cannot produce a faultless delivery, a disruption may ensue in the form of a **repair**. Such disruptions have a fixed structure that includes a **suspension point**, often indicated by a pause, a word cut-off, or a lengthened syllable; a **hiatus** during the period in which the speaker plans how to resume, often occupied with fillers or editing sequences; and a **resumption** performing an operation that Clark calls **replacement** which may involve simply continuing what came before the suspension point, repeating it, substituting it, deleting it, or adding new material to it (Clark 1996, p. 264).

The listener, as well, may interrupt the performance of an utterance, either by initiating a repair when a problem is perceived (see Clark's 'Principle of Repair', Clark 1996, p. 284) or by initiating a **collaborative completion** (Lerner, 1987; Wilkes-Gibbs, 1986) providing their own completion of the utterance, as in (2.1) (from (Lerner, 1987), reported in (Clark 1996, p. 238)) or a **truncation** –interrupting the speakers because they think they already understand the rest of the utterance (Conversely, a speaker may decide to **fade out** an utterance when she perceives that the rest isn't needed.)

- (2.1) Marty: now most machines don't record that slow. So I'd wanna-when I make a tape,  
Josh: **be able to speed it up**  
Marty: Yeah

Completions are the focus of this paper As we will see in Section 4, completions may be performed for a variety of reasons, including signaling understanding, or the desire to be cooperative.

## 2.3 Completions in the Bielefeld Toy Plane Corpus

### 2.3.1 The Bielefeld Toy Plane Corpus

The Bielefeld Toy Plane Corpus (BTPC) is a collection of 22 filmed, speech recorded and transcribed construction dialogues between two agents, the Instructor and the Constructor, whose task is to interactively construct a "Baufix" toy airplane (see Fig. 2.1). The instructor (Inst) explains the constructor (Cnst) how to assemble the airplane. Instructor and Constructor sit at separate tables, have the same collection of "Baufix"-components and can freely communicate. The dialogues were recorded in different conditions so as to set up different contexts for the production and understanding of referring expressions. One dimension of variation was sight condition: total screen, face to face, half-screen allowing eye contact. Different conditions concerning the instructions to be followed by the instructor were also tested. (In some cases the Instructor had to direct the Constructor on the basis of a building plan, in others using an already completed model.)

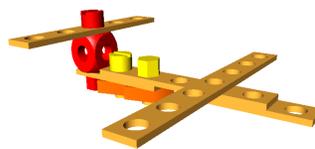


Figure 2.1: Constructor's state of assembly at the beginning of 1.1

The corpus consists of 3675 agents' contributions, counting everything except non-verbal events like groans or laughter.

### 2.3.2 Completions and Continuations in the BTPC

Skuplik (1999) classified 126 cooperative contributions from the BTPC along the following dimensions: (1.) producer, (2.) structural units, (3.) grammatical function, (4.) syntactical category, (5.) resulting syntactical construction, (6.) gapping construction, (7.) acceptance or denial wrt part added by other agent, (8.) wording of acceptance or denial, (9.) indications for change of speaker. Skuplik used the following heuristic definitions to identify completions and continuations. She treated a contribution as a SENTENCE COOPERATION if at least two dialogue participants contributed to its production. In total, 160 contributions (4.34 %) were classified as cooperatively produced turns, most of them other-initiated (95%). Following Wilkes-Gibbs (1995), Skuplik distinguished between two types of sentence extension: a COMPLETION arises if a sub-sentential structure is filled up by obligatory constituents; a CONTINUATION adds material to an already existing sentence.

For our purposes parameters 1, 2, 4, 5, 7, and 9 are of primary interest.

#### (1.) Producer

There were 54 completions (43%) and 72 continuations (57%). In 79% of the cases Cnst produced the completing or the continuing part, in 21% Inst provided the expansions.

#### (2.) Structural Units

61% of sentence cooperations are complete phrases (German "Satzglieder");

#### (4.) Syntactical Category.

Prepositional phrases (37%) are preferred, followed by noun phrases (24%), adverbial phrases (7%), nouns (6%) and sentence parts with finite verbs (5%).

#### (5.) Resulting Syntactical Construction.

Most sentences resulting from cooperation do not show extraposition to the right (27%). Furthermore, we have facultative extraposition (25%), first order extraposition of obligatory constituents (19%), gapping (14%), recursive (second order) extraposition of an obligatory constituent (4%), and subordinate clauses (3%).

#### (7.) Acceptance or Denial of Added Part by other Agent.

84% of the completing or continuing parts were accepted by the previous speaker. 41% of the completing or continuing parts are not explicitly accepted or denied. Explicit acceptance is indicated by e.g. *ja/yes* (28%) and other affirmative particles. In addition, acceptance can be indicated by various forms of resumption or by paraphrase.

#### (9.) Structural Indications for Change of Speaker.

Only in 31% of the sentence cooperations is the change of speaker indicated by prosodic or other means such as various forms of hesitation.

### 2.3.3 The example dialogue at the light of Skuplik's study

The statistical evidence collected by Skuplik suggests that (1.1) is an extremely typical illustration of sentence cooperations in the Bielefeld Toy Plane Corpus. 79% of completions are performed by Cnst. Cnst's completion in this example, *eine Schraube*, is an obligatory NP (30% of the cases), yielding a single complete sentence unit (as in 46% of the cases), making up a sentence if merged with Inst's production (as happens in 50% of the cases). In 31% of the cases a request for a turn release is signalled by prosodic means (lengthening of German *du* and level tone, 31%). Here the evidence is not conclusive that the completing part is not accepted by Inst

(4%), who extends it with *an <-> orange one with a slit* (9%) . Inst's contribution is in turn accepted by Cnst.

In the second joint construction, 2.1-2.4, we find a use of (German) *also* in 2.1, which might indicate that a change of the speaker role will be accepted by current speaker. German *also* frequently indicates a planning pause. However, this time we have a continuation by Cnst, forming a single complete sentence unit of category AdvP, adding finally up to a sentence without extraposition. Cnst acknowledges by resuming the phrase and extending it with a subordinate clause. In 1.2 Cnst's extension acts as a repair, in 2.2 it provides the description of a causal consequence in the domain.

### 3 Incremental Meaning Composition in Dialogue: The PTT Approach

An essential prerequisite of any account of completions is a theory of semantic interpretation in dialogue explaining how the meaning of fragmentary utterances performed by different speakers is incrementally combined to derive more complex interpretations. Our own account is based on such a theory, PTT (Poesio, 1994; Poesio and Traum, 1997; Poesio and Muskens, 1997; Poesio and Traum, 1998; Matheson, Poesio, and Traum, 2000), a theory of dialogue semantics and dialogue interpretation that originated from work on the TRAINS-93 system (Allen et al, 1995). PTT was developed to explain how utterances are incrementally interpreted in dialogue, crucially considering both their semantic impact (e.g., how the occurrence of the pronoun “Sie” in utterance 2.1 of our example dialogue is interpreted) and their impact on other aspects of dialogue interaction traditionally considered as outside the scope of semantic theory (e.g., the role of the two “Ja”s in 1.4 and 2.4), building on the work of Clark (1992, 1996) and on ideas from Situation Semantics (Barwise and Perry, 1983; Cooper and Poesio, 1994; Cooper, 1996, Ginzburg, 2008).

The first distinctive feature of PTT is the assumption—derived from ideas developed in Situation Semantics but also central to Clark’s work, as seen in Section 2—that the common ground doesn’t simply record the propositions asserted or the questions raised, but the whole variety of facts about the *discourse situation* shared between conversational participants (Barwise and Perry, 1983; see also Ginzburg, 2008; Ginzburg, To appear). Among these facts is the occurrence of certain *utterances* of sub-sentential constituents in a certain order. Furthermore, the theory assumes that the occurrence of these so-called *micro-conversational events* also leads to immediate updates of the participants’ *information states* (Larsson and Traum, 2000; Stone, 2004) which in turn leads to the initiation of semantic and pragmatic interpretation processes.

Other facts recorded in the common ground include what Clark called ‘meta-communicative’ acts: dialogue acts which have to do with coordination issues such as turn-taking and the process by which the common ground is established, or *grounding* (Clark and Schaefer, 1989; Brennan, 2005; Traum, 1994). A second feature of PTT that is key for the purposes of this paper is that it provides an explicit account of such meta-communicative acts, and in particular of grounding, building on (Traum, 1994). It is assumed in PTT that new utterances result in the introduction of new *Discourse Units*—the formal correspondent of Clark and Schaefer’s (1989) contributions—which only become part of the common ground as a result of explicit or implicit **Acknowledgments**, and may be cooperatively repaired or revised (as in utterances 1.3 and 2.3 of the example dialogue).

These ideas are formalized using tools derived from Discourse Representation Theory (DRT) (Kamp and Reyle, 1993) – specifically, from Muskens’ Compositional DRT (Muskens, 1996), with additional axioms for specifying the anaphoric behavior of dialogue acts and a simple formalization of events based on (Muskens, 1995). (A brief discussion of Compositional DRT can be found in Appendix B.1.) This means that PTT shares many features with SDRT (Asher, 1993; Asher and Lascarides, 2003); but as we will see, there are crucial differences between the two theories crucially with respect to issues central to this paper such as incrementality and grounding.

#### 3.1 Conversational Events and Discourse Situations

The idea that the shared ‘conversational score’ in a conversation consists only of information about the propositional content of assertions, on which much modern work on the

semantics of discourse rests, is clearly an idealization (Stalnaker, 1978; Barwise and Perry, 1983; Clark, 1992, 1996; Allwood et al, 1993; Traum, 1994; Ginzburg, 2008, to appear). The participants in a conversation also share a great deal of information that they need to coordinate: e.g., whose turn it is to speak, how what is being said fits in within the structure of the rest of the conversation, and whether what has been said has been properly understood (Clark, 1996).

As a result, an ordinary conversation does not consist only of actions performed to assert or query a proposition, but also of actions whose function is to acquire, keep, or release a turn, to signal how the current utterance relates to what has been said before, or to acknowledge what has just been uttered. Bunt (1995) proposed for these utterances the term DIALOGUE CONTROL ACTS. (See also Clark and Schaefer, 1989; Poesio and Traum, 1997; Ginzburg, 1997, 2008). The execution of these actions may involve both linguistic and non-linguistic tools. The linguistic tools include CUE PHRASES such as *so* or (one sense of) *okay*; KEEP-TURN SIGNALS such as filled pauses (*umm*) or *in a minute* in the following fragment reported by Coulthard (1977):

- A They have at their disposal enormous assets //  
and their policy  
B //look can I just come in on that// last year  
A //YES IN A MINUTE IF YOU MAY AND WHEN I'M FINISHED //  
then you'll know  
B // yes I'M SO SORRY

and GROUNDING SIGNALS such as *okay* again, *right* or *uhuh*. Non-verbal means include gaze, gestures such as nods or other head movements, and pointing.

The context update potential of non-assertional speech acts and of dialogue control acts is easy to formalize in terms of a speech act-based theory (Bunt, 1995; Traum and Hinkelman, 1992; Traum, 1994), particularly one in which speech acts are viewed as components of a joint project (Clark, 1996). Poesio and Traum (1997) proposed that the conversational score consists of a record of all actions performed during the conversation, i.e., what in Situation Semantics is called the DISCOURSE SITUATION (Barwise and Perry, 1983; Cooper, 1992; Ginzburg and Sag, 2000; Ginzburg, 2008, to appear). Furthermore, Poesio and Traum argued that this view of the conversational score could be formalized using the tools already introduced in DRT (Kamp and Reyle, 1993; Muskens, 1996), because speech acts-- CONVERSATIONAL EVENTS, in PTT terms-- are in many respects just like any other events, and because conversational events and their propositional contents can serve as the antecedents of anaphoric expressions. So, whereas the ordinary DRT construction algorithm would assign to the text in (3.1.1) an interpretation along the lines of (3.1.2) – i.e., a single DRS containing the merged propositional content of both assertions (using the syntax from Muskens (1996) and his equality operator **is** for equality in DRSs) Poesio and Traum hypothesized that the common ground resulting from (3.1.1) would be as in (3.1.3).<sup>1</sup>

- (3.1.1) a. A: There is an engine at Avon.  
b. B: It is hooked to a boxcar.

- (3.1.2)  $[x,w,y,z,s,s' \mid$   
**engine(x), Avon(w), s: at(x,w), boxcar(y), s':hooked-to(z,y), z is x]**

---

<sup>1</sup> In this paper, discourse referents will be named according to the following conventions. We will use terms with the prefix *ce* (ce1, ce2, etc) for conversational events; terms with the prefix *u* for utterances; terms denoting other events will be indicated by the prefix *e*. We will indicate terms denoting states by the prefix *s*; all other terms will have prefixes *x*, *w*, *y*, and *z*.

(3.1.3) [ce1,ce2]

ce1: **assert**(A,B,[x,w,s | **engine**(x), **Avon**(w), s: **at**(x,w)])  
ce2: **assert**(B,A,[y,z,s' | **boxcar**(y), s':**hooked-to**(z,y), z **is** x)])

(3.1.3) records the occurrence of two conversational events, *ce1* and *ce2*, both of type **assert** (Poesio and Traum, 1998; Matheson, Poesio, and Traum, 2000) whose propositional contents are separate DRSs specifying the interpretation of the two utterances in (3.1.1). The discourse entities *ce1* and *ce2* can serve as antecedents both of implicit and explicit anaphoric references. Implicit anaphoric references include 'backward' acts like **accepts**, as we will see shortly; certain types of clarification requests (Ginzburg, 2008, to appear), and grounding acts, as we will see later in the section. One example of explicit anaphoric reference is the following example, where the *that* uttered by B in 2 appears to be referring to the action of insulting, as opposed to the propositional content of A's utterance (as in 2') or to the locutionary act (as in 2'').

1. A: You're an idiot.
2. B: *That* was uncalled for.
- 2' B: *That's* not true.
- 2'' B: I didn't hear *that*.

As in (Kamp and Reyle, 1993; Muskens, 1995), a Davidsonian treatment of events is assumed, in which each event- or state-describing predicate **p** such as **hooked-to** or **assert** has an additional argument for the event (or state). (We follow Kamp and Reyle's (1993) notation and write  $e: p(x,y)$  rather than  $p(x,y,e)$  for these predicates.)

One immediate advantage of this view is that it can be used to explain the updates to the common ground resulting from conversational events other than assertions, as well as to the all too common situation in which an utterance performs more than one type of conversational event (Traum and Hinkelman, 1992). Even if we ignore the fact that interrogatives and imperatives have non-propositional contents (Poesio and Traum, 1997; Ginzburg and Sag, 2000, Portner 2007)—in this paper, we will for simplicity assume that all contents of conversational events are propositional, as done, e.g., in SDRT (Asher and Lascarides, 2003)—clearly such contents cannot be viewed as providing restrictions on the same set of assignments / worlds as assertions. In (3.1.3'), for example, neither the content of the **open-option** conversational event generated by the first utterance in (3.1.1') nor the content of the **info-request** resulting from (3.1.1'b) express statements about the current state of the world, and should not therefore be evaluated at that index, as they would be using standard DRT semantics if the dialogue were to be assigned the interpretation in (3.1.2'). Notice that (3.1.1'b) may be viewed as achieving at least two purposes: accepting the option proposed with *ce1*, and performing an **info-request**. Notice also that **accept** is implicitly anaphoric to a previous conversational event (*ce1*), as is generally the case with **backward-looking** acts --one argument for assuming that conversational events introduce discourse markers just like normal events do.

- (3.1.1') a. A: We should send an engine to Avon.  
b. B: Shall we use engine E3?

(3.1.2') [x,w,e, y,e']

**engine**(x), **Avon**(w), e: **send**({A,B},x,w)  
**engine**(y), **E3**(y), e':**use**({A,B},y)]

(3.1.3') [ce1,ce2,ce3]

ce1: **open-option**(A,B,[x,w,e | **engine**(x), **Avon**(w), e: **send**({A,B},x,w)])  
ce2: **accept**(B,ce1)  
ce3: **info-request**(B,A,[y,e' | **engine**(y), **E3**(y), e':**use**({A,B},y)])]

**Assert, open-option** and **info-request** in the example above are all examples of CORE SPEECH ACTS (Traum and Hinkelman, 1992; Poesio and Traum, 1997, 1998) – conversational events that express the primary, domain-oriented intention the participant intends to convey. The repertoire of core speech acts assumed in PTT has been changing over the years; the later versions of the theory have been based on the DAMSL repertoire of dialogue acts (Allen and Core, 1997), and core speech acts for which formalizations are provided in PTT include, in addition to **assert** and **open-option**, the forward-looking acts **statement, influencing-addressee-future-act, directive, committing-speaker-future-action, commit**, and **offer**, and the backward-looking acts **agreement, accept, answer**, and **reject** (Matheson, Poesio and Traum, 2000). In addition to core speech acts, utterances may also be used to perform dialogue control acts and grounding acts; we will see examples of these below.<sup>2</sup>

In subsequent work (e.g., (Poesio and Muskens, 1997)), a revised view of the interpretation of speech acts was introduced in PTT, in which the contents of conversational events are associated with propositional discourse referents (discourse referents whose values are DRSs) as proposed, e.g., in SDRT (Asher, 1993; Asher and Lascarides, 2003) and by Geurts (1995). According to this view, the discourse situation resulting from the two utterances in (3.1.1) being interpreted as performing assertions (and nothing else) would be as in (3.1.4), where the propositional contents of *ce1* and *ce2* also become available for subsequent anaphoric reference.

(3.1.4) [ce1,ce2,K1,K2]

K1 is [x,w,s | **engine**(x), **Avon**(w), s: **at**(x,w)],  
 K2 is [y,u,s' | **boxcar**(y), s': **hooked-to**(u,y), u is x],  
 ce1: **assert**(A,B,K1),  
 ce2: **assert**(B,A,K2)]

This type of representation requires complicating the semantics in order to ensure well-foundedness, but at least in principle, there are a number of ways of doing this (Asher, 1993; Geurts, 1995; Poesio and Muskens, 1997); and this representation has the advantage of providing antecedents for references to contents of conversational events—whether explicit as in, e.g., the case in which B follows *ce1* with a denial like *That's not true*, which in PTT would be interpreted as a reference to proposition K1; or implicit, as in the case of grounding acts. We will assume this type of interpretation here and exploit it in our formalization of grounding acts.

Finally, it is assumed in PTT that dialogue acts are **generated** (Goldman, 1970; Pollack, 1986) by locutionary acts (Austin, 1962), which we represent here as events of type **utter**. These events are assumed to become part of the discourse situation as well, at least for a time;<sup>3</sup> indeed, they play an important role in our account of incrementality, as discussed below. We further assume that locutionary acts have a (conventional) semantics associated to them according to

---

<sup>2</sup> The version of PTT discussed in (Poesio and Traum, 1997) also assumes a class of dialogue acts called **argumentation acts** (Traum, 1994) that capture some of the information expressed in SDRT by rhetorical relations. We will not have the space to discuss intentional structure and rhetorical structure in this paper, but these issues are discussed in the follow-up paper (Poesio and Rieser, in preparation).

<sup>3</sup> An important simplification made in PTT is to ignore the issue of forgetting—i.e., the fact that information, particularly linguistic information, only remains 'activated' for a relatively short period. There is quite a lot of evidence that at least certain types of information disappear after a period (Sacks, 1967) although the speed at which this happens is unclear (Fletcher, 1994). A number of utterances in dialogue – so called informationally redundant utterances (Walker, 1993) are also planned with the goal of preventing important information to be forgotten.

standard compositional semantics rules, as we will see in detail in the next subsection; that this semantics is the value of a **sem** function (in fact, a family of functions **sem**<sub>[i]</sub>, **sem**<sub>[π]</sub>, etc.); and that the content of the core speech act generated by a locutionary act with semantics K is also K. After taking these additional assumptions into account, we obtain the picture of the information in the discourse situation after the second assertion in (3.1.1) shown in (3.1.5).

(3.1.5) [u1,u2,ce1,ce2,K1,K2 |  
 u1: **utter**(A,"there is an engine at Avon"),  
 K1 is [x,w,s | **engine**(x), **Avon**(w), s: **at**(x,w)],  
**sem**(u1) is K1,  
 ce1: **assert**(A,B,K1),  
**generate**(u1,ce1),  
 u2: **utter**(B,"it is hooked to a boxcar"),  
 K2 is [y,u,s' | **boxcar**(y), s': **hooked-to**(u,y), u is x],  
**sem**(u2) is K2,  
 ce2: **assert**(B,A,K2),  
**generate**(u2,ce2)]

Therefore, in first approximation—i.e., simplifying the representation of contributions and leaving aside all sub-sentential locutionary acts (see next)—the interpretation of the example dialogue according to the view just presented is as follows.

(3.1.6) [K1.1, up1.1, ce1.1, K2.1, up2.1, ce2.1 |  
 K1.1 is [e,x,x3 | **screw**(x), **orange**(x), **slit**(x3), **has**(x,x3), e:**grasp**(Cnst, x)],  
**utterance**(up1.1), **sem**(up1.1) is K1.1,  
 ce1.1:**directive**(Inst&Cnst,Cnst,K1.1),  
**generate**(up1.1, ce1.1),  
 K2.1 is [x6,e',s,w,y | x6=x, e':**put-through**(Cnst,x6,hole1),  
 w is wing1, y is fuselage1,  
 s: **fastened**(w,y), **purpose**(e',s)],  
**utterance**(up2.1), **sem**(up2.1) is K2.1,  
 ce2.1:**directive**(Inst,Cnst, K2.1) ,  
**generate**(up2.1, ce2.1)]

### 3.2 Micro Conversational Events

A second distinctive feature of PTT with respect to theories such as SDRT is that it takes as a central fact about dialogue that utterances are interpreted incrementally, as suggested by most psychological work on sentence processing (Frazier, 1987; Swinney, 1979; Tanenhaus et al, 1995), and that many contributions to dialogue are fragmentary and non-sentential, as shown by corpus evidence (Poesio, 1995; Fernandez, 2006).

As a result, one of the fundamental hypotheses underlying PTT is that the information state of a conversational participant is updated whenever any new event is perceived, including events such as sub-sentential or even sub-word utterances, and irrespective of whether this event is verbal or non-verbal: non-verbal events such as gestures or nods (McNeill, 1992; Allwood et al, 1992; Clark, 1996; Rieser, 2004) update the information state, as well. Psychological research suggests that such updates can take place every few milliseconds, and that perceiving a phoneme is sufficient to cause an update—in fact, for interpretive processes to begin (Tanenhaus et al, 1995). Here we will simply assume that the view of the discourse situation contained in the information state is updated at least after every word. We use the term MICRO CONVERSATIONAL

EVENTS (MCEs) to refer to events of uttering sub-sentential constituents (Poesio, 1995). Also, we will only discuss here updates caused by linguistic events, although other types of updates have been examined in past PTT work – e.g., to the focus of visual attention (Poesio, 1993) and pointing gestures (Rieser, 2004).

This incremental update hypothesis is motivated not only by these psychological findings about incremental interpretation in sentential utterances, but also by the fact that in dialogue many types of conversational acts hardly if ever require full sentences (Clark, 1996; Fernandez, 2006; Ginzburg, 2008). Among the core speech acts, answers to questions and clarification questions are all typically non-sentential. The utterances used to perform DIALOGUE CONTROL ACTS such as **take-turn**, **keep-turn** and **release-turn** –actions whose function is to synchronize the two participants in the conversation as to who is holding the floor (Sacks et al 1974; Traum and Hinkelman, 1992; Traum, 1994; Bunt 1995; Clark, 1996)—and GROUNDING ACTS, i.e., acts whose function is to keep the common ground synchronized between the two participants (grounding acts are discussed in greater detail below) are also typically non-sentential. These conversational actions are occasionally generated by sentential utterances that also generate a core speech act (e.g., the second utterance in (3.1.1)), but more commonly they are generated by single-word discourse markers—like *okay*, *well*, *now*, all of which may be used to perform a **keep-turn** dialogue control act, or indeed *jetzt* in 1.1 in the example dialogue (1), which arguably is used to perform a **keep-turn** function as well as (possibly) a temporal sequencing one. Non-words such as filled pauses (*umms* and the like) may also be used (Sacks et al, 1974; Clark 1996). To account for the fact that non-sentential utterances may result in updates of the discourse situation it is necessary a theory of update in which the effects of these utterances can be modelled. In PTT, it is hypothesized that observing such utterances may result in adding to the discourse situation **keep-turn** events (as well as other possible events), via updates like those in (3.2.1):

- (3.2.1)            *well* → [u,ce|u: **utter**(A,"well"), ce: **keep-turn**(A), **generate**(u,ce)]  
                       *now* → [u,ce|u: **utter**(A,"now"), ce: **keep-turn**(A), **generate**(u,ce)]  
                       *umm* → [u,ce| u: **utter**(A,"umm"), ce: **keep-turn**(A), **generate**(u,ce)]  
                       *okay* → [u,ce| u: **utter**(A,"okay"), ce: **keep-turn**(A), **generate**(u,ce)]

In this paper we assume that the utterances of *so* at the beginning of 1.1 in the sample dialogue (1) has primarily a dialog control function, so that observing its occurrence results in the following update to the discourse situation. (See Section 5 and Appendix B.2.3.)

- (3.2.2)            *so* → [u,ce|u: **utter**(A,"so"), ce: **take-turn**(A), **generate**(u,ce)]

We will instead assume that *jetzt* in the same utterance is interpreted solely as contributing to the specification of the propositional content of ce1.1, the core speech act generated in that contribution (see (3.1.6)). Of course one could also hypothesize that *jetzt* has a **keep-turn** function instead, as in (3.2.2'), or perhaps both functions.

- (3.2.2')            *jetzt* → [u,ce|u: **utter**(Cnst,"jetzt"), ce: **keep-turn**(A), **generate**(u,ce)]

### 3.3 Incremental Interpretation with Micro Conversational Events

By combining the hypothesis that the information state is updated after every micro-conversational event with the view of semantic composition found in Compositional DRT (as opposed to that of standard DRT) we can explain how utterances whose main function is to express part of the content of a core speech act, like the rest of the utterances in 1.1 - *nimmst*, *Du*,

etc. - do so incrementally, as well, instead of assuming that (3.1.5) or (3.1.6) are derived all at once after the entire mini-dialog in (3.1.1) or the example dialogue have been syntactically analyzed (as it would be necessary when assuming the construction algorithm from (Kamp and Reyle, 1993)).

As suggested by Muskens (1996), using only technical tools already present in Compositional DRT one could already show how (3.1.5) could be derived compositionally and incrementally by concatenating separately produced interpretations for the two utterances in (3.1.1), as in (3.3.1):

(3.3.1) [ce1, K1 | K1 is [x,w,s | **engine**(x), **Avon**(w), s: **at**(x,w)], ce1: **assert**(A,B,K1)] ;  
 [ce2, K2 | K2 is [y,u,s' | **boxcar**(y), s': **hooked-to**(u,y), u is x], ce2: **assert**(B,A,K2)]

But if we accept the two assumptions that locutionary acts are recorded in the common ground, and that single word utterances can lead to updates as well, we can take advantage of Compositional DRT to explain how the utterances of *nimmst* and *Du* can initiate syntactic and semantic interpretation before an entire sentence has been perceived. Psychological research on priming at different levels (as summarized, e.g., in Pickering and Garrod, 2004) and work on clarification questions such as (Ginzburg and Cooper, 2004; Purver and Ginzburg, 2003) provide evidence concerning these updates. This evidence suggests that perceiving an utterance results in the record of the discourse situation being updated with the fact that this utterance just occurred, as well as the results of lexical access – i.e., that utterance’s syntactic classification, and its conventional meaning, which in PTT is identified with the compositional meaning as specified in Compositional DRT (Poesio, 1995; Poesio and Traum, 1997; Poesio and Muskens, 1997; Poesio, To appear). Observing an utterance of the noun *boxcar*, for example, results in an update of the discourse situation which, in first approximation, we will represent as in (3.3.2). This update records the utterance of a new locutionary act *u* (as said above, we use discourse referents with prefix *u* to indicate utterances), of type **utter** (the type of locutionary acts), syntactically classified as a noun, and with semantic content  $\lambda x [|\mathbf{boxcar}(x)]$ .

(3.3.2) [u | **utter**(A,"boxcar"), **Noun**(u), **sem**(u)=  $\lambda x [|\mathbf{boxcar}(x)]$ ]

(We will often use the abbreviated notations *u*: "*boxcar*": *Noun* to indicate the information added by the utterance of a word to the discourse situation, omitting its lexical semantics, and *u*: "*boxcar*":  $\lambda x [|\mathbf{boxcar}(x)]$  that specifies its lexical semantics but omits its syntactic interpretation). A key assumption that PTT derives from psycholinguistic results on lexical access (Swinney, 1979; Tanenhaus et al, 1979) is that lexical access—in fact, all of utterance interpretation—is a process of defeasible inference (Perrault, 1990; Hobbs et al, 1992; Hwang and Schubert, 1993; Poesio, 1994, 1995; Jurafsky, 1996; Asher and Lascarides, 2003) in which competing hypotheses are activated, one of which is rapidly selected, whereas the other ones are discarded (Poesio, 1994, 1995, To appear). The lexicon in PTT is therefore modeled as a default theory (in the sense of Prioritized Default Logic (Brewka 1989)) consisting of default rules specifying lexical updates activated by each utterance, such as (3.3.2).<sup>4</sup>

The PTT view of the interpretive processes that follow lexical access – syntactic interpretation (parsing) and semantic composition—is very much inspired by current work on grammar in frameworks like Tree Adjoining Grammars (Abeille and Rambow, 2000; Joshi, 2004) for what concerns incremental interpretation, and Categorical Grammar (Pereira, 1990; Carpenter,

---

<sup>4</sup> A much more complex view of defeasible inference is adopted in SDRT, with multiple logics for different types of interpretation. For discussion see (Poesio, To Appear).

1994) in that syntactic interpretation is viewed as an inferential process, the only difference being that in PTT interpretation is viewed as an inferential process resulting in updates of the discourse situation. Specifically, parsing is viewed in PTT as a process during which hypotheses about the results of lexical access combine together in phrasal hypotheses. Such phrasal hypotheses are viewed as hypotheses about utterances: e.g., the occurrence of contiguous utterances of type Det and N results in a phrasal hypothesis about the occurrence of an utterance of type NP.

These hypotheses are produced by a second set of defeasible inference rules that encode syntactic competence.<sup>5</sup> As in (Poesio, 1995; Poesio, 2001; Poesio, To appear), we assume here the syntactic framework of Lexicalized Tree Adjoining Grammar (LTAG) (Schabes et al, 1988; Schabes, 1990; Sturt and Crocker, 1996; Abeillé and Rambow 2000), since it lends itself to a very natural account of the process by which syntactic interpretations are constructed incrementally (Sturt and Crocker, 1996).<sup>6</sup> In LTAG, the lexical interpretations of words are *elementary trees*. In the case of sortal nouns like *boxcar* or *Schraube* these trees are *atomic* and take the form in (3.3.3):

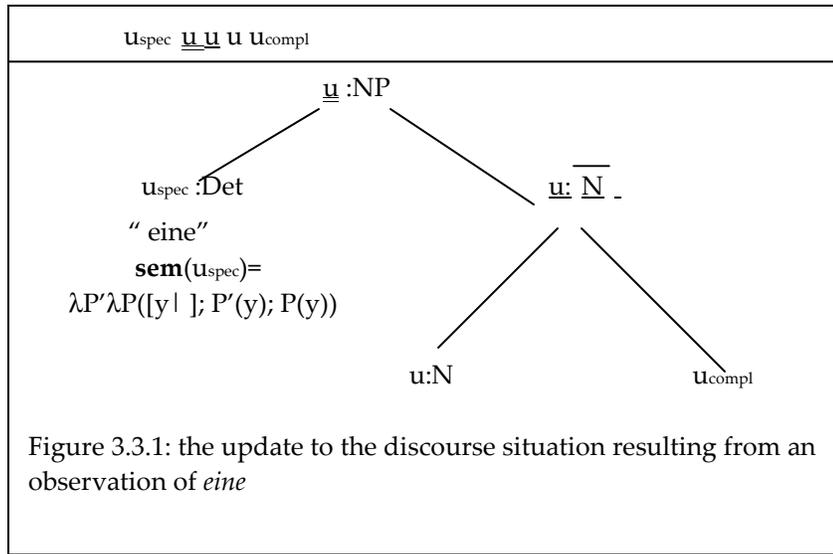
(3.3.3)     N  
               |  
               *Schraube*

but in the case of words whose semantic interpretation takes arguments, such as verbs and determiners, the elementary trees are more complex, and already contain ‘attachment points’ for such arguments. For example, Figure 3.3.1 illustrates the lexical interpretation of the determiner *eine* according to LTAG. In PTT, the utterance of this determiner results in the discourse situation being updated not just by the observation that an utterance  $u_{\text{spec}}$  occurred, but also by the expectation that  $u_{\text{spec}}$  is going to be part of the performance of the utterance  $\underline{u}$  of an NP, of which  $u_{\text{spec}}$  will occupy the specifier position, as well as the performance of an utterance  $u$  of type N and (possibly) of a complement  $u_{\text{compl}}$ . (This node is the attachment point for nominal arguments of relational nouns.)

---

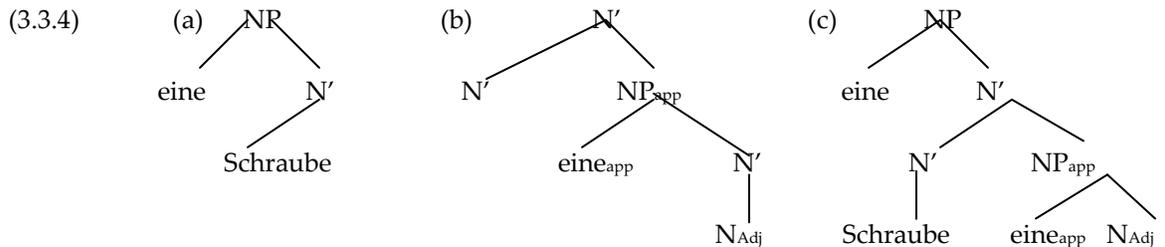
<sup>5</sup> The idea being that the rules encode syntactic *competence* but the choice of one particular syntactic interpretation is the result of performance factors. We believe however that this approach is sufficiently general not to force us to take any particular position wrt the question of whether speakers’ linguistic knowledge includes full syntactic competence or only the competence reflected by their performance.

<sup>6</sup> The LTAG framework has also been adopted in other modern frameworks concerned with semantic interpretation, such as Muskens’ Logical Description Grammar (Muskens, 2001), although it is not used there to provide an account of incremental processing.

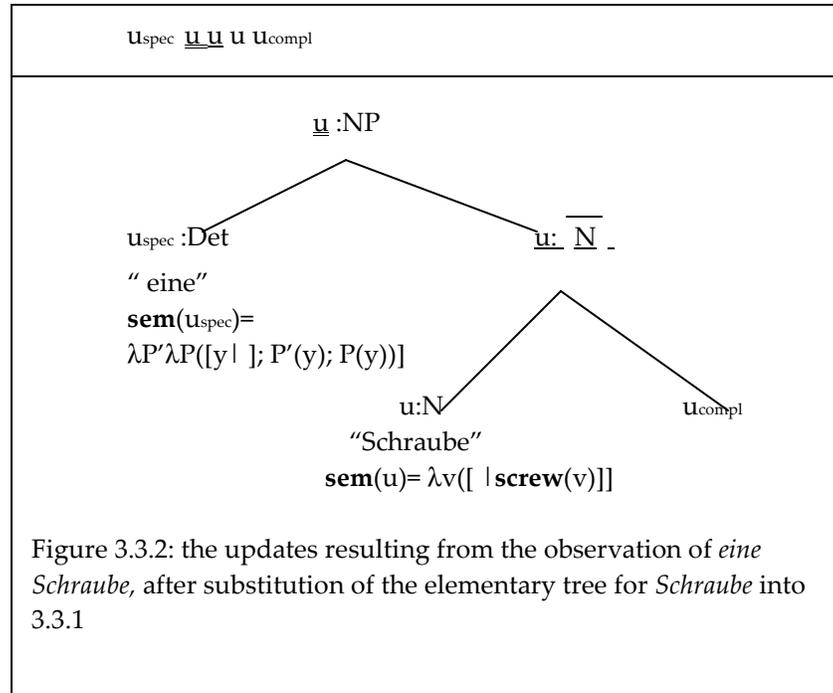


The elementary trees introduced into the discourse situation by the performance of utterances of single words are combined by means of the two basic TAG operations: *substitution* and *adjunction*. Substitution is the operation by which arguments—the ‘expected’ components of the syntactic structure of an utterance, such as the nominal head of the NP expected as a result of the observation of *eine*—are ‘slotted into’ non-atomic elementary trees. For example, assuming an LTAG translation for *Schraube* analogous to that for *boxcar*, observing an utterance of *Schraube*, and subsequently substituting the associated elementary tree into the interpretation in Figure 3.3.1, results in the interpretation in Figure 3.3.2.

Adjunction is the operation by which adjuncts and non-arguments are incorporated into the syntactic interpretation being inferred. In our fragment for the example dialogue, we propose that adjunction is used to incorporate appositions, i.e., to attach *eine orangene, mit einem Schlitz* to *eine Schraube*. We hypothesize for this apposition the syntactic structure in (3.3.4)(b); this structure is attached to the interpretation of NP *eine Schraube* in (3.3.4)(a) by ‘splitting’ the N’ node into two nodes, resulting in the structure in (3.3.4)(c).



TAG in general, and LTAG in particular, have become widely used by psychologists as a syntactic framework in that adjunction provides an explanation for how modifiers can be incrementally attached to a single syntactic interpretation (Sturt and Crocker, 1996). Ferreira and colleagues have argued that LTAG also provides an ideal framework to account for disfluencies (Ferreira et al, 2004). A complete LTAG fragment for the example dialogue is provided in Appendix B; our treatment of appositions is discussed in B.4.2.



The remaining aspect of the interpretation process, semantic composition—the process by which phrasal utterances receive an interpretation—is also viewed in PTT as an inference process (Poesio, 1995; Poesio, To Appear), in accordance with the view adopted in Categorical Grammar and related work on ‘Parsing as inference’ (Pereira, 1990; Carpenter, 1994), where the combination of utterances in larger utterances and the specification of the meaning of these larger utterances are provided by inference rules. PTT hypothesizes that semantic composition, as well, is the result of defeasible inferences over the DRS obtained by concatenating the updates resulting from the utterances of single words (Poesio, To Appear). These default inference rules have the effect of the semantic composition rules introduced by Muskens (1996) for Compositional DRT. For example, the rule **BINARY SEMANTIC COMPOSITION** below specifies that if  $u_1$  and  $u_2$  are the (only) two constituents of  $u_3$ , one of them (say,  $u_1$ ) has a semantic interpretation of type  $\langle \alpha, \beta \rangle$ , and the other has a semantic interpretation of type  $\alpha$ , the semantic interpretation of  $u_3$  is derived by applying the semantic interpretation of  $u_1$  to that of  $u_2$ . (Cfr. Muskens’ **Application** rule (Muskens, 1996, p. 166).) We also assume a **Unary Semantic Composition** inference rule achieving the effect of Muskens’ **Copying** rule.

**BINARY SEMANTIC COMPOSITION (BSC):**

$$\begin{array}{l}
 u_1 \uparrow u_3, u_2 \uparrow u_3, \\
 \text{sem}(u_1) = \varphi_{\langle \alpha, \beta \rangle}, \\
 \frac{\text{sem}(u_2) = \psi_{\alpha}}{\text{sem}(u_3) = \varphi(\psi)_{\beta}}
 \end{array}$$

We can now return to our example dialogue (1) and see how it is interpreted according to PTT, using the LTAG + Compositional DRT grammar from Appendix B together with the micro-conversational events hypothesis. We will deal here with a slightly simplified version of the series of utterances that result in the first **directive**: *Jetzt nimmst Du eine orangene Schraube mit*

*einem Schlitz*, ignoring *so*, merging Cnst's completion (*eine Schraube*) with Inst's refashioning of it via an apposition (*eine orangene ... mit einem Schlitz*) (completions and refashioning are discussed in Section 5; appositions in Appendix B.2). Each of these micro conversational events causes an incremental update of the discourse situation; the concatenated sequence of updates is shown using linear notation in (3.3.5), where we have used  $\hat{\uparrow}$  to indicate dominance, the prefix *up* to indicate phrasal utterances and the prefix *ub* to indicate Xbar utterances.<sup>7</sup> (See Appendix B for graphical representations of these LTAG trees.)

- (3.3.5) [u1.2, up1.1, up1.2 |  
u1.2: **utter**(Inst, "jetzt"), **Advp**(u1.2),  
**sem**(u1.2) is  $\lambda p. \text{now}(p)$ ,  
**S**(up1.1), **S**(up1.2), u1.2  $\hat{\uparrow}$  up1.1, up1.2  $\hat{\uparrow}$  up1.1];
- [u1.3, up1.4, up1.2', up1.3, up1.5, up1.6 |  
u1.3: **utter**(Inst, "nimmst"), **Verb**(u1.3),  
**sem**(u1.3) =  $\lambda Q \lambda x (Q(\lambda x' [e \mid e: \text{grasp}(x, x')]))$ ,  
**VP**(up1.4), u1.3  $\hat{\uparrow}$  up1.4, **S**(up1.2'), up1.4  $\hat{\uparrow}$  up1.2',  
**NP**(up1.3), up1.3  $\hat{\uparrow}$  up1.2', **NP**(up1.5), up1.5  $\hat{\uparrow}$  up1.4,  
**NP**(up1.6), up1.6  $\hat{\uparrow}$  up1.4];
- [u1.4, up1.5' |  
u1.4: **utter**(Inst, "Du"), **Pro**(u1.4), u1.4  $\hat{\uparrow}$  up1.5', **NP**(up1.5'),  
**sem**(u1.4) =  $\lambda \mathcal{P}. \mathcal{P}(\text{you})$ ];
- [u1.5, up1.6', ub1.6, u1.6' |  
u1.5: **utter**(Inst, "eine"), **Det**(u1.5),  
**sem**(u1.5) =  $\lambda P' \lambda P ([y1 \mid ]; P'(y1); P(y1))$ ,  
**NP**(up1.6'), **Nbar**(ub1.6), **Noun**(u1.6'),  
u1.5  $\hat{\uparrow}$  up1.6', u1.6'  $\hat{\uparrow}$  ub1.6, ub1.6  $\hat{\uparrow}$  up1.6'];
- [u1.8, ub1.6', ub1.6'' | u1.8: **utter**(Inst, "orangene"), **AdjP**(u1.8),  
**sem**(u1.8) =  $\lambda P \lambda z ([ \mid \text{orange}(z)]; P(z))$ ,  
u1.8  $\hat{\uparrow}$  ub1.6', ub1.6''  $\hat{\uparrow}$  ub1.6',  
**Nbar**(ub1.6'), **Nbar**(ub1.6'')];
- [u1.6 | u1.6: **utter**(Cnst, "Schraube"), **Noun**(u1.6),  
**sem**(u1.6) =  $\lambda v ([ \mid \text{screw}(v)])$ ];
- [u1.9, ub1.6''', ub1.6'''' , up1.9, up1.10 |  
u1.9: **utter**(Inst, "mit"), **Prep**(u1.9),  
**sem**(u1.9) =  $\lambda \mathcal{P} \lambda y (\mathcal{P}(\lambda x [ \mid \text{with}(x, y)])$ ,  
**PP**(up1.9), **Nbar**(ub1.6'''), **Nbar**(ub1.6''''), **NP**(up1.10),  
u1.9  $\hat{\uparrow}$  up1.9, ub1.6''''  $\hat{\uparrow}$  ub1.6''''', up1.9  $\hat{\uparrow}$  ub1.6''''', up1.10  $\hat{\uparrow}$  up1.9];
- [u1.10, up1.10', ub1.10, u1.11' |  
u1.10: **utter**(Inst, "einem"), **Det**(u1.10),  
**sem**(u1.10) =  $\lambda P' \lambda P ([y2 \mid ]; P'(y2); P(y2))$ ,  
**NP**(up1.10'), **Nbar**(ub1.10), **Noun**(u1.11'),  
u1.10  $\hat{\uparrow}$  up1.10', ub1.10  $\hat{\uparrow}$  up1.10', u1.11'  $\hat{\uparrow}$  ub1.10];
- [u1.11 | u1.11: **utter**(Inst, "Schlitz"), **Noun**(u1.11),  
**sem**(u1.11) =  $\lambda v ([ \mid \text{slit}(v)])$ ]

---

<sup>7</sup> The complete lexical and grammatical rules for the fragment of German we are considering, broadly based on Muskens (1996), are given in Appendix B.

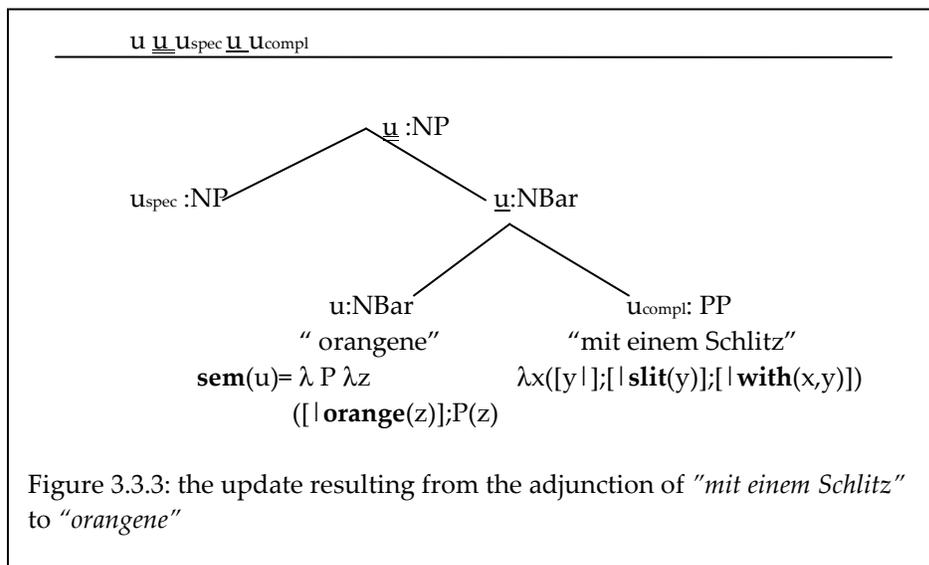
After each of these updates, parsing and semantic composition inference rules apply to combine phrasal utterances together through substitution and adjunction and to assign an interpretation to such phrasal utterances. Let us consider for instance how the updates due to the observation of “*einem*” (micro conversational event *u1.10* in 3.3.5) and “*Schlitz*” (mce *u1.11*) result in the hypothesis that an utterance of the NP “*einem Schlitz*” was observed, and in the assignment of an interpretation to that phrasal utterance. The hypothesis that the utterance of *u1.10* and *u1.11* in (3.3.5) is part of the utterance of an NP leads to a substitution inference: the hypothesis that *u1.11'* (expected after observing an utterance of *einem*) is the same as *u1.11*. The formulation of this hypothesis results in the following update of the discourse situation:

[ | *u1.11'* is *u1.11*]

Applying BSC to the discourse situation thus updated results in the assignment of the conventional meaning  $\lambda P.[y2 | ];[ | \text{slit}(y2)]; P(y2)$  to *up1.10'*, which results in the following update:

[ | **sem**(*up1.10'*)=  $\lambda P.[y2 | ];[ | \text{slit}(y2)]; P(y2)$ ]

A similar process leads to hypothesizing the rest of the structure; the only difference is that attaching *orangene* and *mit einem Schlitz* requires adjoining the PP resulting from processing *mit einem Schlitz* into the NBar resulting from *eine orangene*: the result can be seen in Figure 3.3.3.



At the end of this process of unification of the phrasal utterances, we obtain the structure shown in Figure 3.3.4, from which meanings have been omitted (but see Appendix B and Section 5).

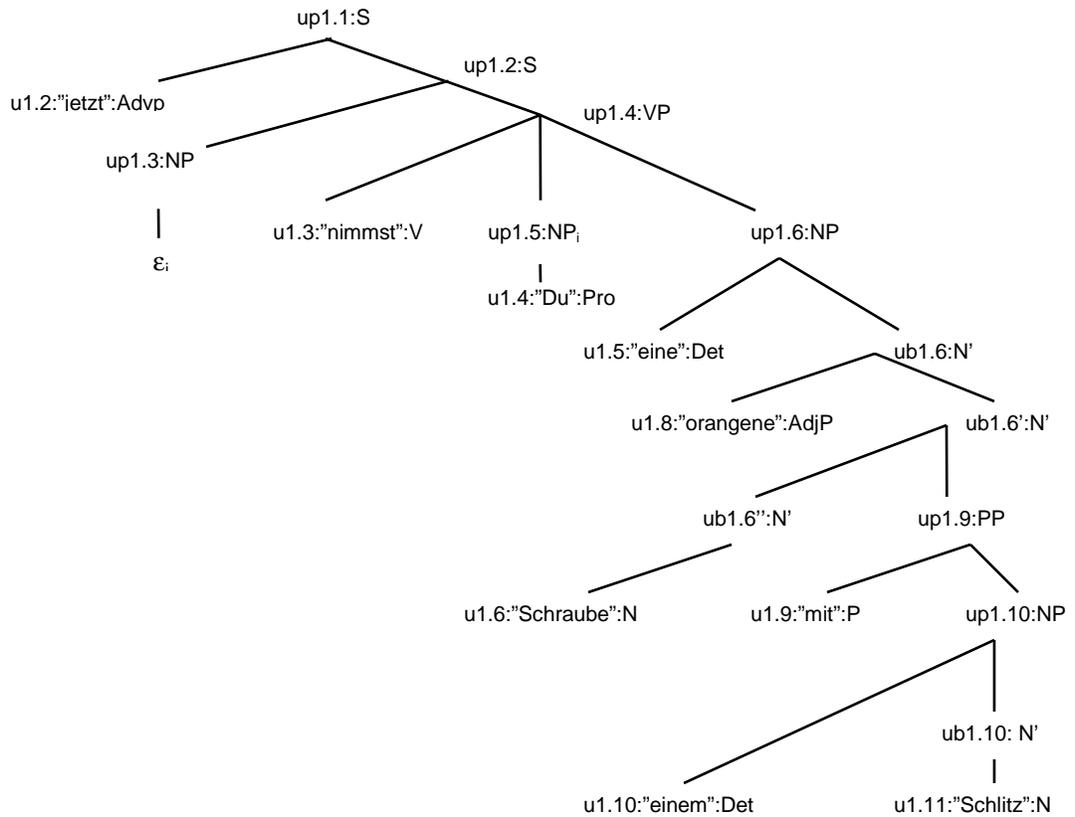


Figure 3.3.4 – simplified representation of the syntactic structure assumed for the first contribution

Inference processes in PTT typically result in augmenting a DRS with new material. We will use the notation  $K += K'$  for indicating DRS update, defined as follows:

**$K += K'$**  Let  $K$  be a proposition-valued discourse referent, and  $K'$  be a proposition (DRS) or a proposition-valued discourse referent. Then

$$K += K' =_{\text{def}} [ \text{Tmp}K \mid \text{Tmp}K \text{ is } K ]; [ \mid K \text{ is } \text{Tmp}K; K' ]$$

(where  $\text{Tmp}K$  is an unused proposition-valued discourse referent)

### 3.4 Grounding and Discourse Units

Unlike other dynamic theories of interpretation, PTT does not rely on the assumption that every utterance automatically becomes part of the common ground; instead, it includes an explicit formalization of the GROUNDING process discussed in Section 2 (Clark and Schaefer, 1989; Brennan, 1991; Traum and Hinkelman, 1992; Traum, 1994; Clark, 1996). Many of the utterances in the example dialogue were most likely intended to play a role in this process; and completions themselves can be viewed as a particularly explicit form of acknowledgment, as discussed in Section 2.

The formalization of grounding developed in PTT is based on Clark and Schaefer's (1989) proposal discussed in Section 2, as modified by Traum (1994). According to Clark and Schaefer, a

conversation consists of a series of CONTRIBUTIONS which have to be ACKNOWLEDGED, possibly implicitly (thus becoming part of the common ground) or may need further CLARIFICATIONS and REPAIRS.

Traum (1994) and Matheson *et al* (2000) developed this proposal by providing an account based on the assumption that grounding is achieved through a particular type of dialogue control acts called GROUNDING ACTS. Following (Traum, 1994), we use the term DISCOURSE UNIT (DU) to refer to a contribution. And again as in (Traum, 1994), it is assumed in PTT that every utterance in a conversation either **initiates** a new DU, **continues** a DU, **acknowledges** a DU, performs a **repair**, or **requests** the other participant to perform one of the grounding acts above. At any point in a conversation a conversational participant may begin a new contribution, i.e., initiate a new DU; this new DU gets added to the semi-public part of the information state. This is the case for example with 1.1 or 2.1 in the example dialogue. As we will discuss in detail in Section 5, in 1.2 Cnst simultaneously acknowledges the part of the new contribution which has already been introduced, grounding it—possibly in response to a perceived request from Inst—and adds new material to the contribution. In 1.3, Inst performs what Clark and Wilkes-Gibbs called a **refashioning** of the contribution—adding new material which may also lead to a revision (e.g., to the choice of a new screw). (In PTT, this type of operation is viewed as a type of **Repair**, as proposed by Levelt (1989) and Clark (1996). In fact, the introduction of new material in 1.2 is seen as a **Repair** as well, as discussed in Section 5.) Finally, Cnst **acknowledges** the remaining part of the contribution, which then is fully grounded, and accepts the **directive** specified by the full contribution.<sup>8</sup> The **accept** itself is viewed in PTT as a second contribution. In 2.1, Inst (implicitly) grounds the **accept** by initiating a new contribution (a new directive).

As in (Poesio and Traum, 1997; Matheson, Poesio, and Traum, 2000), we assume that discourse units are dynamic propositions about the discourse situation, i.e., DRSs containing the type of information about conversational events (whether micro conversational events, core speech acts, and other types of dialogue acts) that we have discussed in the rest of this section. In fact, in recent work we adopted the position that each of the micro-conversational update to a discourse situation we discussed earlier in this section constitutes a discourse unit, but here we will stick with the position adopted in (Poesio and Traum, 1997; Poesio and Traum, 1998), in which all the updates of the discourse situation related to the process of grounding a particular contribution are part of the same discourse unit.

In (Poesio and Traum, 1997, 1998) the agent' information state itself was formalized as a DRS. This DRS includes, in addition to information about the private attitudes of the agent, a record of the contributions (DUs) so far, as well as a special DRS G containing the material which has already been grounded. For example, the information state after interpreting and grounding the first contribution in (3.1.1), and after interpreting the second sentence (i.e., creating a DU DU2 for it) but before grounding it, would be as follows (ignoring micro-conversational events): (3.4.1) [DU1,DU2|

```

DU1 is [ce1, K1|
    u1: utter(A,"there is an engine at Avon"),
    K1 is [x,w,s| engine(x), Avon(w), s: at(x,w)],
    sem(u1) is K1,
    ce1: assert(A,B,K1),
    generate(u1,ce1)],

```

---

<sup>8</sup> Notice that in PTT **acknowledgments**—a type of grounding acts—and **accepts**—a backward act resulting in the speaker's assuming the obligation to perform a certain action—are distinct.

G is [ce1, K1 |  
 u1: **utter**(A, "there is an engine at Avon"),  
 K1 is [x,w,s | **engine**(x), **Avon**(w), s: **at**(x,w)],  
**sem**(u1) is K1,  
 ce1: **assert**(A,B,K1),  
**generate**(u1,ce1)],  
 DU2 is [ce2,K2 |  
 u2: **utter**(B, "it is hooked to a boxcar"),  
 K2 is [y,u,s' | **boxcar**(y), s':**hooked-to**(u,y), u is x],  
**sem**(u2) is K2,  
 ce2: **assert**(B,A,K2),  
**generate**(u2,ce2)]]],

There are two main reasons for viewing the information state as a DRS. First of all, all grounding acts are implicitly anaphoric, in that they refer to particular DUs, as we will see below. Secondly, in Compositional DRT the modifications to G and the DUs resulting from grounding acts can be modeled quite simply as updates to the values of discourse markers. For example, the effect of an **acknowledgment** on the information state was formalized by Poesio and Traum (1998) as replacing the previous value of G with a new DRS which is the merge of G and DU1 using the += operator just introduced:

(3.4.2)  $G += DU1$

In this paper, however, we will use a modal operator G to assert that a particular DU is grounded, for a number of reasons, the main one being that this formulation is closer to that adopted in more recent work by Traum (1999) in which grounding is not viewed as an all-or-nothing affair, but as a matter of degree. This type of theory is more easily modeled using one or more modal operators to specify grounding. The modal operator G expresses a stronger form of mutual knowledge, in that everything that is grounded (acknowledged) is mutually believed, but not viceversa. We will not provide a full axiomatization of G here, but we will assume the following:

[ **AX-G-1**]  $\forall DU G(DU) \rightarrow MK(DU)$

According to this new view, the information state resulting from the first contribution in (3.1.1) being grounded, while the second one still isn't, is as in (3.4.3), instead of as in (3.4.1):

(3.4.3) [DU1,DU2 |  
 DU1 is [ce1, K1 |  
 u1: **utter**(A, "there is an engine at Avon"),  
 K1 is [x,w,s | **engine**(x), **Avon**(w), s: **at**(x,w)],  
**sem**(u1) is K1,  
 ce1: **assert**(A,B,K1),  
**generate**(u1,ce1)],  
 G(DU1),  
 DU2 is [ce2,K2 |  
 u2: **utter**(B, "it is hooked to a boxcar"),  
 K2 is [y,u,s' | **boxcar**(y), s':**hooked-to**(u,y), u is x],  
**sem**(u2) is K2,  
 ce2: **assert**(B,A,K2),  
**generate**(u2,ce2)]]]

According to this version of the theory, in the case of the example dialogue we obtain the following information state after the first contribution (the directive jointly produced by Inst and Cnst in 1.1 – 1.3) and the second (the acceptance produced by Cnst in 1.4) are grounded, but before the second directive is grounded, again ignoring micro-conversational events and phrasal utterances (compare with (3.1.6)):

(3.4.4) [DU1.1, ce1.6, DU1.4, ce2.2, DU2.1 |  
 DU1 is [K1.1, up1.1, ce1.1, |  
 K1.1 is [e,x,x3 | **screw(x), orange(x), slit(x3), has(x,x3), e:grasp(Cnst, x)],**  
**utterance(up1.1), sem(up1.1) is K1.1,**  
 ce1.1:**directive(Inst&Cnst,Cnst,K1.1),**  
**generate(up1.1, ce1.1)]**  
 ce1.6 :**ack(Cnst,DU1.1),**  
 G(DU1.1),  
 DU1.4 is [ce1.7, s1.1 | ce1.7: **accept(Cnst,ce1.1), s1.1 : obl(Cnst,K1.1) ],**  
 ce2.2 :**ack(Inst,DU1.4),**  
 G(DU1.4),  
 DU2.1 is [K2.1, up2.1, ce2.1 |  
 K2.1 is [x6,e',s,w,y | x6 is x, e':**put-through(Cnst,x6,hole1),**  
 w is wing1, y is fuselage1,  
 s: **fastened(w,y), purpose(e',s)],**  
**utterance(up2.1), sem(up2.1) is K2.1,**  
 ce2.1:**directive(Inst,Cnst, K2.1 ) ,**  
**generate(up2.1, ce2.1)]**

Notice that the grounding acts ce1.6 and ce2.2 are *not* included in the discourse units. This is the solution proposed by Traum (1994) to the 'bottoming out' problem present in Clark and Schaeffer's work – explaining how information about the occurrence of grounding acts is unlike other information in the discourse situation in that it doesn't seem to require grounding (else we would have an infinite regress).

Notice also that in all accounts of grounding derived from Clark and Schaeffer's work, planning a core speech act really amounts to planning a contribution, i.e., opening a new DU. Hence such theories need to stipulate an additional step of intentional reasoning codified by the following schema:

$$\text{Int}_{cp1}([\text{ce} \mid \text{ce:core-act}(cp1,cp2,\Phi)]) \rightarrow \text{Int}_{cp1}([\text{DU} \mid \text{DU is} [\text{ce} \mid \text{ce:core-act}(cp1,cp2,\Phi)]])$$

We will see examples of application of this schema in Section 5.

Grounding acts can be simply formalized in terms of operations on the information state. In the version of PTT adopted here,

- i. Performing an **Init**(DU) simply means introducing a new DU in the information state. We assume an **Init**(DU) every time a new contribution is initiated; such grounding acts are not explicitly recorded in the information state.
- ii. Performing a **Cont**(DU) means adding to an existing DU; again, we do not explicitly represent this grounding act in the information state.
- iii. **Acknowledging** a DU has the effect of grounding the DU, i.e., adding G(DU) to the information state;
- iv. **Repairing** DU1 with DU2 means updating the discourse referent DU1 by assigning to it DU2 as value.

- v. **Requesting** a grounding act means adding to the information state an obligation to address that request.

This formulation does not specify preconditions for such acts—e.g., what it means for a contribution to be understood, and therefore when are acknowledgments warranted (Clark’s **grounding criterion**). Clark points out that a contribution may fail to be understood at several levels: e.g., the addressee may have not heard what the speaker said, or may have heard it but not know its meaning. In this paper we assume that an utterance has been understood when values for all the functions that specify its linguistic classifications –phonetic, syntactic, and semantic (**sem**)—can be recovered either from the lexicon or from the context; but we will not provide update rules for grounding acts including such preconditions. (Ginzburg’s (2008) proposals for information state update rules for grounding acts specifying such information and covering clarification requests as well are discussed in Section 7.)

Matheson, Poesio and Traum (2000) discuss **accepting** in PTT for the case of core speech acts. Update rules for core speech acts are conditional upon acceptance—meaning that for each core speech act there is a *conditional* update rule specifying the part of the update caused by that speech act that depends upon acceptance. For instance, for the case of **directives**, the conditional update rule specifies that in case *ce* is a directive by A to B with content K, then if B accepts *ce*, B assumes the obligation to bring it about that K: schematically,

$$\text{UP.DIRECTIVE } ce:\text{directive}(A,B,K) \rightarrow (\text{accept}(B,ce) \rightarrow [o \mid o:\text{obl}(B,K)])$$

(Notice that an update rule is not a material implication, but a rule to update the information state by concatenating new information.) We assume here that update of the information state as the result of grounding acts works in the same way, except that in the case of grounding acts, accepting leads to updates of the information state that affect DUs. We use two such rules in this paper:

$$\text{UP.REPAIR } ce:\text{repair}(A,DU1,DU2) \rightarrow (\text{accept}(B,ce) \rightarrow [ \mid DU1 \text{ is } DU2])$$

(with DU1 a propositional discourse referent)

$$\text{UP.ACK } ce:\text{ack}(A,DU) \rightarrow (\text{accept}(B,ce) \rightarrow [ \mid G(DU)])$$

(with DU a propositional discourse referent)

So, for instance, the result of the *Ja* in 1.4 is to add to the information state  $G(DU1.1)$  when the acknowledgment is accepted (see (3.3.4)), whereas the result of accepting the repairs in 1.2 and 1.3 is to add to  $DU1.1$  the new material, as discussed in Section 5.

### 3.5 The Information State

As already mentioned, PTT is an information-state based theory in the sense of (Cooper et al 1999, Larsson & Traum 2000, Stone 2004, Ginzburg 2008). Ginzburg (2008, Chapter 4) discusses different views concerning what information states such a theory may model; PTT aims at modeling the information state of a single agent, assumed to consist of three main parts:

- A private part, with information available to the participant, but not introduced in the dialogue. This part includes private beliefs and intentions of that participant, as well as hypotheses about beliefs and intentions of other agents.
- A public part consisting of the DUs that are assumed by that agent to have become part of the common ground.

- A semi-public part, consisting of the information introduced with contributions that haven't yet been acknowledged. This information is not yet grounded, but it is accessible.

So far, we have discussed the information about actions in the discourse situation. To conclude we briefly discuss information about private attitudes and social attitudes (obligations).

In PTT, dialogue acts are performed to achieve **intentions** or to satisfy certain **obligations**. Both the fact that one or more agents have a certain (possibly collective) intention, and that they are under certain obligations, may become part of the private, semi-private and public parts of the discourse situation, e.g., as a result of planning or of intention recognition (Matheson, Poesio, and Traum, 2000).

In previous work, only a partial formalization of obligations and intentions was given. In this earlier work, both obligations and intentions were viewed as relations between agents and action types: for example, the fact that agent A has the intention to perform a particular core speech act is captured by the presence in the discourse situation of intention (3.5.1).

(3.5.1) i: **intend**(A,  $\lambda ce. ce: \mathbf{assert}(A, B, K1)$ )

We adopt here a more standard view of intentions and obligations as predicates indexed by the agent(s) holding their intention. We will also simplify matters concerning the dynamics of conversational events by assuming that intentions and obligations have as their contents propositions describing the state of affairs to be achieved –i.e., DRSs—rather than action types, as in (3.5.2), showing the intention by agent A to perform an assertion with content K1.

(3.5.2) i: **Int**<sub>A</sub>([ce | ce: **assert**(A, B, K1)])

A partial formalization of obligations was provided in (Matheson, Poesio ad Traum, 2000). As far as intentions are concerned, previous work only discussed the assumption, inherited from Grosz and Sidner (1986), that intentions may be related to each other in two ways: by a relation of **dominance** when satisfying a certain intention is part of the satisfaction of another, more complex, intention (more formally: intention *i* dominates intention *i'* if achieving *i'* is part of achieving *i*), and by a relation called **satisfaction-precedes** when satisfying an intention is a prerequisite for satisfying the second (intention *i* **satisfaction-precedes** intention *i'* if achieving *i* is a necessary prerequisite of achieving *i'*). As each intention is (directly) dominated by only one other intention, Poesio and Traum (1997) formalized dominance with a (partial) function **dom** mapping intention *i2* to the intention to which it is subordinated, if any. More controversially, satisfaction-precedence was also formalized as a partial function **sp** mapping intention *i2* to the intention *i1* that must be achieved for *i2* to be achievable.

$\mathbf{sp}(i2) = i1$

No axioms for intentions or intentional structure were proposed by Poesio and Traum; one of the goals of this paper is to use completions and continuations as a source of additional evidence on the properties of intentions, and how they affect interpretation; we discuss a few possibilities and our assumptions in Sections 4 and 5.

## 4 Completions: an Account Based on Shared Plans

In this paper we will propose two analyses of the example dialogue. We begin by presenting in this section and the next a mainstream, ‘intentional’ analysis based on hypotheses about the role of intentions and cooperation in communication developed in artificial intelligence, linguistics, philosophy, and psychology over the past thirty years. In Section 6 we will then discuss a second analysis, based on the recent proposals by Garrod and Pickering on the basis of recent psychological results about interpretation, production and alignment in dialogue (Pickering and Garrod, 2004).

### 4.1 Coordination, shared intentions, partial shared plans: a look at existing paradigms

#### 4.1.1 Assumptions: intention, cooperation, coordination, discourse plans

Most work on dialogue in Artificial Intelligence, Philosophy and Psychology in the last thirty or so years has been based on three main assumptions.

The first assumption is that communication involves a great deal of **intention recognition**: the production of utterances is motivated by (implicit and explicit) intentions, and in order to communicate felicitously agents must recognize other agents’ intentions even when they do not intend to help these agents to achieve them. This hypothesis, generally associated with Grice (Grice 1969, Grice 1975, later revised in Grice 1991), has been the foundation of most theories of dialogue in Artificial Intelligence and Computational Linguistics, such as the work of Allen and Perrault (1980), Cohen and Levesque (1990a, 1990b), Grosz and colleagues (Grosz and Sidner, 1986; Grosz and Kraus, 1996) Sadek (Sadek et al, 1994) Asher and Lascarides (2003) and Stone (2004) among others; in Philosophy—e.g., in work by Bratman (1992) or Tuomela (2000); and in Psychology (Clark, 1992, 1996; Bara & Tirassa, 2000)—and was thoroughly examined in the seminal *Intentions in Communication* (Cohen, Morgan and Pollack, 1990).<sup>9</sup> The theories of intentions developed in AI are typically also concerned with how such intentions can be achieved: hence many such theories, and particularly those of intention recognition, are formulated in terms of **plans** to achieve a particular goal. The intentional account we propose will be formulated in this way, as well.

This hypothesis that intention recognition is central to communication is usually supplemented by two further hypotheses. Much work on dialogue in Artificial Intelligence relies on the further assumption that at least in some contexts, communicating agents do not simply recognize other agents’ intentions; they are also **cooperative**, in the sense that they attempt to help other agents’ achieve their intentions even when they are not explicitly expressed. E.g., a genuinely helpful clerk at a ticket counter in a train station will not simply answer the question “From which platform does the train to Montreal leave” by saying the platform, if he / she knows that the train has been cancelled. This view is central both to Herbert Clark’s theory of dialogue discussed in Section 2 and to, e.g., Allen’s and Sadek’s theories of intention recognition in dialogue systems (Allen and Perrault, 1980; Sadek, 1992). Completions and continuations are viewed by, e.g., Clark as some of the best evidence for cooperative behavior in dialogue (Clark, 1996, p. 238).

---

<sup>9</sup> In recent years, the role of intentions in communication has become of interest to developmental psychologists (Homer & Tamis-Le Monda, 2005) and neural scientists (Bara & Tirassa, 2000), as well.

As discussed in Section 2, the theory of dialogue developed by Herbert Clark and summarized in (Clark, 1992, 1996) is based in addition on a third hypothesis: that conversation is a form of **joint activity** just like playing football or playing in an orchestra, i.e., driven by **joint intentions** and in which it is necessary for agents to **coordinate**. Clark uses “joint activities” as a foundational stratum on which coordination on content or process, on setting up common ground, on signalling, establishing (complex) joint projects, on communication using parallel tracks etc. is then based. Clark defines **joint project** quite simply as “... a joint action, projected by one of its participants and taken up by the others” (Clark (1996), p. 191). The notion of joint project or shared plan has been further developed in the Artificial Intelligence, especially by Grosz and Sidner (1990) and by Grosz and Kraus (1996).

Taken together, these three assumptions lead to the view on which our first analysis is based: that what happens in dialogue, particularly task-oriented dialogue, can be explained in terms of the joint intentions of dialogue participants and the shared plans developed to achieve them. In the case of the construction dialogues in the BTPC, this view can be summarized as follows. Inst and Cnst have both shared and private domain plans. Inst and Cnst’s private domain plans overlap to a degree with the shared plan; the difference between the shared plan and the private plans may lead to discrepancies and negotiations (e.g., as when Inst adds more details to Cnst’s proposal in 1.3). Simplifying things drastically, we will assume here that Inst’s private domain plan is a fully specified plan for building the toy airplane (either instructions or a model). The shared domain plan is a partial plan to build a toy airplane, which at the beginning of a dialogue is virtually empty, but gets progressively refined through the construction dialogue. (More on this below.) Cnst’s private plan is a refinement of the shared domain plan, likely to include at least local further specifications based on expectations.

Crucially, dialogue involves shared plans both at the domain level (how to build a toy airplane) and at the discourse level (how to convey a particular intention or plan) (Litman and Allen, 1990). For example, Inst and Cnst also share a discourse plan: that the conversation will consist of a series of instructions by Inst to Cnst aiming at building a toy airplane according to Inst’s model. This view is clearly formulated, e.g., in the following citation from Grosz and Sidner ((1990), p. 418):

*Discourses may exhibit two types of collaborative behaviour: collaboration in the domain of discourse [...] and collaboration with respect to the discourse itself. Although we cannot yet define [...] “collaboration with respect to a discourse”, it includes not only surface collaboration (such as coordinating turns in a dialogue) or use of appropriate referring expressions ([...]) but also collaborations related to the discourse purpose. For example, the participants collaborate to ensure that the utterances of the discourse itself provide sufficient information to make possible the satisfaction of the discourse purpose.*

In terms of shared plans, Cnst’s completion in 1.2 can be interpreted as indicating that Cnst has recognized Inst’s intentions both at the domain level and at the discourse level, and one possible explanation (although not the only one, as we will see below) for her decision to complete is that she is being cooperative. Our first, ‘intentional’ account of Completions and Continuations is based on this Shared Plans hypothesis. We follow here what is perhaps the best-known development of this idea, due to Bratman in his work on SHARED COOPERATIVE ACTIVITY (SCA) (Bratman, 1992) and further refined by Tuomela (2000). These two theories are briefly introduced next.

#### 4.1.2 Bratman

Arguably the most influential modern account of intentions (private and shared) is that proposed by Bratman (1992). Bratman's theory aims at providing a formalization of the notion of "shared cooperative activity" (SCA, p. 327) that underlies actions characterised as "being done together" like singing a duet, painting a house or starting an attack in a basketball game — i.e., joint actions in Clark's sense, i.e., most actions performed in dialogue, including completions.

Bratman identifies three main characteristics of SCAs: *mutual responsiveness*, *commitment to the joint activity* and *commitment to mutual support*. The conditions that according to Bratman must hold in order for us to be said to perform a shared cooperative activity (SCA) J are listed in (4.1.1). First of all, we need individual intentions to J together: (1)(a)(i) and (1)(b)(i). Secondly, it must be the case that you and I develop these intentions because of the other's intentions ((1)(a)(ii), (1)(b)(ii)). Third, (1)(c) is intended to exclude forced cooperation, e.g. the "Mafia sense" (1992, p. 333) of doing something together. Fourth, the intentions must hold for the matching subplans involved and they must be stable. (An intention is *minimally cooperatively stable* if "there are cooperatively relevant circumstances in which the agent would retain that intention" (p. 338). This notion is similar to Cohen and Levesque's notion of 'commitment'.) Finally, we must mutually know (1). Mutual knowledge is used in the "fixed point" sense (Fagin et al. , 1995, p. 402; Tuomela 2000, p. 78).

- (4.1.1) Our J-ing is a SCA only if
- (1)(a) (i) I intend that we J.
    - (ii) I intend that we J in accordance with and because of meshing subplans of (1)(a)(i) and (1)(b)(i).
  - (b) (i) You intend that we J.
    - (ii) You intend that we J in accordance with and because of meshing subplans of (1)(a)(i) and (1)(b)(i).
  - (c) The intentions in (1)(a) and in (1)(b) are not coerced by the other participant.
  - (d) The intentions in (1)(a) and (1)(b) are minimally cooperatively stable.
- (2) It is common knowledge between us that (1).

This first version of Bratman's characterization of an SCA already captures some key elements of the 'dialogue as a joint action' view discussed in Sections 2 and 4.1.1: *commitment to joint activity*, *meshing subplans*, *interdependent intentions*, *mutual support*, and *connecting attitudes*. Parts (ii) of (4.1.1)(1)(a) and (4.1.1)(1)(b) mean that it is not sufficient for an action to be independently intended to be a joint action by me and you for it to be a SCA; these intentions have to be causally connected. Parts (ii) also make explicit reference to the shared plans required to achieve a joint intention, via the notion of *meshing subplans*, described by Bratman as follows: "[...] our individual sub-plans concerning J-ing *mesh* just in case there is some way we could J that would not violate either of our subplans" (p. 332). (As we will see, Inst and Cnst's private plans are meshing in this sense.)

According to Bratman, however, one ingredient is still missing from (4.1.1): *mutual responsiveness*. Mutual responsiveness amounts to the following: "In an SCA each participating agent attempts to be responsive to the intentions and actions of the other, knowing that the other is attempting to be similarly responsive" (p. 328). (As we will see below, mutual responsiveness, or more precisely, its development by Tuomela into the hypothesis that participants in a SCA may perform 'unrequited contributory actions,' provides us with one explanation for the performance of completions.) The full definition of SCA is therefore as in (4.1.2).

- (4.1.2) For cooperatively neutral J, our J-ing is a SCA if and only if
- (a) we J
  - (b) we have the attitudes specified in (4.1.1) (1) and (2), and
  - (c) (b) leads to (a) by way of mutual responsiveness (in the pursuit of our J-ing) of intention and in action.

Given Bratman's notion of SCA, an intentional account of completions in the example dialogue would go as follows. The relevant J-ing at the stage of example dialogue (1.1) just preceding the completion in 1.2 is that Inst and Cnst we-intend that Cnst fix the 'wing' to the 'fuselage' according to Inst's directives. The intended and mutually known meshing sub-plans are the domain plan for building the toy plane in front of Inst, and the discourse plan that Inst give the directives and Cnst carry out the actions. (Both plans are discussed in more detail in Section 4.2.) In order then for the action of joining wing and fuselage to count as a SCA, the following must hold:

- (4.1.3) Inst and Cnst joining wing and fuselage is a SCA only if:
- (1)(a)(i) Inst intend that Inst and Cnst join wing and fuselage.
    - (ii) Inst intend that Inst&Cnst join wing and fuselage in accordance with and because of meshing subplans of (1)(a)(i) and (1)(b)(i).
  - (b)(i) Cnst intend that Inst and Cnst join wing and fuselage .
    - (ii) Cnst intend that Inst and Cnst join wing and fuselage in accordance with and because of meshing subplans of (1)(a)(i) and (1)(b)(i).
  - (c) The intentions in (1)(a) and in (1)(b) are not coerced by the other participant.
  - (d) The intentions in (1)(a) and (1)(b) are minimally cooperatively stable.
  - (2) It is common knowledge between Inst & Cnst that (1).

The necessary and sufficient conditions for the action to be an SCA are specified by adding the mutual responsiveness requirement from (4.1.3), obtaining (4.1.4):

- (4.1.4) For cooperatively neutral joining of wing and fuselage, Inst&Cnst joining wing and fuselage is a SCA if and only if:
- (a) Inst&Cnst join wing and fuselage
  - (b) Inst and Cnst have the attitudes specified in (1) and (2), and
  - (c) (b) leads to (a) by way of mutual responsiveness (in the pursuit of Inst & Cnst's joining of wing&fuselage) of intention and in action.

We'll discuss the meshing subplans at stake in some detail in Section 4.2.

### 4.1.3 Tuomela

Bratman's (1992) concept of shared cooperative activity is not formalized in terms of a logic, but several such formalizations have appeared, most famously by Cohen and Levesque (1990a, 1990b). Our own use of **we-intention** ( $Int_{Inst\&Cnst}$ ) in the rest of the paper will be based on the logical reconstruction of Bratman's theory by Tuomela (2000). Tuomela also proposes several revisions of Bratman's theory, that we briefly review here.

One problem with Bratman's account of plan-based joint action identified by Tuomela is that the meshing requirement should not be part of the definition of **we-intention**. Rather, we should take it as an entailed conceptual presupposition along the lines of 'If we have a joint intention to see to it that p, then as a rule we also have meshing subplans concerning p'. In addition, according to Tuomela, condition (4.1.1)(1)(c) is too strong: coercion should be admitted

as long as a subject's intentional agency is not completely by-passed. Tuomela proposes, therefore, a stripped-down characterization of the **we-intention** going into the construction of the joint of wing and fuselage at the point in which the first completion takes place that can be characterized in first approximation as follows.

(4.1.5) Inst and Cnst **we-intend** that Cnst join wing and fuselage is equivalent to:

It is Inst's and Cnst's mutual knowledge<sup>10</sup> that  
 Inst intends that Cnst join wing&fuselage because  
 Cnst intends that Cnst join wing&fuselage;  
 and  
 Cnst intends that Cnst join wing&fuselage because  
 Inst intends that Cnst join wing&fuselage

The two clauses cover much of clauses (1a) and (1b) of in (4.1.3), the 'because' connective providing the 'meshing' required by Bratman, but without explicit mention of plans and of coercion. In Tuomela's notation (4.1.5) is expressed as follows, where 'because' gets translated by the **reason** relation  $/r$ , which is factual in the sense that  $X/rY \rightarrow X \& Y$ :

(4.1.6)  $\text{Int}_{\text{Inst}\&\text{Cnst}}(\text{join}(\text{Cnst}, W, F))$   
 $\leftrightarrow$   
 $\text{MK}((\text{Int}_{\text{Inst}}(\text{join}(\text{Cnst}, W, F)) /r \text{Int}_{\text{Cnst}}(\text{join}(\text{Cnst}, W, F)))$   
 and  
 $(\text{Int}_{\text{Cnst}}(\text{join}(\text{Cnst}, W, F)) /r \text{Int}_{\text{Inst}}(\text{join}(\text{Cnst}, W, F))))$

According to (4.1.6) the we-intention  $\text{Int}_{\text{Inst}\&\text{Cnst}}$  that Cnst join  $W$  and  $F$  amounts to mutual knowledge that each agent's intention that Cnst join  $W\&F$  is caused by the other agent's intention that this be done.

We said above that agents' disposition to helping, of which at least some forms of cooperative completions are an illustration, is meant to be covered in Bratman's proposal by the notion of mutual responsiveness which, however, is not fully developed, as pointed out by Bratman himself. Tuomela (2000, p. 107) provides a more explicit explanation of the fact that Inst gets extra help by Cnst in formulating a directive for which there can't be a SCA yet. To do this, Tuomela developed an alternative to Bratman's definition of  $\text{Int}_{\text{Inst}\&\text{Cnst}}$ . This new definition cannot be fully discussed here as it is extremely complex and requires several pages. Instead, we will simply highlight the following condition from the definition (2000, p. 95):

“[...] the participants are also assumed to be disposed willingly to perform *unrequired contributory actions* [our italics], thus being disposed to incur extra costs (this being rational as long as the costs of performing them are less than the gross gains accruing from their performance)”.

Tuomela's notion of unrequired contributory action underlies the 'intentional' account of completions given here. Briefly, the idea is that if Cnst and Inst have **we-intentions** and the associated shared plans, Cnst can infer that a screw is needed at the point of the dialogue we are analyzing. She is therefore able to produce an unrequired contributory action by producing a completion, if circumstances demand it. (We will return on this shortly.) Tuomela's formalization of helping is of course related to the traditional notion of cooperativeness, as formalized, e.g., in

---

<sup>10</sup> In many contexts mutual knowledge will be too strong. We will however stick to mutual knowledge for the purposes of this paper.

Cohen and Levesque's 'Cooperative' axiom (Cohen and Levesque, 1990b) or in Sadek's theory (Sadek, 1992). The distinctive aspect of Tuomela's notion of help is that it is embedded into joint intentions and actions. Tuomela argues that participants acting on the basis of a cooperative attitude must be disposed to "strong" helping (p. 249).<sup>11</sup> Helping in the "full" sense means 'helping in all circumstances in which help is contributive to the others' part-performances'. It is exactly reliance on part-performances (making up a SCA eventually) which distinguishes Tuomela's notion of help from Cohen and Levesque's (1990b) and Sadek's (1992), which capture taking over of some agent's intention.

In the rest of the paper, we will use Tuomela's  $\text{Int}_{\text{Inst}\&\text{Cnst}}$  predicate to express we-intentions, adopting the formalization in (4.1.6).

---

<sup>11</sup> Tuomela distinguishes between 'helping in the weak sense' and 'helping in the strong sense'. Help in the weak sense means that agent A is responsive to agent B's performance of her part of the SCA, in order to successfully complete her own part; it is therefore a form of alignment. Helping in the strong sense is supporting the other agent by performing part of their performance of an SCA in an essential way. 1.2 and 2.2 in our example are examples of strong helping, whereas Cnst's acceptances are examples of helping in the weak sense.

## 4.2 An intentional analysis of the example dialogue: first pass

In order to maintain the complexity of the presentation manageable, we divided the presentation of our intentional analysis of the example dialogue between this section and the next. In this section we concentrate on we-intentions and shared plans, showing how the Bratman / Tuomela framework just introduced can explain how completions are produced, without however discussing how completions can be incrementally produced and interpreted, and how the common ground gets updated as a result. In fact, we abstract away from the details of the logical formalism and from DRT and adopt a vanilla logical form. In the next Section we present a more complete account in which these issues are considered as well, using the tools from PTT.

As said above, our intentional analysis is based on the assumption that at the beginning of the dialogue Instructor and Constructor have a **we-intention** –i.e., an SCA in the sense of Bratman and Tuomela—that Constructor assemble a toy airplane identical to the one that the Instructor has. (As discussed in Section 3, the argument of the we-intention is a DRS, i.e., a proposition, but we adopt a simplified representation in this subsection.)

### (4.2.1) $\text{Int}_{\text{Inst}\&\text{Cnst}}(\text{toy-airplane}(x) \wedge \text{assemble}(\text{Cnst}, x))$

The Instructor (henceforth, Inst) has a complete<sup>12</sup> plan for assembling the toy airplane, shown in Figure 4.2.1. The dialogue is driven by the goal of making this a shared plan by discussing it with the Constructor (henceforth: Cnst), so that Cnst can then execute the relevant actions. (For the purposes of this paper, we'll assume that the case in which Inst receives no instructions, just a completed model of the airplane, can be subsumed under this case as well.)

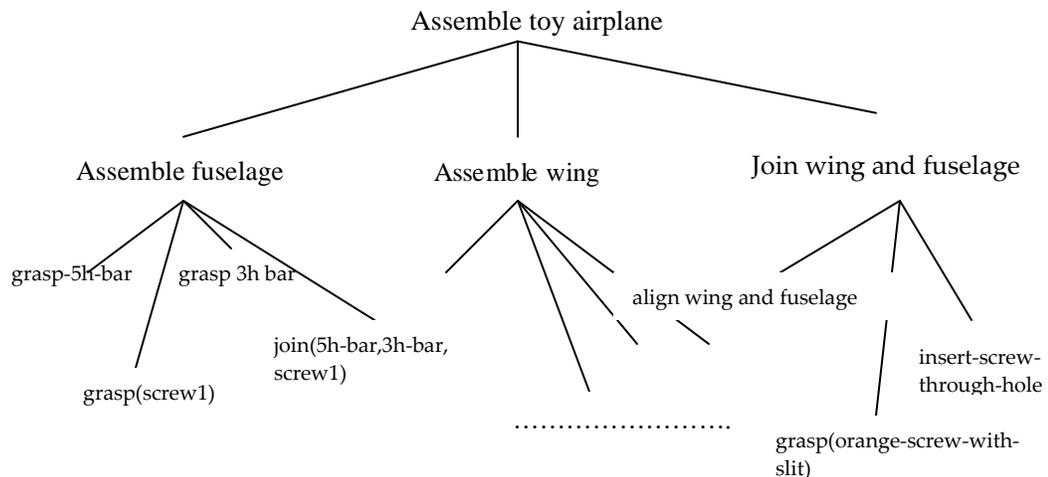


Fig 4.2.1: Inst's private plan before the conversation

According to the intentional view, Inst and Cnst develop a shared plan to achieve this we-intention of assembling the toy airplane. This shared plan is initially highly underspecified

<sup>12</sup> Arguably, a more plausible view is that Inst doesn't start with a complete plan for the construction of the model—instead, he/she develops her private plan incrementally as well. We do not think however that this more plausible, but also greatly more complex, view would have any implications for an account of completions.

and simply specifies that Inst and Cnst are going to assemble a toy airplane; but as the conversation progresses, the shared plan gets progressively more refined as each subplan becomes an SCA as well. (Note that the agreed upon actions in the shared plan are immediately executed, mixing planning and execution, unlike in the TRAINS conversations (Gross et al, 1993; Heeman and Allen, 1995), for example.)

At the point in which Inst begins utterance 1.1 in (1.1), the state of Cnst's assembly is as shown in Figure 2.1, repeated here for convenience as Figure 4.2.2. Cnst has built the tail and the rear part of the fuselage. She has already picked up a 7-hole bar which is to become the 'wing' (henceforth, W), has laid it across the 'fuselage' (henceforth, F), and has aligned the wing's and the fuselage's holes, which will make it possible for a fixing mechanism such as a screw to join wing and fuselage. Each of these steps has been acknowledged by Inst and Cnst. (A full listing of the dialogue prior to the fragment being analyzed is in Appendix A.)

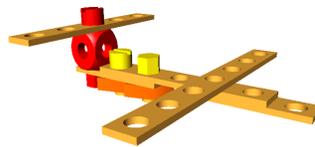


Figure 4.2.2: Instructor's and Constructor's respective situations at the beginning of 1.1

Given this state of assembly, the partial shared plan at this point is presumably similar to the one shown in sketchy form in Figure 4.2.3. The parts of Inst's private plan devoted to the assembly of fuselage and wing are now shared (in fact, they have already been executed). Now Inst and Cnst have a we-intention that Cnst join wing and fuselage; this we-intention is indicated in italics as it is currently at the top of the agenda. (More details on the shared plan in a moment.)

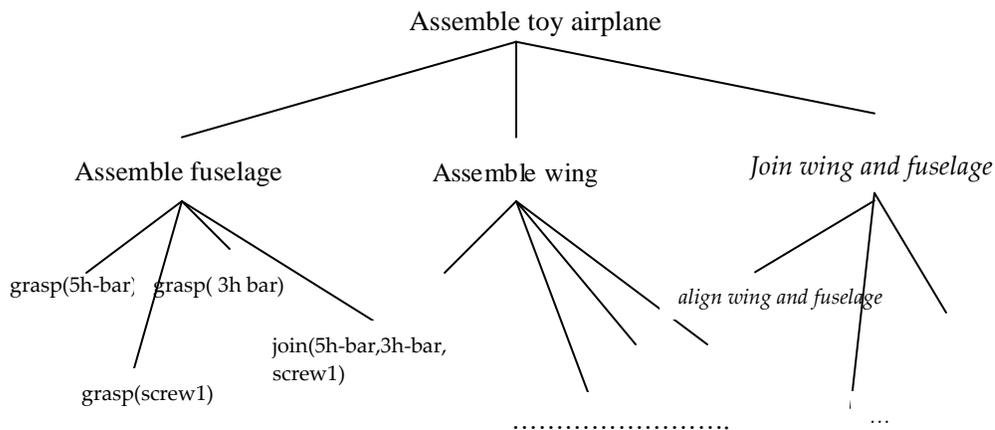


Fig 4.2.3: The shared plan at the point the completion takes place

At this point in the dialogue Cnst has a private plan as well, a refinement of the partial shared plan in Fig. 4.2.3. This private plan probably already contains the information that she will need to get a screw in order to join wing and fuselage, as sketchily shown in Figure 4.2.4.

However, Cnst has lots of spare parts left, above all nine screws which she can in principle use for the join as shown in Figure 4.2.5; hence, she cannot refine her own private plan any further.

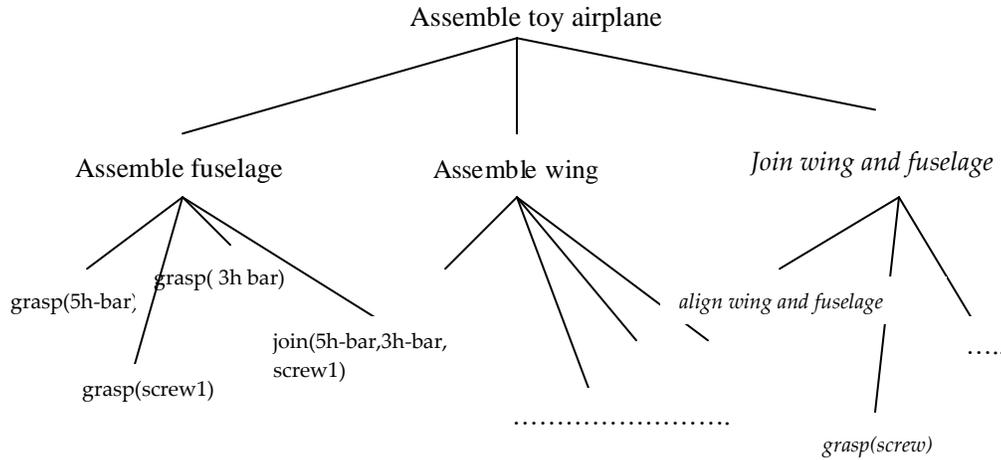


Fig 4.2.4: Inst's private plan at the point the completion takes place.

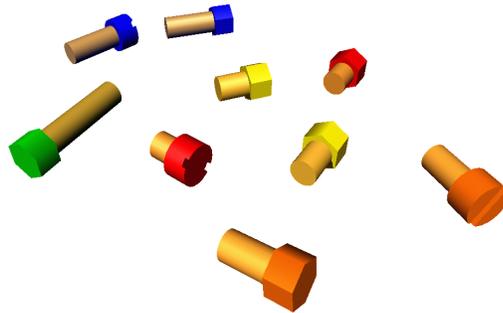


Figure 4.2.5: Screws available to Cnst in at the point when 1.1 is uttered. The intended screw is the one orange screw with a slit.

As said above, in the Tuomela-derived logical notation just introduced, the currently active **we-intention** to join W and F can be expressed as follows:

$$(4.2.2) \quad \mathbf{Int}_{Inst\&Cnst}(\mathbf{join}(Cnst,W,F))$$

As shown in (4.1.6), repeated below for convenience, Inst and Cnst having a **we-intention** to join wing and fuselage, in Tuomela's framework, is equivalent to them mutually knowing that each of them has an intention to do this action because of the other agent's intention to do it:

$$(4.1.6) \quad \mathbf{Int}_{Inst\&Cnst}(\mathbf{join}(Cnst, W,F)) \\ \leftrightarrow \\ \mathbf{MK}((\mathbf{Int}_{Inst}(\mathbf{join}(Cnst, W,F)) /_r \mathbf{Int}_{Cnst}(\mathbf{join}(Cnst, W,F))) \\ \text{and} \\ (\mathbf{Int}_{Cnst}(\mathbf{join}(Cnst, W,F)) /_r \mathbf{Int}_{Inst}(\mathbf{join}(Cnst, W,F))))$$

In order to join two objects Cnst needs, first of all, to find a screw and a nut. She must then stick the screw through the holes and screw it into the nut in order to produce a stable join of three bars (that we will call here the Wing&Fuselage-join). This partial *domain* plan, that we assume to be shared, can be summarized as follows:<sup>13</sup>

(4.2.3)            **join**(Agent, Bar1,Bar2)  
                   **Step 1:**            **align**(Agent, Bar1, Bar2, Hole1,Hole2),  
                   **Step 2:**            **grasp**(Agent, Screw),  
                   **Step 3:**            **grasp**(Agent, Nut),  
                   **Step 4:**            **put-through**(Agent, Screw, Hole1,Hole2),  
                   **Step 5:**            **screw-into**(Agent, Screw, Nut),

As in (Asher and Lascarides, 2003), we assume that the relation between intentions and actions (here formalized in terms of plans) is non-monotonic: i.e., that a plan such as (4.2.3) is a *default* way of executing a particular action. (Asher and Lascarides formalization of intentions and actions does not involve the notion of plans.) This can be formalized as follows, where we use a generic **normally involves** operator to state that the conclusions are non-monotonic:<sup>14</sup>

(4.2.4) **join**(Agent, Bar1, Bar2) **normally involves** b & c & d & e & f  
           where  
           (b) **align**(Agent, Bar1,Bar2, Hole1,Hole2)  
           (c) **grasp**(Agent,Screw),  
           (d) **grasp**(Agent, Nut)  
           (e) **put-through**(Agent, Screw, Hole1,Hole2),  
           (f) **screw-into**(Agent, Screw, Nut)

In the conversations of the Bielefeld corpus, things are a bit more complex, in two respects: (i) actions are performed by Cnst under Inst's instruction; and (ii) actions are immediately executed. As it is not our goal here to provide an account either of the interleaving of planning and execution, or of the interleaving of discourse plans and domain plans (see e.g., (Litman and Allen, 1990)), we simplify matters by assuming that in this domain, shared plans / meshing subplans combine discourse actions, domain actions, and execution. For the particular example under discussion, the resulting plan is assumed to be as shown in (4.2.5). According to this plan for Cnst to join wing and fuselage, in order to jointly perform the action currently **we-intended**, Inst has to produce the directives triggering these actions; each of these results in actions by Cnst; and at the end, Cnst must communicate that she completed the actions demanded.

---

<sup>13</sup> We are using a very simplified view of plans here, ignoring the distinction between preconditions –conditions that have to hold in order for the plan to be executable—and steps of the plan proper. Normally, having the Nut and the Screw would be considered to be a precondition. (This difference is not unlike that between presupposition and assertion.)

<sup>14</sup> (4.2.4) is meant to be a neutral notation which could be reconstructed in different nonmonotonic formalisms. In SDRT, the relation between intentions and plans would be expressed by a conditional, and **normally involves** would be Asher and Morreau's >Asher&Morreau operator. In PTT defaults are inference rules, formulated using Brewka's (1990) prioritized version of Reiter's (1980) Default Logic).

(4.2.5) **join**(Cnst, W, F) normally involves b & c & d & e & f & g, where

- (b) 1. **directive**(Inst, Cnst, **align**(Cnst, W, F, Hole1, Hole2)),  
2. **align**(Cnst, wing, fuselage, Hole1, Hole2),  
3. **assert**(Cnst, Inst, **aligned**(Cnst, W, F, Hole1, Hole2))
- (c) 1. **directive**(Inst, Cnst, **grasp**(Cnst, Screw)),  
2. **grasp**(Cnst, Screw),  
3. **assert**(Cnst, Inst, **grasped**(Cnst, Screw ))
- (d) 1. **directive**(Inst, Cnst, **grasp**(Cnst, Nut)),  
2. **grasp**(Cnst, Nut),  
3. **assert**(Cnst, Inst, **grasped**(Cnst, Nut))
- (e) 1. **directive**(Inst, Cnst, **put-through**(Cnst, Screw, Hole1,Hole2)),  
2. **put-through**(Cnst, Screw, Hole1,Hole2),  
3. **tell**(Cnst, Inst, **put-through**(Cnst, Screw, Hole1,Hole2))
- (f) 1. **directive**(Inst, Cnst, **screw-into**(Cnst, Screw, Nut)),  
2. **screw-into**(Cnst, Screw, Nut),  
3. **assert**(Cnst, Inst, **screwed-into**(Cnst, Screw, Nut))
- (g) **assert**(Cnst, Inst, **joined**(Cnst, W,F).

We will assume throughout these two sections that intention recognition for Cnst amounts to recognizing the directives and performing the required actions.<sup>15</sup>

Bratman's and Tuomela's theories of intention make the stronger assumption that a we-intention to bring-about a W&F join distributes over the conjuncts: i.e., having a we-intention to bring about an action entails we-intentions for all parts of the plan. With this assumption, we would then also derive from (4.2.1) and (4.2.5) (c) the following:

- (h) **Int**<sub>Inst&Cnst</sub> (**directive**(Inst, Cnst, **grasp**(Cnst, Screw)))
- (i) **Int**<sub>Inst&Cnst</sub> (**grasp**(Cnst, Screw))
- (j) **Int**<sub>Inst&Cnst</sub> (**assert** (Cnst, Inst, **grasped**(Cnst, Screw))).

Notice that even with this assumption, Inst and Cnst would still have different private plans and different private intentions: Inst, seeing the fully built up airplane model on his side, subscribes to

(4.2.6) **Int**<sub>Inst</sub>(**directive**(Inst, Cnst, **grasp**(Cnst, orange-slit-screw)))<sup>16</sup>

whereas for Cnst we have

(4.2.7) **Int**<sub>Cnst</sub>(**grasp**(Cnst, Screw)).

(The difference between (4.2.6) and (4.2.7) is due to the fact that **grasp**(Cnst, Screw) can be satisfied by Cnst in various ways due to the nine screws she has.) The point is that if we make the further assumption that **Int**<sub>Inst&Cnst</sub> is distributive, we get (4.2.7) irrespective of any inferences

---

<sup>15</sup> Intention recognition could also be formalized in such a way as to bypass the directive recognition stage, and assuming simply that Cnst's recognizes Inst's intention of Cnst performing a particular action. An account along these lines would not, however, differ from the one we are adopting for our purposes.

<sup>16</sup> We are not being very precise concerning notation here. Quantificational formulae are sometimes represented as constants for convenience's sake.

due to the verbal exchange: in other words, with this assumption, Cnst already wants to grasp a screw *before* Inst starts his directive. (Cnst's intention can be satisfied because there are screws on his side of the screen.) The only difference between the two plans would then be that Cnst doesn't know which particular elements have to be used. However, we will remain non-committal on whether Cnst derives (4.2.7) from distributivity of we-intentions or from the directive and cooperativity.

We now consider how an intentional account explains what may have caused Cnst to produce a completion in 1.2. Given that  $\text{Int}_{\text{Inst}}(\text{directive}(\text{Inst}, \text{Cnst}, \text{grasp}(\text{Cnst}, \text{Screw})))$  is the next step in the interleaved discourse / domain plan in (4.2.5) to achieve SCA (4.2.2), we may *ceteris paribus* assume that by uttering *Well, now you grasp* Inst has started the production of the next directive in this shared plan. What about Cnst's contribution? According to the mutually intended step (d) of (4.2.5), Cnst should by default wait until Inst has fully produced his directive, and only then she should do the grasping. But she intervenes. So we must explain:

- a. the cause for the completion, including the information used for the intervention;
- b. how the old shared plan changes, and what the new shared plan looks like;
- c. the relation between old and new shared plan.

Let us start with the cause for the completion and the information used for the intervention. The step in the plan which is we-intended at this point in the conversation is:

$\text{join}(\text{Inst}\&\text{Cnst}, W, F).$

Inst begins to perform the next step, (c):

$\text{directive}(\text{Inst}, \text{Cnst}, \text{grasp}(\text{Cnst}, \text{orange-slit-screw})).$

However, Cnst 'jumps in' before the directive is completed, and possibly without being prompted by some sort of request for help (as we saw in Section 2, there is some evidence that in this particular example a request for completion may have been made using prosody, but the matter is not entirely clear, and these signals are not encountered with all completions). Clearly, Cnst has been trying to compare her expectations (based on the plan she assumes to be shared) with the actions she observes (as proposed, e.g., in the MOSAIC model (Wolpert et al, 2003)) and she has been doing this incrementally. As a result, even before Inst's utterance of the directive is completed, Cnst has already related the incomplete utterance she is observing to the next step in the shared plan,  $\text{directive}(\text{Inst}, \text{Cnst}, \text{grasp}(\text{Cnst}, \text{Screw}))$ . We will discuss in more detail how this can happen in the next section, in which PTT is used to provide a detailed account of incremental semantic interpretation in this example.

As we said above, in a theory in which we-intentions are assumed to distribute over the subactions of plans, recognizing this directive is predicted to be fairly straightforward, as intention (4.2.6) would already be near the top of Cnst's agenda. Otherwise, some sort of search through the space of possible plans to achieve (4.2.2) must be assumed.

By comparing the part of the directive that has already been produced by Inst with the directive she expects, Cnst can hypothesize that what's missing so far is an utterance of the fact that the object to be grasped is a screw. (There is no more information about this in the shared plan, and as there are nine screws, there are nine ways to make the plan more specific, so there is no way for Cnst to make the recognition more precise, except by guessing.)

We see at least four ways of explaining why Cnst follows up this recognition with a decision to utter "a screw":

- a. *Responding to a Request*: Cnst may have interpreted 1.1 (more precisely, its prosodic lengthening) either as a Request to perform a Continue (DU), or as a Request to Acknowledge the DU. In either case, in PTT it is assumed that the result is an obligation:

**obl(Cnst,cont(DU)) or obl(Cnst, ack(DU)).**

Notice how an account of this type presupposes a theory of obligations and their discharge of the type of the theory in (Matheson, Poesio and Traum, 2000).

- b. *Voluntary coordination-level control*: Even without having interpreted Inst’s utterance as a request, Cnst may nevertheless intend to signal her understanding of the directive: i.e., to acknowledge Inst’s directive. Acknowledging a directive cannot be done simply by repeating the part already uttered: this might be interpreted simply as a partial acknowledgment of only the part of the directive already performed. Hence, Cnst acquires the (private) intention to perform the missing part of the directive (as above). However, the directive is still considered by Cnst as performed by Inst only. Such an account presupposes a theory of grounding acts like the one discussed in Section 3.
- c. *Cooperativeness*: Cnst intends to help Inst to perform the directive:

**directive(Inst,Cnst,grasp(Cnst,Screw)).**

We formalize this as Cnst making performing the directive into a shared plan:<sup>17</sup>

**Int<sub>Cnst&Inst</sub> (directive(Inst&Cnst,Cnst,grasp(Cnst,Screw))).**

This leads to Cnst performing an “unrequired contributory action”: performing the part of the directive that is missing. As a result, Cnst assumes the private intention:

Int<sub>Cnst</sub> (utter(“a screw”))

- d. *Blurting out*: Cnst feels that Inst is taking too long to perform a simple directive. As soon as Cnst recognizes the directive that Inst is trying to perform, the plan (in terms of utterance actions) Inst is using to do so, and what is still missing for the completion of this plan, Cnst acquires the intention to utter the missing part, without necessarily making the directive into a joint action, and without necessarily intending to acknowledge Inst’s contribution. (Although both intentions could be there.)

All these are valid explanations of what may have happened, but we won’t have enough space to pursue all of them in detail. In a way, explanation d. is the most interesting, but also the most difficult to formalize. Explanations a.-c. all involve reasoning about information states—i.e., the type of explanation that is most in contrast with the alignment model proposed by Pickering and Garrod—so we will concentrate on them. Further, explanations a. and b. are examples of the type of interaction most closely studied in previous work on PTT, particularly in (Matheson *et al*, 2000), so we will primarily concentrate on providing a detailed account of the Cooperativeness explanation, c., providing however the bare bones of an account according to a. and b.—and in particular, of the interaction between such account and the grounding process.

---

<sup>17</sup> This notion of cooperativeness is different from Cohen and Levesque’s in two respects. First of all, Cnst is deriving a *joint* intention from a *shared* plan, instead of deriving a *private* intention from Inst’s plan. Secondly, as we will see in Section 5, Cnst is adopting an intention to perform *part* of a plan whose execution would satisfy an obligation of Inst, instead of performing an action from scratch.

After the completion, as a result of Inst and Cnst's coordinated action we get a complete directive: *Well, now you take a screw*. However, there is a difference between this directive and the version in Inst's private plan:

**Int<sub>Inst</sub>(directive(Inst, Cnst, grasp(Cnst, orange-slit-screw)))**

Inst may react to Cnst's contribution by either

- a. Accepting the completion (either because the complete speech act reflects Inst's initial intention, or because Inst considers the new directive a valid alternative);
- b. Rejecting it,
- c. Performing a literal resumption,
- d. Paraphrasing,
- e. Refashioning it, by modifying some aspects of it..

In this particular case, Inst appears to be performing a refashioning, as discussed in Sections 2 and 3 (Clark and Wilkes-Gibbs 1986). Inst knows that orange-lit-screw  $\subset$  screw; this can be captured by the following meaning postulate:

$\text{orange slit-screw}(x) := \text{screw}(x) \wedge \text{orange}(x) \wedge \exists y(\text{slit}(y) \wedge \text{with}(y,x))$ .

So he accepts some aspects of the contribution, but adds more material to ensure the right screw is identified. Inst's intention is of refashioning the directive currently under elaboration into the more specific directive:

**Int<sub>Inst</sub>(directive(Inst, Cnst, grasp(Cnst, screw(x)  $\wedge$  orange(x)  $\wedge$   $\exists y(\text{slit}(y) \wedge \text{with}(y,x))$ ))))**

Again, Inst compares the directive already produced with his intended directive. A *screw* has already been produced; the difference between Cnst's  $\exists x(\text{screw}(x))$  and Inst's  $\exists x(\text{screw}(x) \wedge \text{orange}(x) \wedge \exists y(\text{slit}(y) \wedge \text{with}(y,x)))$ , is

$\exists x(\text{orange}(x) \wedge \exists y(\text{slit}(y) \wedge \text{with}(y,x)))$ .

We get the more economical

**Int<sub>Inst</sub>(directive(Inst, Cnst, grasp(Cnst, orange(x)  $\wedge$   $\exists y(\text{slit}(y) \wedge \text{with}(y,x))$ ))))**

yielding *an orange one with a slit*.

Notice how performing the type of reasoning discussed here requires the ability to talk about incomplete speech acts, i.e., a theory such as PTT. We next show how PTT allows us to make the account just presented more precise.

## 5 An Intentional Analysis of the Example Dialogue: Second Pass

In this section we analyze the example dialogue, utterance unit by utterance unit,<sup>18</sup> supplementing the intentional analysis just provided with a discussion of how the utterance units are incrementally interpreted and produced according to the PTT view of interpretation, generation and the grounding process, on the basis of the grammar given in the CDRT fragment in Appendix B. The analysis of these aspects we will provide is very detailed. On the other hand, PTT, unlike SDRT, does not provide (yet) a complete set of default rules specifying the inferential connection between utterances and their interpretation, so we will only discuss these inferences in an informal way. (See (Poesio, 1994; Poesio, 1996; Poesio, to appear) for a discussion of the underlying nonmonotonic reasoning formalism and a formalization of other types of inferences within it.)

### 5.1 Discourse situation updates up to the point just before 1.1

We discussed in Section 4 how according to the intentional analysis, a conversation like the one from which fragment (1.1) is extracted is driven by Inst and Cnst's we-intention to build a toy plane according to the model given to Inst, which in the simplified view adopted here becomes the private plan in Figure 4.2.1. By way of his instructions, Inst turns the private plan into a progressively more specified shared plan interleaving discourse and domain actions as in the example in (4.2.5). As a result of Cnst's performing the domain actions in this shared plan, Inst and Cnst reach the situation whose relevant aspects are depicted in Figure 4.2.1 (state of assembly) and Figure 4.2.4 (leftover screws).

In PTT, performing a directive amounts to performing a contribution which, ultimately, amounts to a series of micro-updates as in (3.3.5), each of which has to be properly grounded. The discourse situation just before (1.1) is a record of both core speech acts and micro conversational events performed up to that point.

### 5.2 Production of 1.1: So, jetzt nimmst Du ....

We saw in section 4.2 that at this point in the dialogue, Inst and Const **we-intend** to join two specific objects, that we will call here *wing1* and *fuselage1*.<sup>19</sup> In PTT notation, and using DRSS as intentional contents, the fact that we-intention (4.2.2) is part of the common ground is captured by including (5.2.1) among the conditions characterizing the discourse situation.<sup>20</sup>

(5.2.1) i1.1a:Int<sub>Inst&Cnst</sub> ([e1 | e1:join(Cnst,wing1,fuselage1)])

We also saw that in the BTPC conversations such goals are achieved by following plans along the lines of (4.2.5), whose next step (step (c)) is a directive by Inst to Cnst to grasp a screw. In PTT terms, performing such a directive amounts to planning (i.e. intending) a conversational event *ce1.1* of type **directive** and with content proposition K1.1: the occurrence of

---

<sup>18</sup> As in (Traum, 1994, Traum & Nakatani, 1999, Poesio and Traum, 1997, Poesio and Traum, 1998) we use the term UTTERANCE UNIT to refer to basic units of dialogue processing, corresponding roughly to prosodic units.

<sup>19</sup> We remind the reader that in CDRT, unlike standard DRT, there are regular constants; *wing1* and *fuselage1* are such constants, and so is *orange-slit-screw*.

<sup>20</sup> We will generally assume *de dicto* (wide scope) intentions.

an event  $e$  of  $Cnst$  grasping *orange-slit-screw*. (The event should take place in the future, but we will omit tense here for brevity).

(5.2.2)  $i1.1b: \text{Int}_{\text{Inst}}([\text{ce}1.1, K1.1 \mid K1.1 = [x, e \mid x \text{ is orange-slit-screw}, e: \text{grasp}(Cnst, x)],$   
 $\text{ce}1.1: \text{directive}(\text{Inst}, Cnst, K1.1)])$

A couple of remarks about (5.2.2). Notice, first, that embedded DRs are essential to formulate this intention. Secondly,  $\text{Inst}'$ 's intention is very likely to be about a specific screw (represented here as the constant *orange-slit-screw*), instead of the more general intention to grasp an  $x$  which is a screw (in (5.2.2')) which can be expected to be in the shared plan. (That  $\text{Inst}'$ 's intention is about a specific screw is suggested, e.g., by the subsequent repair.)

(5.2.2')  $i1.1b': \text{Int}_{\text{Inst}}([\text{ce}1.1, K1.1 \mid K1.1 = [x, e \mid \text{screw}(x), e: \text{grasp}(Cnst, x)],$   
 $\text{ce}1.1: \text{directive}(\text{Inst}, Cnst, K1.1)])$

(In passing, note also that if we assume that we-intentions distribute over the plan in (4.2.5),  $Cnst$  could derive (5.2.2') from the non-specific *joint* intention in the shared plan shown in (5.2.2'').)

(5.2.2'')  $i1.1b'': \text{Int}_{\text{Inst} \& \text{Cnst}}([\text{ce}1.1, K1.1 \mid K1.1 = [x, e \mid \text{screw}(x), e: \text{grasp}(Cnst, x)],$   
 $\text{ce}1.1: \text{directive}(\text{Inst}, Cnst, K1.1)])$

Assuming that intentions to perform domain actions lead to the adoption of domain plans to achieve such intentions via von Wright's (1963) Practical Syllogism,<sup>21</sup> (5.2.2) leads to  $\text{Inst}'$ 's plan of performing an utterance generating the directive in the sense discussed in Section 3.1, example (3.1.5). To simplify matters, we assume here, as in (Poesio and Traum, 1997), that this plan amounts to intending to perform an utterance  $up1.1$ <sup>22</sup> whose conventional meaning (the value of  $\text{sem}(u1.1)$ ) is the same as the content of the directive.

(5.2.3)  $i1.1c: \text{Int}_{\text{Inst}}([\text{up}1.1 \mid \text{utterance}(\text{up}1.1), \text{sem}(\text{up}1.1) = K1.1, \text{generate}(\text{up}1.1, \text{ce}1.1)])$

(5.2.3) is the starting point for the generation process. And at this point, the hypothesis that the participants' record of the discourse situation includes information about the occurrence of micro conversational events – events of uttering sub-sentential constituents – does begin to do some work. Specifically, the MCE hypothesis (i) connects intentions such as (5.2.3) with their syntactic realization, and (ii) provides us with an account of the information that is available to  $Cnst$  when she starts planning her completion.

In PTT, performing an utterance with the conventional meaning represented by  $K1.1$  is viewed as an action (of type **utter**). We can then view the results of surface realization, the process of determining the surface structure to be used to realize an utterance  $up1.1$  with content  $K1.1$  (Reiter and Dale, 1995), as a plan just like those produced in other types of planning. (The assumption being of course that specialized planners, called SURFACE REALIZERS in natural language generation, are responsible for this task.) As shown in (3.3.5), the plan chosen by  $\text{Inst}$  for producing conventional meaning  $K1.1$  in this particular occasion involves choosing  $up1.1$  to be an utterance of type S. This action is decomposed in the performance of an utterance  $u1.2$  of the temporal adverbial "Jetzt" and an utterance of a second phrasal utterance of type S, here called  $up1.2$ . Performing  $up1.2$  involves performing a (phonetically silent) utterance  $up1.3$  of type

<sup>21</sup> Intending  $\phi$  and believing that performing action  $\chi$  has as a consequence  $\phi$  leads to performing  $\chi$ :  $\text{Intend}(\phi) \wedge \text{Bel}(\chi \rightarrow \phi) \rightarrow \text{Do } \chi$  (from Wright, Georg Henryk von. 1963. "Practical Inference." *Philosophical Review*, 72:159-179. \_\_\_\_\_. 1972. "On the So-Called Practical Inference." *Acta Sociologica*, 15:39-53.)

<sup>22</sup> Terminological note: we use discourse referents named  $up_i.j$  to indicate phrasal utterances, where  $i$  is the number of the turn. We use discourse referents named  $ui.j$  to indicate lexical utterances.

NP and an utterance *up1.4* of type VP, which in turn involves three subactions: an utterance *u1.3* of type V (“nimmst”), a subject utterance *up1.5* of type NP (“Du”), and a complement utterance *up1.6* of type NP (“eine orangene Schraube mit einem Schlitz”).

Another advantage of the micro-conversational events hypothesis is that we do not need to assume that every utterance that gets produced is a constituent of a single syntactic tree—a big ‘sign’ in the HPSG sense. In the case of 1.1, for example, it is plausible to view at least the first “So” (indicated here as *u1.0*) and possibly even “jetzt”, as a separately planned utterance intended to achieve quite distinct goals from the other utterances which generate the directive: specifically, as a turn-taking action, whose goal is to take the turn and keep it (Traum and Hinkelman, 1992; Traum, 1994; Poesio and Traum, 1997).<sup>23</sup> The hesitation that may or may not have occurred at the end of 1.1 would play a turn-taking oriented goal, as well, as well as possibly a role in the grounding process. To show what such an analysis would be, we assume here this interpretation for “So,” whereas we interpret the utterance of “jetzt” as an utterance of type Advp with a sentence adjunct interpretation. Omitting many of the details already shown in (3.3.5), including lexical interpretations and many sub-utterances, and using the notation *u:“word“:Cat* to stand for:

$$u:\text{utter}(A, \text{“word“}), \text{Cat}(u)$$

Inst’s plan to satisfy *i1.1c* in (5.2.3) becomes a plan of performing the actions shown in tree format in Figure 5.2.1 and in linear format in (5.2.4).<sup>24</sup>

(5.2.4) *i1.1d:Inst*([*u1.0, ce1.0, up1.1, u1.2, u1.3, up1.2, up1.3, up1.4, u1.4, up1.5, up1.6* |  
*u1.0: utter*(Inst, “so”), *ce1.0: take-turn*(Inst),  
*generate*(*u1.0, ce1.0*),  
*S*(*up1.1*) , *sem*(*up1.1*)= *K1.1*,  
*u1.2:“jetzt“:Advp*, *S*(*up1.2*), *u1.2* ↑ *up1.1*, *up1.2* ↑ *up1.1*,  
*NP*(*up1.3*), *VP*(*up1.4*), *up1.3* ↑ *up1.2*, *up1.4* ↑ *up1.2*,  
*u1.3:“nimmst“:V*, *u1.3* ↑ *up1.4*, *NP*(*up1.5*), *up1.5* ↑ *up1.4*,  
*u1.4:“Du“:Pro*, *u1.4* ↑ *up1.5*,  
*up1.6:“eine orangene Schraube mit einem Schlitz“:NP*,  
*up1.6* ↑ *up1.4*)

(Again, this intention is dominated by *i1.1c*.) The overall mental state of Inst after acquiring intentions *i1.1b*, *i1.1c*, and *i1.1d* is portrayed in ‘box’ format in Figure 5.2.1, where constituency relations are also displayed in a graphical format.

<sup>23</sup> “So” could also be interpreted as a READY signal as in the MapTask scheme (Carletta et al, 1997).

<sup>24</sup> We have assumed for simplicity that the entire utterance is planned at once. However, In a more plausible theory, the generation process would be incremental, as well.

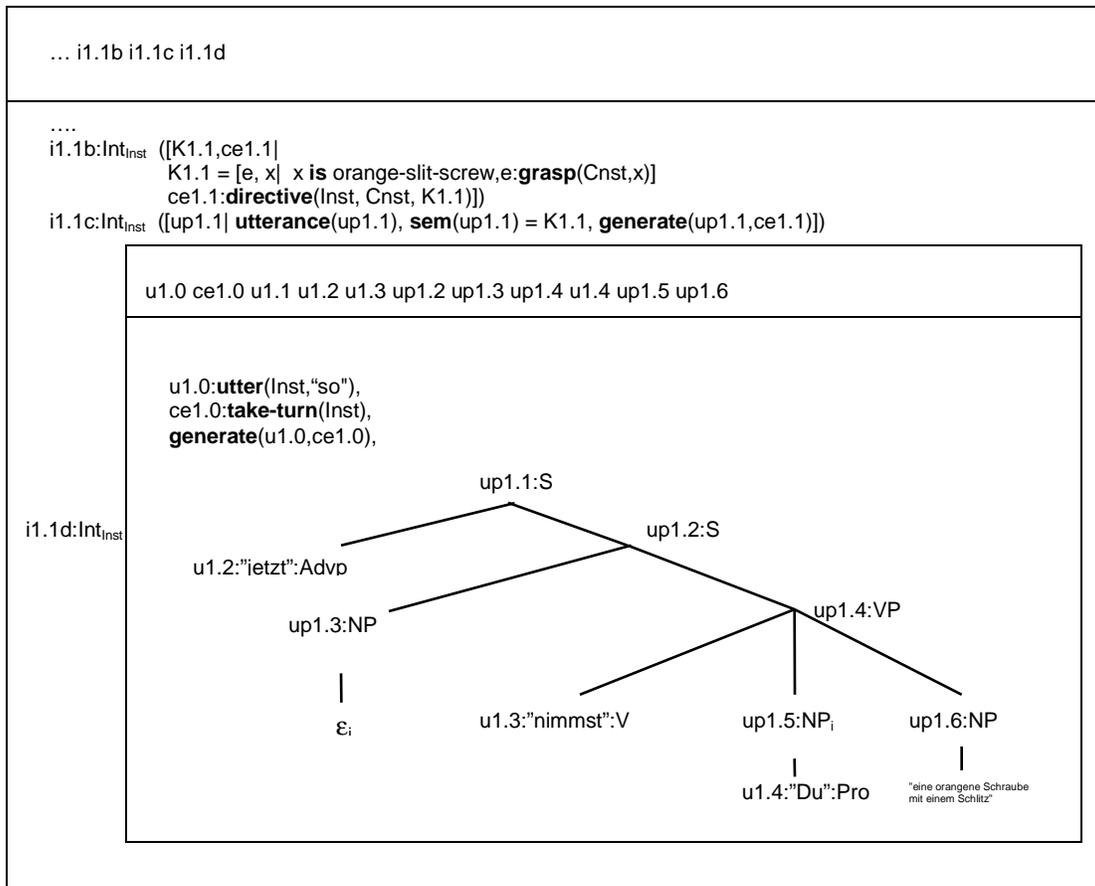


Figure 5.2.1 – Inst’s mental state after sentence planning

As said in Section 3, in the PTT model—derived from Clark and Schaefer’s proposals as modified by Traum—the intention to perform a dialogue act becomes an intention to make a new contribution, i.e., to open a new DU, according to the general schema:

$\text{Int}_{\text{cp1}}([\text{ce} \mid \text{ce:dialogue-act}(\text{cp1}, \text{cp2}, \Phi)]) \rightarrow \text{Int}_{\text{cp1}}([\text{DU} \mid \text{DU is } [\text{ce} \mid \text{ce:dialogue-act}(\text{cp1}, \text{cp2}, \Phi)])]$

i.e., (5.2.2), (5.2.3) and (5.2.4) lead to (5.2.5):

(5.2.5) i1.1e:Inst<sub>inst</sub> ([DU1.1<sub>inst</sub> |  
 DU1.1<sub>inst</sub> is [ce1.1 K1.1  
 up1.1,u1.0,ce1.0,u1.2,u1.3,up1.2,up1.3,up1.4,u1.4,up1.5,up1.6 |  
 K1.1 is [x,e | x is orange-slit-screw, e:grasp(Cnst, x)],  
 ce1.1:directive(Inst, Cnst,K1.1),  
 utterance(up1.1), sem(up1.1) = K1.1, generate(up1.1,ce1.1),  
 .... remaining conditions from (5.2.4) ...  
 ]])

Intentions such as i1.1e in (5.2.5) lead to Inst performing the planned action, i.e., beginning a new contribution. Inst starts executing this intention, succeeding to perform u1.2, u1.3, and u1.4. It’s not clear to us whether the lengthening observed at this point is meant to indicate a problem – e.g., that Inst has forgotten which screws are unused, and therefore does not know what description would be adequate (Dale and Reiter, 1995; Stone, 2004) – a request for acknowledgment, or whether it is just accidental.

### 5.3 The completion: 1.2, “eine Schraube”

The process that leads Cnst to produce the completion begins when she observes Inst performing the first four micro-conversational actions of his planned contribution. As discussed in Section 3, according to PTT observing these utterances leads Cnst to create a number of new Discourse Units, each of which represents “material to be grounded”; one ‘micro-DU’ per utterance event is assumed. For simplicity, however, we represent here Cnst’s interpretation of the utterance events in 1.1 as a single new Discourse Unit, that we will call DU1.1<sub>Cnst</sub> (to emphasize that it need not be the same as the DU from Inst’s perspective). The update to the discourse situation resulting from the observation of DU1.1<sub>Cnst</sub> is shown in (5.3.1). This new contribution minimally contains a record of the occurrence of each utterance, together with the results of lexical access and (incremental) parsing (see (Poesio, 1995; Poesio and Muskens, 1997; Poesio, To Appear) for details). Assuming that there are no interpretation problems, DU1.1<sub>Cnst</sub> would be a partial representation of the DU planned by Inst, as shown in (5.2.5). The difference between the two is that in the contribution observed by Cnst, NP *up1.6* hasn’t yet been observed, but it is already expected.<sup>25</sup> (In addition, a new dialogue act *ce1.1* is presumably hypothesized, but at this point there is no evidence that it has already been classified as a directive, although that is likely as well given the information from word order and the pending **we-intention**) We show this interpretation of Cnst in linear format in (5.3.1) and in tree format in Figure 5.3.1.

(5.3.1) [ DU1.1<sub>Cnst</sub> | DU1.1<sub>Cnst</sub> is [u1.0,ce1.1,u1.2,up1.1,u1.3,up1.2,up1.3,up1.4,up1.5,up1.6,u1.4 |  
u1.0: “so”:take-turn,  
S(up1.1) ,  
u1.2:“jetzt”:Advp, S(up1.2),  
u1.2 ↑ up1.1, up1.2 ↑ up1.1,  
u1.3:“nimmst”:V, NP(up1.3), VP(up1.4),  
up1.3 ↑ up1.2, up1.4 ↑ up1.2,  
u1.3 ↑ up1.4, NP(up1.5), up1.5 ↑ up1.4, NP(up1.6),  
up1.6 ↑ up1.4,  
u1.4:“Du”:NP, u1.4 ↑ up1.5 ]]

---

<sup>25</sup> We assume here that Inst and Cnst have the same mental grammar.

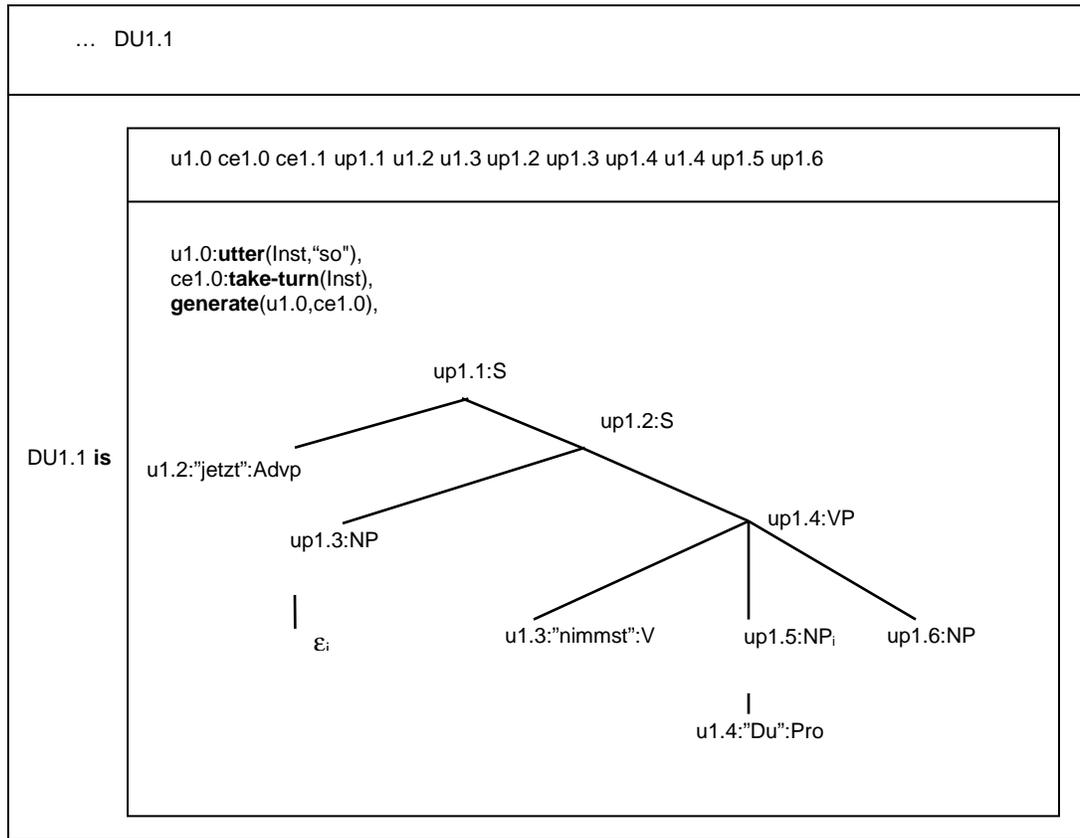


Figure 5.3.1 – Cnst's interpretation after perceiving 1.1

(5.3.1) is the starting point for Cnst's semantic interpretation process. The extent to which semantic construction and speech act interpretation take place immediately is still an open question (our views are discussed in more detail in (Poesio, To Appear)). The interpretation in (5.3.1) simply encodes the results of lexical access and preliminary syntactic interpretation, both of which are known to at least begin very early (Swinney 1979, Simpson 1994, Frazier 1987, Tanenhaus et al 1995). There is also evidence that the observation of the occurrence of an event of uttering an anaphoric NP such as a pronoun is sufficient to start the processes by which such expressions are interpreted, as discussed in (Poesio, 2001; Poesio, To Appear). How much more interpretation takes place? As discussed in Section 4, under the intentional account the fact that a completion takes place is an indication that Cnst somehow manages to recognize Inst's intention to perform a directive. (We represented this intention there using the simplified form  $\text{Int}_{\text{Inst}}(\text{directive}(\text{Inst}, \text{Cnst}, \text{grasp}(\text{Cnst}, \text{Screw})))$ .) This recognition is what prompts Cnst to utter "eine Schraube." A more explicit PTT representation of the result of the first of these inferential processes is shown in (1), whereas the intention to utter a screw would be represented as in (2).

- (1)  $\text{Int}_{\text{Inst}}([\text{ce1.1}, \text{K1.1} \mid$   
 $\text{K1.1}=[x, e \mid \text{screw}(x), e:\text{grasp}(\text{Cnst}, x)]$   
 $\text{ce1.1}:\text{directive}(\text{Inst}, \text{Cnst}, \text{K1.1})]$
- (2)  $\text{Int}_{\text{Cnst}}([\text{u} \mid \text{u}:\text{utter}(\text{Cnst}, \text{"eine Schraube"})])$

Note that it's hard to tell from the dialogue whether Cnst has already identified a particular screw, so we will not assume anything in this respect here.

We consider two hypotheses concerning the way Cnst reaches conclusion (1). As for the way recognition of (1) leads to the adoption of intention (2), we already saw in Section 4 that this can be explained in a number of ways, and a number of these alternative hypotheses seem equally plausible, so that choosing among them would amount to mind reading. We will nevertheless attempt to show that several such hypotheses could be expressed in terms of PTT.

### 5.3.1 Identifying intention (1)

*Existential Closure* One hypothesis concerning the process through which Cnst reaches conclusion (1) is based on the operation of EXISTENTIAL CLOSURE proposed by Chater, Pickering and Milward (1995). According to Chater *et al*, existential closure takes place every time a new input is perceived; its purpose is to produce propositions out of partial syntactic interpretations, so that the new input can be immediately evaluated against the current situation. In PTT terms, the existential closure hypothesis amounts to hypothesizing that Cnst attempts to derive a proposition as the conventional meaning of *up1.1* by existentially closing the missing argument, even when the only information at her disposal is what is presented in (5.3.1). More precisely, the hypothesis is that as a result of observing the contribution represented in (5.3.1) as  $DU1.1_{Cnst}$ , Cnst hypothesizes a new DRS  $K1.1$  as the argument of an intention attributed to Inst:  $K1.1$  is the proposition that an event of the Constructor taking an object  $x$  yet to be determined will have to take place. (In PTT, the result of all inferences performed by Cnst on the basis of (5.3.1), such as existential closure, is viewed as material to be grounded separately, i.e., as a separate DU; for simplicity here we will however assume that all such inferences are added to the same DU,  $DU1.1_{Cnst}$ .) We remind the reader that we indicate DRS updates using the notation  $K += K'$ , defined as follows: (see Section 3.2):

**$K += K'$ :** Let  $K$  be a proposition-valued discourse referent, and  $K'$  be a proposition (DRS) or a proposition-valued discourse referent. Then

$$K += K' =_{\text{def}} [ \text{TmpK} \mid \text{TmpK is } K ]; [ \mid K \text{ is } \text{TmpK}; K' ]$$

(where  $\text{TmpK}$  is a new proposition-valued discourse referent)

Using this notation, the result of existential closure is the update of Cnst's view of the discourse situation shown in (5.3.2).

$$(5.3.2) \quad DU1.1_{Cnst} += [K1.1 \mid K1.1 \text{ is } [x,e \mid e:\text{grasp}(Cnst,x)]]$$

The recognition by Cnst that an action of grasping something is being discussed is the basis for her inferring that the grasping action in  $DU1.1_{Cnst}$  is the first step in performing the action currently we-intended. We recall that under the intentional account, Inst and Cnst are operating under the shared plan for bringing about the required join of the three pieces in (4.2.5), involving both domain and discourse actions. The next step in the 'existential closure hypothesis' is that inferring (5.3.2) leads Cnst to infer that Inst is performing the second directive (c.1) in the shared plan in (4.2.5). (The acceptance by Cnst of that directive will result in Cnst's adopting the intention of performing the action indicated by the directive, (c.2) in (4.2.5).)<sup>26</sup> This inference leads to Cnst updating her view of the content  $K1.1$  of the directive with the information [ | **screw**( $x$ )]:

$$(5.3.2') \quad DU1.1_{Cnst} += (K1.1 += [ | \text{screw}(x) ])$$

<sup>26</sup> In the 'standard' version of PTT, as presented, e.g., in (Matheson, Poesio, and Traum, 2000), adopting this intention would be a way of addressing the obligation raised by the directive. As already said above, we are not concerned with obligations here.

And then to update DU1.1<sub>Cnst</sub> with (1), i.e., to attribute to Inst a directive whose content is the updated K1.1. This is shown in (5.3.3). Notice that this intention is the intention discussed above in (5.2.2'). (As above, we assume Cnst ascribes to Inst a *de dicto* intention. Also, as discussed in Section 4, we leave it undetermined here whether this private intention attributed to Inst is directly derived from we-intention (h) derived from (4.2.5) via distributivity of we-intention.)

(5.3.3) DU1.1<sub>Cnst</sub> += [i1.1b' | i1.1b':Int<sub>Inst</sub>([ce1.1, K1.1 |  
**sem**(up1.1) **is** K1.1,  
K1.1 **is** [x,e | **screw**(x),e:**grasp**(Cnst,x)],  
**generate**(up1.1,ce1.1),  
ce1.1:**directive**(Inst,Cnst,K1.1))]]

*Direct inference from the lexicon to the plan* An alternative analysis of the inference process leading Cnst to conclude (5.3.3) is that Cnst recognized that Inst is performing directive (c.1) in (4.2.5) directly from the lexical meaning of mce1 in (5.3.1), which is a mention of a grasping action. In other words, mentioning a grasping action is sufficient by itself to activate step (c.1) of the shared plan in (4.2.5), without the step of existential closure resulting in (5.3.2). Although the end result would again be (5.3.3), this second hypothesis has the advantage of avoiding the need to stipulate that existential closure takes place after each and every utterance; and there is increasing evidence for this type of 'surface' semantic reasoning (see, e.g., the work by Ferreira *et al* on 'good enough' representations, (Ferreira, Ferraro, and Bailey 2002)).

### 5.3.2 Acquiring intention (2) to perform a completion

As discussed in Section 4, there are at the very least four possible explanations for the fact that Cnst acquires intention (2) to perform a completion: responding to a request, voluntary coordination control (Clark's 'collaborative completions'), cooperativeness, and 'blurring out'. All these four explanations assume that Cnst recognizes (5.3.1) as a partial plan for performing the contribution in (5.3.3). We discuss each in turn.

*Cooperativity* What we called the 'cooperativity explanation' in Section 4 is that Cnst has decided to change the current contribution by turning the directive in a joint action (and adding some details). Revising a contribution in this way is a sort of *refashioning* (Clark and Wilkes-Gibbs, 1990, p. 481-486) of the DU that represented the contribution. As discussed in Section 3, refashioning can be formalized in terms of the system of grounding actions proposed by Traum (1994) and adopted in PTT as a **repair** grounding act: Cnst proposes to replace DU1.1<sub>Cnst</sub>, according to which it is Inst alone who performs the directive, with the concatenation of DU1.1<sub>Cnst</sub>;DU1.2 of DU1.1<sub>Cnst</sub> and DU1.2 (shown in (5.3.4)), in which the directive is a joint action of Inst and Cnst. DU1.1;DU1.2 still contains all the information in Fig. 5.3.1, including the information about utterances and the fact that **sem**(up1.1) = K1.1.

(5.3.4) [i1.2a | i1.2a:Int<sub>Cnst</sub>([DU1.2,ce1.2 | DU1.2 **is** [ce1.1, K1.1 |  
K1.1 **is**[x,e | **screw**(x),e:**grasp**(Cnst,x)],  
ce1.1:**directive**(Inst&Cnst,Cnst,K1.1)]  
ce1.2:**repair**(Cnst,DU1.1<sub>Cnst</sub>,DU1.1<sub>Cnst</sub>;DU1.2))]]

If the repair is accepted, (each participant's view of contribution) DU1.1 gets replaced by DU1.2, as seen in Section 3 and discussed below. (Arguably, Cnst is *also* acknowledging the partial contribution represented by DU1.2; we will discuss this additional function of the completion shortly.)

After deciding to turn the directive into a joint action, Cnst has to decide how best to complete it—i.e., which additional utterance actions to perform in order to generate ce1.1 as in

(5.3.4). Again, Cnst may have determined which utterances are missing from (5.2.4) in a number of ways. The simplest explanation is that a sort of structural alignment took place: Cnst develops her own plan for performing the directive in (5.3.4), and then compares this plan with the actions she observed. Cnst's plan to perform ce1.1 is a series of utterances very much like those in (5.2.4), except that it includes the generic "eine Schraube" instead of the more specific NP "eine orangene Schraube mit einem Schlitz", so the conventional meaning she associates to *up1.1* is the generic proposition to grasp a screw. Cnst's utterance plan is shown in (5.3.5).

(5.3.5) [ce1.1, u1.2, up1.1, u1.3, up1.2, up1.3, up1.4, up1.5, up1.6, u1.4 |  
 S(up1.1), **sem**(up1.1) **is** K1.1,<sup>27</sup>  
 u1.2:"jetzt":Advp, S(up1.2),  
 u1.2 ↑ up1.1, up1.2 ↑ up1.1,  
 u1.3:"nimmst":V, NP(up1.3), VP(up1.4),  
 up1.3 ↑ up1.2, up1.4 ↑ up1.2,  
 u1.3 ↑ up1.4, NP(up1.5), up1.5 ↑ up1.4, NP(up1.6),  
 up1.6 ↑ up1.4,  
 u1.4:"Du":NP, u1.4 ↑ up1.5,  
 u1.5:"eine":Det, u1.6:"Schraube":N,  
 u1.5 ↑ up1.6, u1.6 ↑ up1.6]

Having produced this utterance plan / syntactic structure, Cnst compares it with the utterances she has observed Inst performing.<sup>28</sup> Cnst then decides to include in the new contribution she is planning, DU1.2 (shown in (5.3.4)), the missing information about NP *up1.6*, shown in reduced form in (5.3.6).

(5.3.6) [i1.2b | i1.2b:IntCnst(DU1.2 += [ | up1.6:"eine Schraube":NP])]

(by the definition of +=, this is equivalent to:

[i1.2b | i1.2b:IntCnst([TmpK | TmpK **is** DU1.2]; [ | DU1.2 **is** TmpK; [ | up1.6:"eine Schraube":NP]])]

The fact that Cnst decides to utter an NP instead of producing a complete utterance such as, say, "Ich nehme eine Schraube" might also be explained in terms of cooperativity: reusing as much as possible of Inst's output would make ce1.1 a genuinely joint action.

In case existential closure were assumed, alignment at the *domain* level (rather than at the syntactic one) could also be used to explain how Cnst plans her contribution to the joint directive (see (4.2.5)(c)). That is, Cnst might have decided to utter an NP by comparing the content of the jointly intended directive in (5.3.4) –the proposition to grasp a generic screw– with the proposition of grasping some object in (5.3.2), derived from the partial directive that Inst has performed by means of existential closure. Cnst would then conclude that this latter proposition should be augmented with K1.1d:

(5.3.7) K1.1d=[x | **screw**(x)]

<sup>27</sup> We are forcing a bit the notation here and use the **is** notation to indicate the condition  $\lambda$  i.  $\text{sem}(v(\text{up1.1})(i)) = v(K1.1)(i)$ .

<sup>28</sup> This assumption that participants in a conversation always compare their expectations with what they observe is well-accepted in cognitive science: for instance, it is at the basis of Wolpert, Doya and Kazato's (2003) 'MOSAIC' model of language comprehension, and also appears to be broadly consistent with Pickering and Garrod's (2004) 'alignment' model (which will be discussed at greater length in a following section).

Cnst could then decide to produce an utterance with K1.1d as content – i.e., an indefinite NP, very much as in (5.3.6). (Cfr. Pickering and Garrod’s “I/O constraint”.)

*Acknowledgment Completions* are often produced in order to acknowledge an (as yet incomplete) DU (see Clark’s ‘collaborative completions’, Clark 1996, p. 238). In other words, Cnst’s intention in uttering “eine Schraube” may have been to acknowledge (the complete version of) Inst’s directive, instead of, or in addition to, repairing it, as proposed in (5.3.4). This possible interpretation of Cnst’s intention is shown in (5.3.8). The content of intention i1.2a in (5.3.8) is an acknowledgment of the *completed* DU1.1<sub>Cnst</sub> in (5.3.3) –i.e., not simply the utterance, but the utterance together with the results of inferences:

(5.3.8) [i1.2a | i1.2a:Int<sub>Cnst</sub>([ce1.3 | ce1.3:**ack**(Cnst,DU1.1<sub>Cnst</sub>)])]

The effect of this acknowledgment is to make DU1.1<sub>Cnst</sub> grounded: i.e., to update the discourse situation with G(DU1.1<sub>Cnst</sub>) (see below).

A few aspects of this interpretation of what happened in the sample dialogue are worth discussing. First of all, one might argue that what gets acknowledged by Cnst is not just DU1.1<sub>Cnst</sub>, but the proposition resulting from the concatenation of DU1.1<sub>Cnst</sub> with the record of the completion taking place:

DU1.1<sub>Cnst</sub> += [ | up1.6:“eine Schraube”:NP]

I.e., that what Cnst actually intends is the following:

(5.3.8’) [i1.2a | i1.2a:Int<sub>Cnst</sub>( (DU1.1<sub>Cnst</sub> += [ | up1.6:“eine Schraube”:NP]);  
[ce1.3 | ce1.3:**ack**(Cnst,DU1.1<sub>Cnst</sub>)])]

which, as K += K’ is short for [TmpK | TmpK is K];[ | K is TmpK;K’], is equivalent to:

[i1.2a | i1.2a:Int<sub>Cnst</sub>( [TmpK | TmpK is DU1.1<sub>Cnst</sub>];  
[ | DU1.1<sub>Cnst</sub> is TmpK; [ | up1.6:“eine Schraube”:NP];  
[ce1.3 | ce1.3:**ack**(Cnst,DU1.1<sub>Cnst</sub>)])]

(Note that acknowledgments are *not* parts of discourse units (Traum, 1994; Matheson Poesio and Traum, 2000), so (5.3.8’) is a well-formed proposition as ce1.3 is not part of DU1.1<sub>Cnst</sub>. )

Secondly, it might be argued that an intention to acknowledge is part of Cnst’s intentions even under the ‘cooperativity’ interpretation discussed above; i.e., that intention i1.2a in (5.3.4) should actually incorporate parts of (5.3.8), as follows:

(5.3.4’) [i1.2a | i1.2a:Int<sub>Cnst</sub>([DU1.2,ce1.2, ce1.3 |  
DU1.2 is [ce1.1, K1.1 |  
K1.1 is [x,e|**screw**(x),e:**grasp**(Cnst,x)],  
ce1.1:**directive**(Inst&Cnst,Cnst,K1.1),  
up1.6:“eine Schraube”:NP ]  
ce1.2:**repair**(Cnst,DU1.1<sub>Cnst</sub>, DU1.1<sub>Cnst</sub>;DU1.2),  
ce1.3:**ack**(Cnst,DU1.2)])]

The distinction is subtle, which is as it should be as interpretation seems to proceed in the same way irrespective of the ‘true’ intention. The main remaining difference between the acknowledgment interpretation in (5.3.8’) and this new repair + acknowledgment interpretation in (5.3.4’) is that in the case of (5.3.8’), the directive would be still considered Inst’s action.

Just as in the case of the cooperativity interpretation, in the case of acknowledgment interpretation in (5.3.8) Cnst would have a number of ways to realize her intention. Notice that a repetition of the partial directive would be appropriate as an acknowledgment of that part only (i.e., without making a completion). As in the case of the cooperativity explanation, to decide

how to execute the Acknowledgment Cnst would have to determine what's missing from Inst's directive - either by a structural or by a domain-level match. Either way, the result of this decision process is that Cnst acquires an intention along the lines of (5.3.6) or (5.3.8').

*Responding to a Request* As pointed out before, another possible reason for the completion is that Cnst interpreted the pause in 1.1 as a request. The system of grounding acts developed by Traum (1994) and adopted in PTT (see section 3.6) includes three types of Requests that Inst may have intended to perform: Request for Acknowledgment, Request for Continuation, and Request for Repair. In all three cases, if Cnst adopts this interpretation she will update her information state by adding an obligation to address the request (the mechanics of obligations in PTT is discussed in some detail in (Matheson Poesio and Traum, 2000)). After that, it will be this obligation (rather than cooperativity, or a voluntary desire to acknowledge) that will lead Cnst to acquire intentions such as (5.3.4) or (5.3.8). Once these intentions are acquired, however, the result will not be any different from that discussed above.

*Blurting Out* Finally, it is possible that Cnst's action is not motivated either by a decision to be cooperative at the domain level (i.e., by acquiring intention (5.3.4)) or by an intention to acknowledge (as in (5.3.8)), but by impatience – i.e., by a view that Inst is taking too long to perform the action. According to this explanation, Cnst quickly recognizes that Inst has intention i1.1b' in (5.3.3), and realizes what is missing by a structural or domain level match, just as explained above. However, Cnst is also moved by other goals, such as completing the task quickly, and it is as a result of these goals that she acquires intentions (5.3.4) or (5.3.8). We do not have a full account of what these other goals might be and how they lead to the acquisition of one of these intentions, but we do expect these intentions to be expressible in terms of the formalism assumed here.

#### **5.4 The repair: 1.3, "eine orangene mit einem Schlitz"**

Upon hearing the completion, Inst's view of the discourse situation is predicted by the analyses above to be as depicted in Figure 5.4.1.

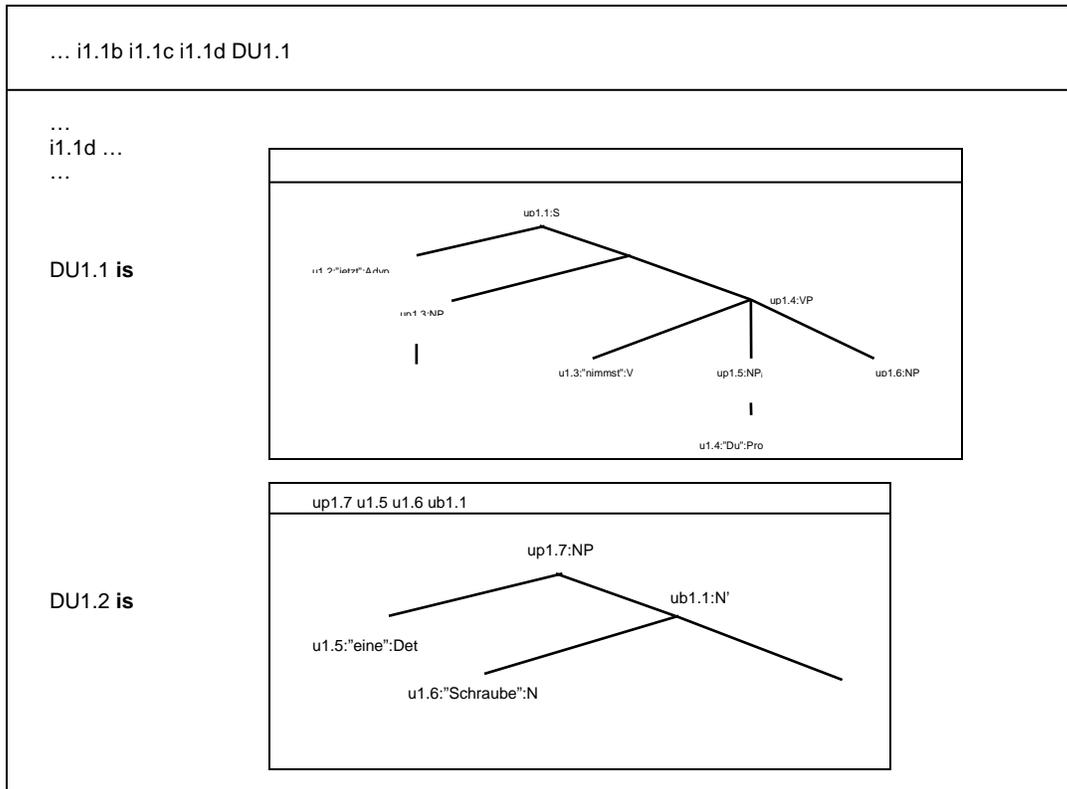


Figure 5.4.1 – Inst's view of the discourse situation after perceiving Cnst's completion in 1.2

Inst's behavior upon hearing the completion is consistent with him, first of all, recognizing that the NP uttered by Cnst is intended as a completion of the utterance Inst himself previously performed: i.e., that he is meant to update DU1.2 with the information in (5.4.1)—that the NP uttered by Cnst is meant to fill the position of the missing NP in the tree in (5.2.4):

(5.4.1) DU1.2 += [ | up1.7 is up1.6]

He then appears to ascribe to Cnst, on the basis of the information in Figure 5.4.1, one of the intentions in (5.3.4), (5.3.8), or (5.3.4'). In all of these cases Cnst intends to perform a grounding act having as one of its arguments discourse unit DU1.1 from (5.3.1) augmented with the information that a screw is needed (as well as, possibly, with changes to the directive, and/or information about the utterance). In the case of (5.3.4), the intention is of repairing DU1.1; in the case of (5.3.8), the intention is to acknowledge it; in the case of (5.3.4'), to do both. (We might call this DU the DU 'under discussion', to borrow a term from Ginzburg.)

We will proceed here under the assumption that Inst attributes to Cnst the most complex intention, that in (5.3.4')—i.e., the intention both to be cooperative and to acknowledge the contribution in DU1.1 augmented with the information that the object that Inst wants Cnst to pick up is a screw. (In this way, we will cover most of the ideas that would be needed to analyze what happens in the dialogue under the other interpretations.) As a result, Inst's view of the discourse situation is updated as intended by Cnst, as shown in (5.4.2).

(5.4.2) DU1.2 += [ce1.1, K1.1|

K1.1 is [e,x|screw(x),e:grasp(Cnst,x)],  
ce1.1:directive(Inst&Cnst,Cnst,K1.1)];

[ce1.2, ce1.3|  
 ce1.2:**repair**(Cnst,DU1.1<sub>Inst</sub>, DU1.1<sub>Inst</sub>;DU1.2),  
 ce1.3:**ack**(Cnst,DU1.1<sub>Inst</sub>)]

At this point, Inst appears to ‘accept’ the repair proposed by Cnst, but also to have decided that Cnst’s description of the screw may not ensure that Cnst chooses the correct screw (the one which we have called **orange-slit-screw**) and so decides to perform a *continuation*: an expansion of the description of the object to be grasped. Also, Inst may or may not ground DU1.1, depending on whether he recognizes an Acknowledgment.

As said in Section 3, we need two conditional update rules for grounding acts here:

UP.REPAIR ce:**repair**(A,DU1,DU2) → (**accept**(B,ce) → [ | DU1 is DU2])  
 UP.ACK ce:**ack**(A,DU) → (**accept**(B,ce) → [ | G(DU)])

(A reminder that an update rule is not a material implication, but a rule to update the information state by concatenating new information.) The result of inferring the intentions in (5.4.2), then, is that in case Inst accepts the repair he will proceed to update his information state by adding to DU1.1 the content of DU1.2, as follows:

(5.4.3) [ | DU1.1 is DU1.1;DU1.2]

And if he also recognizes and accepts an acknowledgment, he will perform the further update:

(5.4.4) [ | G(DU1.1) ]

Inst then proceeds to plan his continuation. Such expansions of an object’s description are also viewed in PTT as a type of refashioning; this time, of the DU which is the result of the repair performed by Cnst in 1.2, the just updated DU1.1<sub>Inst</sub>. With his continuation, Inst intends to repair DU1.1<sub>Inst</sub> with a new DU (called here DU1.3) in which the argument of the directive, K1.1, is updated by adding to its previous content (in 5.4.2) the additional information about the screw x which Inst believes it will be necessary in order for Cnst to identify the appropriate screw. This intention is represented in (5.4.5) as intention i1.3a of performing a **repair** ce1.5, replacing DU1.1<sub>Inst</sub> with the new DU1.3, in which K1.1 is updated with the information in K1.3.

(5.4.5) [i1.3a| i1.3a:Int<sub>Inst</sub>([DU1.3, ce1.4, ce1.5 |  
 DU1.3 is DU1.1<sub>Inst</sub> ;  
 [K1.3 | K1.3 is [ | **orange**(x)];[ x3| **slit**(x3), **has**(x,x3)]];  
 (K1.1 += K1.3),  
 ce1.4:**ack**(Inst,DU1.1<sub>Inst</sub>),  
 ce1.5:**repair**(Inst,DU1.1<sub>Inst</sub>, DU1.3)]

(Notice, incidentally, that the content of directive *ce1.1*, proposition K1.1, is not anymore the content that Inst had originally planned, shown in (5.2.5). I.e., the PTT account of this episode of collaboration in dialogue makes it this core speech act a genuine case of ‘joint production’ in the sense of Clark (1996). ) Furthermore, Inst reaches the decision to realize the planned refashioning in (5.4.5) as an apposition, as it is often the case in dialogue. This makes a treatment of appositions essential for a theory of incremental dialogue processing, which has to be based on a syntactic theory like LTAG in which such incremental construction can be modeled. We return to appositions shortly.

In the PTT framework there are at least two ways to explain how Inst reaches intention i1.3a in (5.4.5). One explanation is that this intention is the result of *plan* matching: matching Cnst’s version of the directive in (5.4.2) with Inst’s originally planned directive in (5.2.5). According to this view, Inst compared the content K1.1 of directive ce1.1 that would follow from Cnst’s repair, with the content of ce1.1 according to his own original intention, shown in (5.2.5). As a result, Inst identifies the need to produce the additional information in K1.3. A second

explanation is that Inst acquires the intention as the result of *situation* matching: Inst considers the situations that may result from the execution of the content K1.1 of the directive as proposed by Cnst, and realizes that there are 9 such situations, one for each of the screws in Figure 4.2.4. This appears to be what Pickering and Garrod (2004) have in mind. (Such simple planning can be done quite efficiently, as shown, e.g., by the work on planning in the TRAINS project (Allen and Ferguson, 1994, Ferguson et al, 1996).)

As an aside, the fact that utterance unit 1.3 was actually produced in two prosodic units suggests that Inst identifies the properties that have to be added to the description of the screw produced by Cnst in order to uniquely identify the required screw incrementally, in two steps (Levelt, 1989). First of all, Inst realizes that the property [ | **orange**(x)] is required; then, that the property of having a slit

[ x3 | **slit**(x3), **has**(x,x3)]

is also needed. Each of these two properties is going to be independently realized using an apposition, further argument for the need for a theory of appositions.

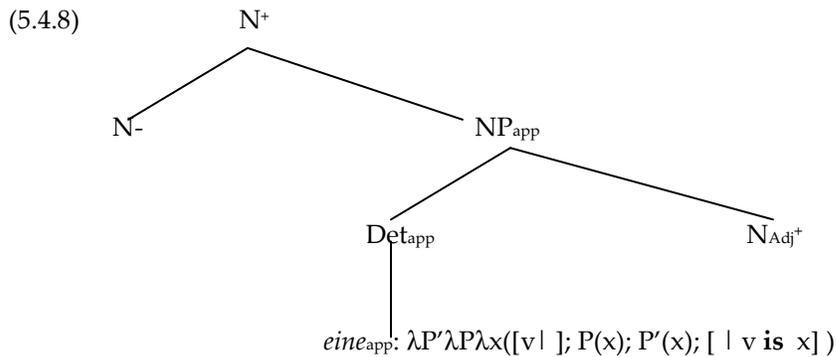
In PTT, this two-stage process would be modeled by attributing to Inst two distinct intentions: an intention to enrich K1.1 with the information that screw *x* is orange (shown in (5.4.6)) and an intention to enrich this new description with the additional information that *x* has a slit, shown in (5.4.7).

(5.4.6) [i1.3a' | i1.3a':Int<sub>Inst</sub>([DU1.3,ce1.4 |  
 DU1.3 **is** DU1.1<sub>Inst</sub> ;  
 [ K1.3 | K1.3 **is** [ | **orange**(x)]] ;  
 (K1.1 += K1.3) ,  
 ce1.4:**repair**(Inst, DU1.1<sub>Inst</sub>, DU1.3)]]

(5.4.7) [i1.3a'' | i1.3a'':Int<sub>Inst</sub>([DU1.3', ce1.4' |  
 DU1.3' **is** DU1.3 ;  
 [ K1.3' | K1.3' **is** [ x3 | **slit**(x3), **has**(x,x3)]] ;  
 [ | **orange**(x)]] ;  
 (K1.1 += K1.3') ,  
 ce1.4':**repair**(Inst, DU1.1<sub>Inst</sub>, DU1.3')

We will however simplify matters in what follows and treat the whole of 1.3 as a single apposition..

Our treatment of appositions, and the derivation of the meaning of 1.3, are discussed in Appendix B.4. Briefly, our treatment is as follows. One assumption one is forced to make, given the properties of LTAG, is that *eine* is ambiguous between the 'normal' interpretation and the form used in appositions, henceforth *eine<sub>app</sub>*, to which it is associated the auxiliary tree and lexical meaning in (5.4.8), which yields a type [[ $\pi$ ],[ $\pi$ ]] ( $\langle\langle e,t \rangle, \langle e,t \rangle\rangle$ ) for the appositional NP:



We further assume that the appositional NP is syntactically adjoined to the previous utterance, resulting in the LTAG analysis for *eine Schraube, eine orangene mit einem Schlitz* shown in (5.4.9).



(ii) to perform a grounding action with respect to its content – informally,  
 $\text{obl}(\text{Cnst}, \text{grounding-act}(\text{ce1.1}))$

There are several ways to address a directive, two of which are accepting it and rejecting it. This is modeled in PTT by assuming that (part of) the formalization of actions such as directives consists of update rules leading to the conclusion that a participant has addressed a conversational action if certain types of responses have been produced. This in turn leads to the removal of the obligation. These conditional update rules take the following form:

$\text{ce:directive}(A, B, K) \rightarrow (\text{accept}(B, \text{ce}) \rightarrow \text{address}(B, \text{ce}))$   
 $\text{ce:directive}(A, B, K) \rightarrow (\text{reject}(B, \text{ce}) \rightarrow \text{address}(B, \text{ce}))$

I.e., if A performs a directive to B, and B accepts that directive, then that counts as having addressed the obligation of addressing the directive. This removes the obligation to address the conversational action, but of course, accepting a directive then leads to a new obligation, to perform the action directed. This is formalized using another update rule which, in a schematic fashion, can be formalized as follows:

$\text{ce:directive}(A, B, K) \rightarrow (\text{accept}(B, \text{ce}) \rightarrow \text{obl}(B, K))$

The same happens with obligation (ii), the obligation to perform a grounding act. The performance of any of the grounding acts in (Traum, 1994) will discharge the obligation; the performance of any of the grounding acts will also lead to further updates – e.g., acknowledging a DU results in that contribution being grounded, as discussed in Section 3:

$\text{ack}(A, \text{DU}) \rightarrow \mathbf{G}(\text{DU})$

In summary, as a result of the performance of 1.4, the discourse situation gets updated as in (5.5.1):

(5.5.1) [ce1.6, DU1.4 |  
 $\text{ce1.6:ack}(\text{Cnst}, \text{DU1.1}_{\text{Cnst}}),$   
 $\mathbf{G}(\text{DU1.1}_{\text{Cnst}}),$   
 DU1.4 is [ce1.7, s1.1 | ce1.7 :  $\text{accept}(\text{Cnst}, \text{ce1.1}),$  s1.1 :  $\text{obl}(\text{Cnst}, \text{K1.1})$ ]  
 ]

Where  $\text{DU1.1}_{\text{Cnst}}$  is the DU resulting from the two repairs in 1.2 and 1.3:

[ce1.1, K1.1 |  
 $\text{K1.1 is [e,x,x3 | screw}(x), \text{orange}(x), \text{slit}(x3), \text{has}(x,x3),$   
 $\text{e:grasp}(\text{Cnst}, x)]$   
 $\text{ce1.1:directive}(\text{Inst}\&\text{Cnst}, \text{Cnst}, \text{K1.1})$ ]

This completes the first contribution in (1).

## 5.6 The rest of the dialogue

An analysis of the second contribution in the sample dialogue –utterances 2.1-2.4– at the level of detail just given for the first one would require a discussion of numerous additional aspects of PTT, including the treatment of anaphoric accessibility, and of a number of additional grammatical issues, such as our treatment of connectives and the German ‘Satzklammer’. As it is just not possible to present all this material in a single journal paper, we will only provide a sketchy discussion of the main issues from the point of view of coordination, and defer the additional discussion to a second paper in preparation (Poesio and Rieser, in preparation).

2.1: “Und steckst Sie dadurch, also”

With utterance 2.1 Inst moves on to the next step in the shared plan in Figure 4.2.3 and in (4.2.4), which turns into step (e) in (4.2.5). Inst wants Cnst to put the screw through the aligned holes of the longer bar serving as ‘fuselage’ and the shorter bar serving as ‘wings’ of the toy airplane. From a production / sentence planning perspective (Levelt, 1989; Dale and Reiter, 2000), there are very few differences between the steps that Inst goes through in planning utterance 2.1 and those that were involved in planning utterance 1.1. Again, Inst has the intention to perform a directive (let’s call it *ce2.1*); and again, Inst decides to plan an utterance *up2.1* generating directive *ce2.1*. (As said above, a full analysis of the particular realization of *ce2.1* chosen by Inst for the example dialogue would require a discussion of three linguistic phenomena: rhetorical relations (*und, also*), the so-called ‘Satzklammer’ (*steckst ... dadurch*) and the pronoun *Sie*, that raises the issue of anaphoric accessibility within micro conversational events. We defer this discussion to a second paper.) And again, Cnst interrupts Inst before the plan for *up2.1* is completely executed.

#### 2.2: “von oben”

What happens at this point is very similar to what happened during the first contribution when Cnst performed the completion in 1.2, and we believe our analysis of that completion applies with relatively few modifications to this case as well, except that there is very little evidence in this case that anything in Inst’s performance could be interpreted by Cnst as a request, so the other explanations (Cooperativity, Voluntary coordination control, and blurring out) look more plausible in this case than the Request for Acknowledgment. In fact, in this case our intuition tells us that there is a stronger sense that Cnst is simultaneously acknowledging Inst’s contribution and refashioning it. The main issue raised by this utterance is one of compositional semantics: how the meaning of the modifier *von oben* is integrated with the meaning of Inst’s contribution. In the later paper we discuss two options: maintaining the simple treatment of VPs as having type <e,t> used so far, and going to the higher <s, <e, t>> type common in work on tense based on a Parsonian semantics (Parsons, 1990).

#### 2.3: “von oben, daß also die drei festgeschraubt werden”

With this utterance, Inst achieves two objectives. First of all, he acknowledges Cnst’s participation to the current DU—this time by explicitly repeating Cnst’s utterance. He then returns to the utterance plan which had been interrupted by 2.2, and continues the production of the purpose clause. The purpose clause also contains a second example of (in this case, deictic) reference: the reference to the three constituents of the airplane previously aligned.

#### 2.4: “Ja”

Finally, the *Ja* in 2.4 is produced (and interpreted) more or less like 1.4. We will just briefly remark that Cnst seems not to have had any problems interpreting the definite description *die drei*.

## 6 A non-intentional analysis

Pickering and Garrod (2004) while encouraging psycholinguists to take up dialogue as a valid area of research, argued that many aspects of dialogue interpretation—particularly *priming* at the lexical, syntactic, and referential level (Morton, 1969; Branigan et al, 2000)—can be explained without resorting to an intentional analysis, and sketched instead a model of interaction centered around the notion of ALIGNMENT between the dialogue participants' models. Although Pickering and Garrod's alignment model is more of a programmatic proposal than a fully worked out theory, it has proven very influential (Purver et al, 2006; Rickheit & Wachsmuth, 2008). As a contribution to the ongoing discussion concerning this model, in this section we briefly discuss which aspects of cooperations could be accounted by an alignment model, and how.

### 6.1 The Pickering and Garrod alignment model

The starting point of Pickering and Garrod's analysis is the consideration that successful communication between dialogue participants A and B requires them to go through the same stages of interpretation and production, illustrated in Figure 6.1.

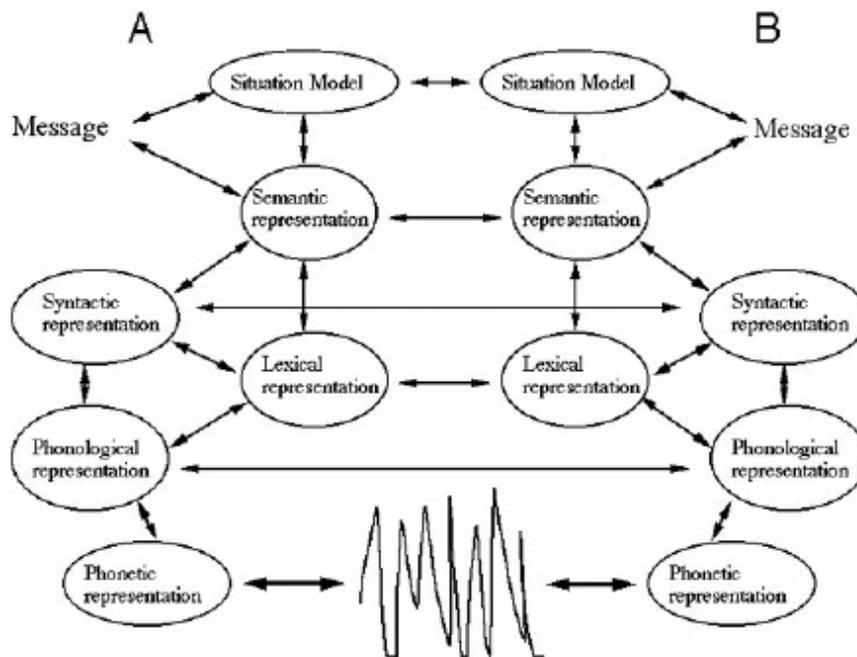


Figure 6.1: Different representations involved in (spoken) language comprehension. (From (Pickering and Garrod, 2004); reprinted with permission.)

Pickering and Garrod (henceforth: P&G) also note that priming effects can be observed at each of these levels of interpretation, from the phonetic (Bard et al, 2000) to the lexical (Garrod and Anderson, 1987) to the syntactic (Bock, 1986); Branigan et al, 2000). Finally, P&G discuss experiments suggesting that in order for dialogue to be felicitous, the participants need to produce aligned 'situation models'. The term 'situation model' is widely used in the psycholinguistic literature on reading and text processing to indicate a variety of representations expressing information about the events and entities described in a narrative (theories of

situation models have been proposed in, e.g., (Sanford and Garrod, 1981; Johnson-Laird, 1983; Garnham, 2000; OTHER REFERENCES). But what P&G have in mind is a much more general notion of situation model—and one which may change from task to task, often being developed on the spot by the participants to the task. For example, in one of the studies they discuss, by Garrod and Anderson (1987), the participants to a ‘cooperative maze game’ have to develop ways of describing each other’s position in a maze in order to follow a route. The subjects were found to develop ‘aligned’ views of the maze in the sense that they would jointly develop a protocol for referring to maze positions, and particular positions, confirming the results of work of Clark and colleagues (Clark and Wilkes-Gibbs, 1986; Clark, 1996).

The conclusion that Pickering and Garrod draw from these observations is, first, that ‘aligning representations’ at all levels is a central aspect of dialogue processing; and that this alignment is the result of fairly simple comprehension mechanisms, rather than of intention-based mental state reasoning. (Such ‘more complex’ types of processing are considered as an option, but not the basic form of processing available to dialogue participants.)

A phenomenon that, according to P&G, can be predicted from the Interactive Alignment Models is the creation of ROUTINES. P&G point out that dialogue is highly repetitive: e.g., Aijmer (1996) suggests that up to 70% of words in the London-Lund corpus occur as part of recurrent sequences of words. According to P&G, this follows from the alignment assumption: during production, participants in the dialogue will tend to choose expressions or syntactic structures which have already been produced as part of the present (or, perhaps, earlier) dialogues.

## **6.2 An alignment-based analysis of the example dialogue**

What would an account of sentence cooperations according to the Interactive Alignment model look like? We find it useful to divide this question in two parts: explaining how Cnst reaches her conclusion about what Inst is trying to achieve; and explaining how she decides to produce a completion. (We concentrate on the completion.)

### **6.2.1 Interpreting Inst’s actions**

The first question that has to be addressed is: what would the situation model be in a task in which the participants are not simply describing reality, but modifying it via actions? We think at least two answers are possible:

(i) the situation model is a representation of the state that Cnst has reached in assembling the toy plane: i.e., what has already been assembled, and which parts are still available.

(ii) the situation model is the shared plan itself: i.e., which of the required actions have already been performed, and which still have to be. Most, if not all, theories of planning view plans such as those discussed in the previous sections as representations—typically, tree structures—which can be manipulated by, e.g., decomposing an action in its subactions. (In fact, the idea that plans can be viewed as mental constructs is fairly recent, dating back to Pollack’s dissertation (1986).

If we assume for this task a situation model of the second type, the process by which Cnst would recognize Inst’s actions under the IAM would not be so very different from the process discussed above for the intentional model—it would simply require Cnst to find a matching action in the plan, as done, e.g., by Kautz (1987). We concentrate therefore on explaining how Cnst could reach such a conclusion if the situation model was of type (i).

One obvious difficulty is that Inst and Cnst’s situation models in this sense are not really ‘globally aligned’ in the sense of being identical until the very end of the conversation, as both

Inst and Cnst are perfectly aware. It is however possible to argue that achieving the desired alignment is the entire point of instruction-giving conversations, which are not therefore an exception to the IA model.

Coming to what could be a 'priming' alternative to recognizing Inst's intentions by looking at a shared plan, we believe that a candidate can be found by generalizing the notion of 'routine' to non-linguistic actions. Namely, it could be argued that during a typical conversation of the BTPC, an AGGREGATE FORMATION ROUTINE is established: a repeated pattern of actions which together lead to the same result (namely, two parts of the toy airplane get joined). In the dialogue from which (1) is excerpted, Inst first instructs Cnst to join the 'side rudder' to the 'fuselage'; then to join the tail to the fuselage; etc. After a while, Cnst may be able to abstract over these repeated actions and produce the routine in Figure 6.2:

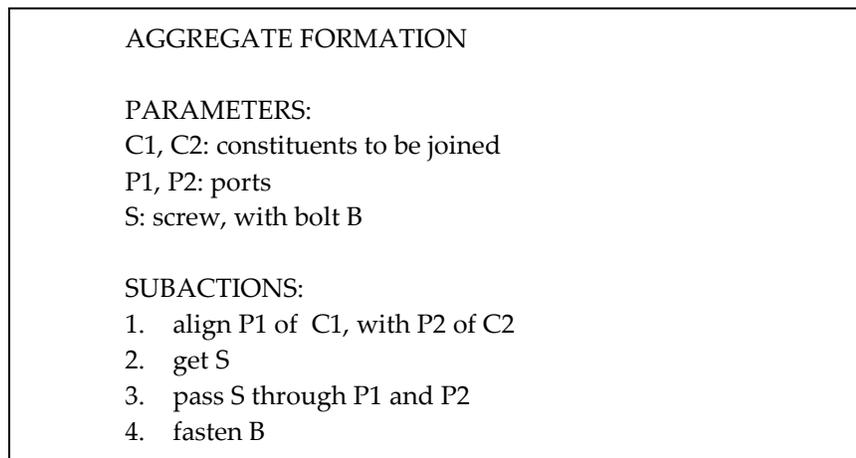


Figure 6.2: routine created during a toy plane assemble dialogue

Once this routine becomes part of the common ground, an instance can be recognized through simple structural match, without need for explicitly reasoning with intentions. In the case under consideration, for example, Cnst may be able to recognize that *nimmst* is an instance of the second step of the routine, and hence, that the missing argument must be a screw. She may also be able to identify the syntactic structure to use for the completion by looking at the realization of previous mentions of this type of object.

### 6.2.2 Deciding to produce a completion

As the sketchy discussion above makes clear, in our opinion it would be possible to develop an alignment account of interpretation. What is not at all clear is how such a model could explain what happens next: i.e., why is it that Cnst decides to produce a completion, instead of waiting for Inst to finish his contribution. In other words, the alignment model as currently stated does not provide an account of dialogue management, an important aspect of what dialogue participants have to do (and one which had a significant influence in the design of PTT as currently conceived).

## 6.3 Conclusions

'Traditional', Grice-inspired models of dialogue processing tend to formalize all aspects of an agent's behavior when involved in a dialogue in terms of intentions (see, e.g., Cohen's account of reference (Cohen, 1979) or Grosz and Sidner's theory of discourse structure (Grosz and Sidner,

1986)). We feel that the alignment model offers a very reasonable alternative to these Gricean models as far as interpretation and production are concerned. However, intentions also play an important role in formalizations of dialogue management—how an agent decides what to do next—and we can't find anything in the alignment model as currently stated which would explain how Cnst reaches her decision to produce a completion, nor to explain what is it that Inst is trying to do (a non-intentional alternative to speech acts, that is).

## 7 Comparisons with related literature

To our knowledge the only other proposal explicitly concerned with completions in the literature is the paper by Purver et al (2005), developed in the Dynamic Syntax framework (Kempson et al, Cann et al). We discuss Dynamic Syntax and Purver et al's proposal in this section, after however discussing the two other main theoretical frameworks for the semantics of dialogue, Ginzburg's KOS (Ginzburg and Sag, 2000; Ginzburg, to appear) and Asher and Lascarides' SDRT (Asher and Lascarides, 2003), both of which provide theoretical insights that have been incorporated in our framework and that could perhaps be extended to provide a treatment of the phenomenon under consideration.

### 7.1 Ginzburg's theory of the Dialogue Game Board

Arguably, the most developed modern theory of semantic interpretation in dialogue is Ginzburg's KOS framework (Ginzburg and Sag, 2000; Ginzburg, to appear). The theory was conceived from the start to deal with dialogue, and provides the widest coverage of dialogue-specific phenomena, including in particular grounding, clarification requests and non-sentential utterances; therefore it could probably be extended to provide an account of completions, at least for what concerns their interpretation.

#### 7.1.1 General features of the model

Like PTT, KOS is inspired by Situation Semantics (Barwise & Perry, 1983) and is founded on the assumption that many aspects of interpretation in dialogue can only be properly explained by providing an account of the discourse situation in its entirety. Another similarity with PTT is that KOS is based on an information state view of dialogue interpretation—and as in PTT, it is assumed that each participant's information state – their **dialogue gameboard** (DGB)—contains both public and private information. In its basic form, a participant's information state according to KOS contains a record of **Speaker**, **Addressee**, the **Moves** so far –with a distinguished slot recording the **LatestMove**-- the (shared) **Facts** about the discourse situation, and a record of the Questions Under Discussion, or **QUD**. The information state is a record in Type Theory with Records (Cooper 2005).

The QUD slot is one of the most distinctive features of this framework. It holds a stack (or perhaps a partial order) of the **issues** raised in a conversation. QUD was hypothesized to account for the pragmatics of queries – and especially to explain how the propositional interpretation of short non-sentential answers is recovered—but, as we will see, it also plays a central role in KOS's treatment of grounding. Simplifying matters considerably, when A asks B a query such as "Who came?", the DGB of both participants gets updated with the information that the latest move was a query: **Ask**(A,B,  $\lambda x. \mathbf{come}(x)$ ). As a result, the issue raised by the query—in the case of a wh-query, an abstract:  $\lambda x. \mathbf{come}(x)$ —gets added to the QUD, as shown in Figure 7.1.1.

Spkr: A Addr: B Moves: ⟨...⟩ LatestMove: <b>Ask</b> (A,B, $\lambda x. \mathbf{come}(x)$ ) Facts: ⟨...⟩ QUD: ⟨ $\lambda x. \mathbf{come}(x)$ ⟩
--

Figure 7.1.1. Update of the DGB after asking a question.

According to KOS, the presence of this issue in QUD is what enables B to provide a short answer to the question: say, “John”. The content  $j$  of this short answer can be combined with the content of the question that is maximal in QUD to get a proposition, “John left”. When grounded, this proposition is added to Facts.

The effect of dialogue acts on the DGB—i.e., how the information state is updated as a result of, e.g., questions and answers—is modeled by **conversational rules**. A conversational rule specifies how information states that met certain preconditions, such as the presence in LatestMove of an **Ask** dialogue act, are updated as a result. For instance, the incrementation of QUD with the issue resulting from a question is specified by a conversational rule called **Ask QUD-incrementation** (Ginzburg 2008, p. 117). The overall series of updates to the information states of all conversational participants resulting from a change in the discourse situation is called a **protocol**; an example of protocol is the protocol for **Cooperative query exchange** specified below (Ginzburg 2008, p. 116).

#### Conversational query exchange

1. LatestMove.Cont = Ask(A,q)
2. A: push A onto QUD; release turn
3. B: push Q onto QUD; take turn; make q-specific utterance.

As this example shows, the updates specified by a protocol generally differ depending on the role of the conversational participant. After asking the question A will (normally) release the turn, whereas B will take it and perform a relevant utterance.

As well as the most extensive treatment of dialogue phenomena, KOS also comes with the most detailed fragment of any theory of dialogue, the HPSG grammar from Ginzburg and Sag (2000), subsequently modified in substantial ways in (Purver, 2004; Fernandez, 2006; Ginzburg, 2008), *inter alia*.

#### 7.1.2 Grounding and clarifications

(Ginzburg, 2008) also includes a full treatment of the grounding process. The crucial advance of this proposal is that it provides a detailed account not just of successful communication, but also of *failure* in communication and how it is handled via clarification requests. Ginzburg’s book also contains an interesting theory regarding what it means to understand an utterance, and provides a formalization of the grounding criterion based on this theory.

In its simplest form, the KOS account of the process by which the DGB is updated as a result of utterance  $u$  is as follows:

1. If the utterance is understood, LatestMove is updated with  $u$  (and the process of update continues as discussed above);
2. If the utterance is not completely understood, so-called **crification** ensues.

Note how utterances are not immediately added to LatestMove upon being perceived. Instead, they go to an information state slot called PENDING, that contains ungrounded utterances (PENDING is very similar to the slot called UDU by Matheson Poesio & Traum (2000).) From there, utterances go to LatestMove if grounded—but may also be revised as a result of the crification process.

The content of PENDING are what Ginzburg calls **locutionary propositions** (p. 148--149). A **proposition** in TTR is a type of record with the two slots

[SIT = S, SIT-TYPE = ST]

For which truth conditions are defined: the proposition above is **true** iff S is of type ST, written S : ST (Ginzburg 2008, section 2.5.4). In locutionary propositions, SIT-TYPE is the utterance type resulting from parsing, whereas SIT is the utterance. The situation type, in turn, is a record of type SIGN that encodes the same information encoded by signs in the earlier (Ginzburg and Sag, 2000): phonetic information in the PHON field, syntactic category in the field CAT, constituency information in the field CONSTITS, and semantic information. As in HPSG, the meaning of an utterance in KOS is viewed as a function from contexts to contents, where contexts are a generalization of the notion of context due to Kaplan (1989): whereas Kaplan's contexts were limited to a few values for indexicals like *i*, *you*, *here*, *now*, in the theory of contexts developed in Situation Semantics, particularly in the work of Robin Cooper (Gawron and Peters, 1990; Cooper and Poesio, 1994; Cooper, XXXX) a great deal of aspects of the interpretation are sought in contexts, including referents for proper names. Hence, semantic information is encoded in two fields: the field C-PARAMS specifies the **contextual parameters** of a semantic object that have to be supplied by the context for the object to be interpretable, whereas the field CONT specifies content proper—how these values and those supplied by lexical semantics are combined. For instance, the locutionary proposition corresponding to discourse unit DU1 in (3.4.3) would be as follows:

(7.1.1) [SIT-TYPE =

[ PHON = "there is an engine at Avon",

CAT = S : syncat,

CONSTITS = { there, is, an, engine, at, Avon, an engine, at Avon } : set (sign),

C-PARAMS = [ spkr: IND, addr: IND, s0: SIT, l: LOC,

a: IND, c1: named(a, "Avon")],

CONT = **assert**(spkr, addr, [SIT = s0,

SIT-TYPE = [x : Ind, c1 = **engine**(x), c2 = **at**(x,a)])]

SIT = [PHON = "derisanenginatevn",

CAT = S: syncat,

CONSTITS = { u1(der), u2(is), u3(an), u4(engine), u5(at), u6(evn), u7(anengin),  
u8(atevn) },

C-PARAMS = [ spkr = A, addr = B, s0 = sit1, l = l1, a = avon, c1 = prop1]

CONT = ]

Comparing these propositions with the PTT encoding of the same contribution, it is clear that leaving aside the semantic differences between records and DRSs—potentially significant, but not entirely clear to us—the first main difference is that whereas in PTT the representation of the information state is more akin to what in Situation Semantics would be called a situation type, just as in DRT and SDRT, the information state in KOS consists of propositions specifying that indexical situational objects are of a certain situation type. It might be argued that the KOS view is more natural, and a similar view of 'anchored DRT' has also been advocated for DRT by Kamp e.g., in (Kamp, 1990)—but the semantic differences become less clear with Compositional DRT, in

which, unlike in vanilla DRT, proper names and indexicals are treated as directly referential (i.e., are encoded as constants), not as existentially quantified.

The one distinction which clearly seems to make a difference is the explicit encoding of situational parameters in KOS. This explicitness is the basis for the formulation of the **grounding criterion** proposed by Ginzburg and colleagues: that understanding an utterance means finding values for the contextually dependent parameters in C-PARAMS within the DGB (Ginzburg, 2008, p. 153). In case of failure to ground—for instance, in case the referent of Avon is unknown, or in case the addressee does not understand what is the relevant meaning of *engine* for this context—crification ensures. By contrast, in PTT we have failure of understanding in case an utterance can't be assigned a meaning. The KOS treatment is more general, offering for instance the potential to identify failure of understanding also in the case of failure of intention recognition, i.e., when no dialogue act can be assigned to an utterance. Encoding contextual parameters in CDRT would not be particularly problematic, as they could be viewed as distinguished discourse referents (as indeed they were in (Cooper and Poesio, 1994)) but no attempt to compare the two types of grounding criterion and / or develop a version of PTT with contextual parameters has been made so far.

The treatment of clarification requests in KOS is based on the extensive empirical analysis of clarifications in the British National Corpus carried out by Purver (2004). Purver classified clarification requests with respect to their form (identifying eight distinct types: wot, explicit, literal reprise, wh-substituted reprise, reprise sluice, reprise fragment, gap, and filler - fillers corresponding to what we are calling here completions) and their content (identifying four distinct categories: repetition, clausal confirmation, intended content, and correction); many of these are treated in (Ginzburg and Cooper, 2004; Ginzburg, 2008).

According to KOS, the process by which contextual parameters are instantiated begins by applying a **Contextual Instantiation** update rule (p. 153, 154) which updates the information state by 'filling in' parameters as soon as suitable instantiations are found. In case the instantiation of the parameters is only partial, **Clarification Context Update Rules** apply, of which two are discussed: Parameter Identification and Parameter Focussing. **Parameter identification** applies whenever the value of one of the contextual parameters is missing and results in questions that can be paraphrased as "what is the intended content of  $u_i$ ?" (e.g., "what do you mean by Avon" or "Who is Bill?"). In KOS terms, asking such a question amounts to putting as maximal in QUD a question of the form  $\lambda x \text{Mean}(\text{sprk}, u_i, x)$ . **Parameter focusing** produces confirmation requests on the basis of the value of the CONT slot. In both cases, the process simply involves putting an appropriate question on QUD—in other words, clarification requests are supposed to work just like any other request.

With these refinements, the process by which the information state is updated can be specified by the more complex protocol that follows (Ginzburg, 2008, p. 180 – (96)):

**Utterance Processing Protocol** For agent A with IS I: if a locutionary proposition

$$lp = [ \text{sit} = u, \text{sit-type} = T ]$$

is maximal in PENDING:

- (a) if  $p$  is true, try to integrate  $p$  in A.DGB using a Moves update rule
- (b) Otherwise: try to accommodate  $p$  as a CR to LatestMove
- (c) Else: seek a witness for  $T$  by asking a Clarification Request.

### 7.1.3 Non-sentential utterances, self-initiated repair and incremental pending

In Chapters 6 and 7 of (Ginzburg, 2008), two ingredients are added that make the KOS framework appropriate for providing a treatment of completions. In Chapter 6 an extensive treatment of Non-Sentential Utterances is provided, based on the work of Fernandez (2006), covering in particular short answers (*yes, no, John*) and sluices of various types (as in, e.g., A: *John left* B: *Who?*). That chapter discusses in detail how utterances providing a complete dialogue act can update and be interpreted with respect to the DGB even when syntactically they are not sentential. In Chapter 7, and in particular in Section 7.3, the possibility that PENDING can be incrementally updated (after every micro conversational event) is considered, which would provide a unified treatment of self-repair. Such moves would make KOS and PTT even more similar, making it virtually certain that solutions developed within one framework could be readily imported in the other.

### 7.1.4 Summary of the discussion

To our knowledge no explicit treatment of completions has been provided in KOS so far, but we don't see any reason why the analysis provided in this paper could not be recast in the KOS framework, once the differences between a treatment based on TTR and one based on DRT are better understood. The main difference between the two frameworks is that PTT attempts to stay closer to mainstream formal semantics in terms of syntactic and logical underpinnings, whereas KOS is closer to the unification-based world using HPSG as syntactic framework and TTR as semantic representation (hence, arguably, closer to treatments of dialogue in computational linguistics). It's not clear to us however that such differences matter much in terms of the proposed treatment of completions; on the other end, the two frameworks share the basic assumptions about information state, common ground, and their update that we argued are crucial to the treatment provided here.

Arguably, the most interesting difference between the frameworks is in the grounding criterion, formulated in terms of lack of meaning for utterances in PTT, and of contextual parameters in KOS. This is one area where the treatment proposed in KOS could be adopted in PTT.

## 7.2 Asher and Lascarides' Logic of Conversation

### 7.2.1 A general comparison between the design philosophies of LOC and PTT

Asher and Lascarides' (2003) monograph contains a wealth of insights concerning the structure of dialogue and its systematic description. It is based on interesting decisions concerning the division of labour of, *inter alia*, information content, lexicon, cognitive states, and world knowledge. The formal apparatus is developed in great detail, integrating much fashionable research in logics, linguistics, philosophy and AI, in a way echoing the "Zeitgeist" in an attractive manner.

An obvious similarity between LOC and PTT concerns the underlying logic: both frameworks are derived from DRT, of which they share the general logic assumptions about the dynamic nature of the common ground. PTT is more conservative, being based on Compositional DRT and hence, ultimately, on type theory in the sense of Montague. LOC is less traditional, both because meaning is characterized in terms of a logic more closely related to DRT as formulated, say, in (Kamp and Reyle, 1993), and, above all, because a substantial effort has been made in LOC to 'modularize' the process of interpretation and to identify distinct logical

frameworks to characterize, for instance, semantic interpretation as opposed to meaning proper. No such effort has been made in PTT.

Another big difference, already pointed out in (Asher and Lascarides 2003, p. 304), is that PTT is speech-act and intention based, whereas LOC's theory of coherence rests on the concept of discourse relations such as *Elaboration*, *Narration* and the like and on various axioms like *Contrast* guaranteeing/explaining local coherence. In PTT discourse relations and the like have played a very minor role so far, although argumentation acts are proposed in (Poesio and Traum, 1997). This is partly due to different preferences concerning the data used for setting up dialogue theory: LOC sticks to the philosophy of language principle of using prototypical examples and pays less attention to the mapping from NL to logical forms, i.e. SDRs, which is more or less taken for granted, whereas PTT proceeds from natural data, especially from task-oriented dialogue, and tries to preserve NL surface using CDRT-style translations. So there are different guiding methodological principles here. Discourse relations are a heavy machinery resting on rich intuitions concerning discourse structure and PTT is minimalistic in this respect. In the long run, PTT needs of course a theory of rhetorical structure--indeed, this is already required for the example transcript; this will be discussed in the follow-up paper. By contrast, in LOC the idea of plan-basedness is boiled down to speech-act related goals (SARGs), there are no large shared partial plan structures used; large structures emerge solely *via* discourse relations encapsulating recursively built up material.

### 7.2.2 Completions in LOC

Asher and Lascarides state quite clearly that coordination of the sort discussed in this paper is beyond the scope of their book—e.g., when discussing their example (4), attributed to Sacks, on p. 297, Asher and Lascarides say (footnote 3): 'Note that Henry's and Mels's utterances contribute a *single* proposition. We gloss over this here, however'

- (4) a. Joe: We were having an automobile discussion ....  
b. Henry: discussing the psychological motives for  
c. Mel: drag racing in the streets.

Nevertheless, LOC has a lot to offer concerning the description of our example, repeated below, and appropriate extensions are not difficult to envisage. We will investigate this in some detail in this section.

(7.2.1)

#### Dialogue example for coordinated utterance production

Inst: So, jetzt nimmst du  
*Well, now you grasp*  
Cnst: eine Schraube  
*a screw,*  
Inst: eine <-> orangene mit einem Schlitz.  
*an <-> orange one with a slit*  
Cnst: Ja.  
*Yes.*

*Coordination in the production of single speech acts.*

Two ingredients seem missing from LOC in order to provide an account of such examples. In example (7.2.1) we have two cooperative dialogue moves, Cnst's production of *a screw* and Inst's other repair *an <-> orange one with a slit*, which is acknowledged by Cnst. To handle cooperative productions one needs some sort of representational underpinning, such as joint intentions, or at least joint plan structures. In the version of PTT proposed in this paper shared plans, common intentions and common situations help to explain why alignment wrt completions and continuations occurs. Misalignment arises simply as a consequence of the difference between shared plans and individual plans. In LOC there is no real basis for the explanation of alignment, due to the lack of a principled treatment of fine grained coordination, despite the use of a cooperativity principle and claims to the contrary (p. 417). If one does not handle aligned structures, one misses a central feature of NL dialogue and neglects established insights, e.g. Clark's track-theory or his presentation-acceptance cycle for common productions. Asher and Lascarides' notion of correction, which might be considered of relevance here at first sight (LOC, pp. 345-354), solely rests on intuitions concerning denial and consistency, which is clearly not what one needs for the repair in (Ex.1), because there is no inconsistency there.

Furthermore, the decision to abstract away from explicit representation of utterances and speech acts means that some other mechanism will have to be introduced in order to account for non-sentential information state updates.

*Indirect Speech Acts and Acknowledgement in LOC*

Nevertheless, if we abstract from completion and coordination and take an idealised version of Ex. 1, 1', we can profitably ask what LOC has to offer for that. As seen in Sections 3 and 5, PTT has incorporated several general insights from SDRT concerning the representation of the content of dialogue acts, and more have been included in PTT's treatment of anaphoric accessibility, discussed in the follow-up paper. In this section we discuss whether specific proposals about indirect speech acts and acknowledgments in SDRT apply to such examples.

(1') Inst & Cnst: *Well, now you grasp a screw, an <-> orange one with a slit.*  
Cnst: *Yes.*

The mundane intuitions concerning an idealized (1') which does not observe completion and coordination as phenomena *sui generis* are that Inst and Const's turn is a request and Const's *Yes* functions either as an acknowledgement, an acceptance act or an indication of the action carried out. We concentrate on the issues of request and acknowledgement here. Before we can start that, a rough picture about the global theory levels going into LOC might be of help (see Fig. 1):

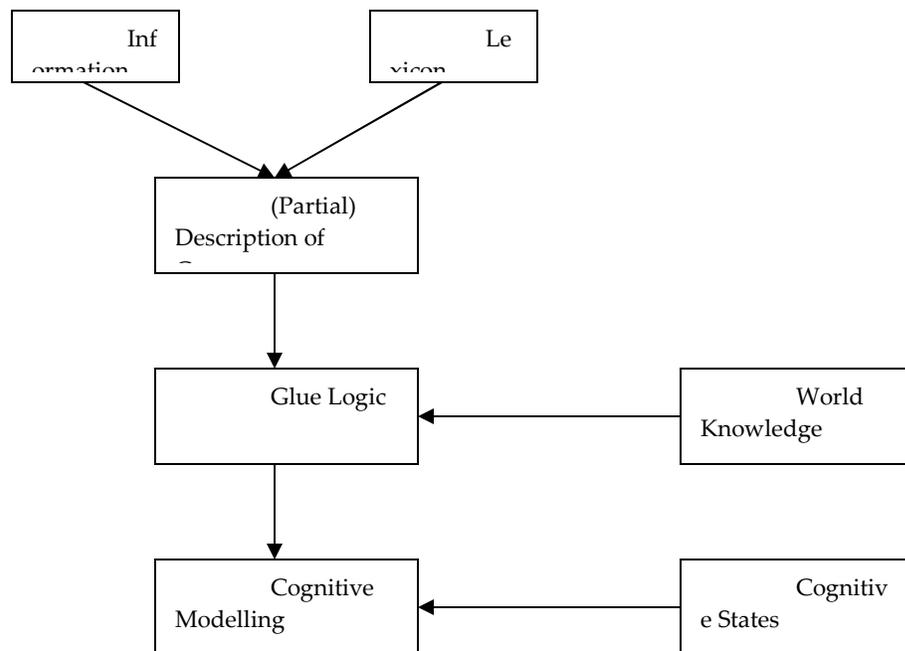


Fig. 1: Interaction of logical modules in LOC (simplified, cf. Asher&Lascarides (2003), p. 431)

Assuming that the first turn represents an unconventionalised indirect speech act, ISA (LOC, pp. 307-311), Gricean style reasoning leads to the inference of an implicit speech act. According to LOC we then get a complex type *assertion • request* for the first turn. Complex types can be expanded either way, here we assume an expansion of *request*. Where are we in the complex modularised structure in Fig. 1 after having applied Gricean style reasoning? Presumably within *Cognitive Modelling*. Grammar alone will give us of course only an SDRS for the declarative sentence, which we set up in “flat Dutch DRS-representation style” as follows:

$$[x, y \mid \text{grasp}(\text{you}, x), \text{orange}(x), \text{screw}(x), \text{slit}(y), \text{with}(y,x)].$$

What we want to have is, however, a request, i.e. it is the grasping that should be done. In LOC-style notation this is ( $\delta$  for *do* and the DRS for the proposition which is requested):

$$\delta([x, y \mid \text{grasp}(\text{you}, x), \text{orange}(x), \text{screw}(x), \text{slit}(y), \text{with}(y,x)]).$$

How one can proceed from grammatical information to cognitive modelling wrt non-conventionalised indirect speech acts is not explained in detail in LOC.

So, let us look what we can do with  $\delta([x, y \mid \text{grasp}(\text{you}, x), \text{orange}(x), \text{screw}(x), \text{slit}(y), \text{with}(y,x)])$ , i.e. which axioms of LOC apply.

Due to the default **Request Related Goals (RRG)** (LOC, p. 394) we arrive at:

$$[x, y \mid \text{grasp}(\text{you}, x), \text{orange}(x), \text{screw}(x), \text{slit}(y), \text{with}(y,x)] ! >$$

$$I_S[x, y \mid \text{grasp}(\text{you}, x), \text{orange}(x), \text{screw}(x), \text{slit}(y), \text{with}(y,x)] (A[x, y \mid \text{grasp}(\text{you}, x), \text{orange}(x), \text{screw}(x), \text{slit}(y), \text{with}(y,x)]).$$

It says: If a request  $\alpha$  is made, then the speaker's goal is typically that the action  $a_\alpha$  it denotes be performed. This is indicated by the intention operator  $I_{S(\alpha)}$ .

We then need **Cooperativity**, condition (a) (LOC, p. 391), in order to model agents' take-over of intentions:

$$(a) \quad I_{\mathcal{A}}(\delta([\mathbf{x}, \mathbf{y} \mid \text{grasp}(\text{you}, \mathbf{x}), \text{orange}(\mathbf{x}), \text{screw}(\mathbf{x}), \text{slit}(\mathbf{y}), \text{with}(\mathbf{y}, \mathbf{x})])) > I_{\mathcal{B}}(\delta([\mathbf{x}, \mathbf{y} \mid \text{grasp}(\text{you}, \mathbf{x}), \text{orange}(\mathbf{x}), \text{screw}(\mathbf{x}), \text{slit}(\mathbf{y}), \text{with}(\mathbf{y}, \mathbf{x})]))$$

$\mathcal{A} = \text{Inst}$  and  $\mathcal{B} = \text{Cnst}$ . Following the commentary LOC, p. 394, we can replace  $I_{\mathcal{S}}([\mathbf{x}, \mathbf{y} \mid \text{grasp}(\text{you}, \mathbf{x}), \text{orange}(\mathbf{x}), \text{screw}(\mathbf{x}), \text{slit}(\mathbf{y}), \text{with}(\mathbf{y}, \mathbf{x})])$

by

$\text{SARG}_{\text{Inst}}([\mathbf{x}, \mathbf{y} \mid \text{grasp}(\text{you}, \mathbf{x}), \text{orange}(\mathbf{x}), \text{screw}(\mathbf{x}), \text{slit}(\mathbf{y}), \text{with}(\mathbf{y}, \mathbf{x})], \text{Done}(\mathbf{a}_{[\mathbf{x}, \mathbf{y} \mid \text{grasp}(\text{you}, \mathbf{x}), \text{orange}(\mathbf{x}), \text{screw}(\mathbf{x}), \text{slit}(\mathbf{y}), \text{with}(\mathbf{y}, \mathbf{x})]}))$ .

In addition, we can assume that Cnst's *Yes* entails that he has accepted or achieved  $\text{Inst}_{[\mathbf{x}, \mathbf{y} \mid \text{grasp}(\text{you}, \mathbf{x}), \text{orange}(\mathbf{x}), \text{screw}(\mathbf{x}), \text{slit}(\mathbf{y}), \text{with}(\mathbf{y}, \mathbf{x})]}$ 's SARG of  $([\mathbf{x}, \mathbf{y} \mid \text{grasp}(\text{you}, \mathbf{x}), \text{orange}(\mathbf{x}), \text{screw}(\mathbf{x}), \text{slit}(\mathbf{y}), \text{with}(\mathbf{y}, \mathbf{x})])$ .

This being the case, the two turns, Inst&Cnst's *Well, now you grasp a screw, an orange one with a slit* and Cnst's *Yes*, satisfy the semantic conditions of **Acknowledgement( $\alpha, \beta$ )** (LOC, p. 466).

*Summary of argument*

LOC does not treat agents' coordination in the production of single speech acts and is therefore not directly applicable. However, if one fuses Inst's and Cnst's contributions as in (1'), one can apply LOC's indirect speech act analysis and derive the speech-act related goal of Inst's, namely that he wants Cnst to grasp a yellow slit bolt. In addition, one gets Cnst's take over of Inst's intention due to cooperativity as well as his acknowledgement.

## 7.3 Dynamic Syntax

### 7.3.1 The Dynamic Syntax Framework

The paradigms from psycholinguistics, linguistics and philosophy we have discussed so far all started from the assumption that agents' joint verbal contributions have to be treated within a theory of dialogue. The DS framework, first summarized in Kempson *et al.* (2001) started with the aim to 'characterize structural properties of language in terms of the incremental process of building up interpretation from the sequence of words' (Kempson *et al.* (2001), p. ix). Still, the empirical boundary of this endeavour was the sentence and, as far as we can see from the published work, at least in the beginning no extension toward dialogues such as e.g. the modelling of turn-constructional units or sequences of turns had been considered. However, in the course of the development of the paradigm since 2001 it turned out that DS modelling techniques for intra-sentential regularities such as anaphora, ellipses, relative clause construction or appositions generalize naturally to similar phenomena in dialogue. Incrementality is, of course, the main methodological assumption DS shares with PTT (see Section 3 and Appendix B.2 in this paper). Meanwhile (see Gargett *et al.* (2008)) the state in DS theorizing appears to be as

follows: There is a deep foundational debate among DS scholars whether some phenomena like, say, corrections or completions presuppose for their treatment a special paradigm, i.e. a dialogical one or whether these can be treated in a theory which is neutral *vis à vis* the sentence, monologue or dialogue distinction. The data of interest in this context are non-repetitive fragment forms of acknowledgements, clarifications and corrections. Their example (7.3.1) below is very near the corpus data concerning completions which we discuss in this paper:

(7.3.1) A: Are you left or  
B: Right-handed.

There is a deep methodological issue here, namely which phenomena should be treated in which theoretical layer and how far sentence bound theories should be stretched. DS theoreticians maintain a two-fold perspective, on the one hand they talk about a DS model of dialogue and on the other hand they treat dialogue features within the frame of a proposition. We briefly look at these two things starting off with some of the basic assumptions and tools of DS, leaving, however, the wealth of technical details aside (concerning these, see Cann, Kempson, Marten (2005), Purver, Cann, Kempson, (2006) or Gargett *et al.* (2008) for a short synopsis).

**Basic assumptions and tools of DS (according to Gargett *et al.* (2008)):**

DS is a parsing based approach using a strictly left-to-right interpretation of linguistic input. It works with growing tree structures and their decorations which are encoded using a special modal logic LOFT. LOFT handles the usual relations on finite trees. Under-specification and update are essential ingredients of DS. Growth of information for tree structures and their decorations is conceived of as a monotonic process. The building up of tree structures is steered by the requirements of nodes, for example, an underspecified subject node in a tree would have to be equipped with the information that an expression of the type of individual is needed. Requirements say which information must be got in order to yield a (more) complete structure. Structures are built up by lexical or computational actions. Computational actions comprise the introduction and the updating of structure. Lexical actions introduce individual lexical items inducing nodes and decorations. Partial trees grow incrementally, driven by the requirements of the words encountered. A pointer  $\diamond$  marks the state of the parsing process. Complete individual trees correspond to predicate-argument structures. A tree adjunction operation is used for generating more complex structures via the fitting together of trees sharing one term. 'Importantly, adjunction as other forms of construction and update, can be employed to model how subsequent speakers may dynamically provide fragmentary extensions in response to the previous utterance'. (Gargett *et al.* (2008), p. 45). DS uses structural and content under-specification, e.g. meta-variables for pronouns and subscripted placeholders for names both of which inform tree construction. Meta-variables can be up-dated if the context yields an appropriate term for substitution. Essential for the matters of this paper is the DS notion of context. A context in DS involves the storage of parse states, i.e. of the partial tree arrived at, the word sequence identified and the actions used. As a consequence, the context provides a tool for switching from parsing to generation. The switch from parsing to generation is needed in turn to model coordination of speakers in dialogue.

**Central assumptions of DS concerning dialogue** (called 'the DS dialogue model' in the some of the papers):

Parsing is the prime mechanism of DS, generation being defined upon it. Generation and parsing are both seen as incremental processes. How is the parsing-generation relation modelled?

Whereas a hearer by needs builds up a succession of partial parse trees, a speaker is also equipped with a goal tree containing the information he wants to produce. The generation process is determined by the parsing process on the goal tree and a subsumption relation existing between the goal tree and the partial parse tree guaranteeing felicity of production, so to speak. The DS dialogue model takes account of the hearer's parse tree as well as of the speaker's goal and parse trees. As a consequence, hearer's activities such as clarification requests, completions and acknowledgements can be handled in an on-line manner. Above all, a hearer can start his generation process right from the parsed string resulting from the on-going production of the speaker. This technique is essential for completions and repairs as well as for the production of feed back indicators. 'In particular, for split/joint utterances, this enables switch from hearer to speaker at any arbitrary point in the dialogue [...]' (Gargett *et al.* (2008), p. 46). In sum, we see that we get the incremental tools to handle completions. In the next chapter we show how completions can be modelled in DS, starting with some remarks tying the DS methodology to general dialogue research.

### 7.3.2 Completions in Dynamic Syntax: Purver et al (2005)

Clark and Wilkes-Gibbs (1990) started from the hypothesis that in dialogue participants produce utterances cooperatively. This led to a sort of schema capturing presentation and mutual acceptance as a recursive process, described here informally and in an abridged version. Clark and Wilkes-Gibbs' idea was that e.g. in order to initiate a reference to, say, some Tangram figure, an agent either presents a referring term  $x$  or invites one from the other agent. If a term is invited by the first agent, the other agent, being cooperative, presents one. If the presented term  $x$  is considered inadequate, then a revision  $x'$  of  $x$ , an expansion  $y$  of  $x$  or a replacement  $z$  of  $x$  is provided. Finally, a term  $x$  held adequate is accepted by both parties and mutual acceptance is inferred. Observe that by the last step  $x$  enters the common ground. Applying this device to example 1.2, we have a production *So, jetzt nimmst du* by Inst, a presentation *eine Schraube* by the other agent, Cnst, most probably invited by Inst's lengthening. Inst reacts with an expansion 1.3 *eine orangene mit einem Schlitz* which gets mutually accepted, see Cnst's *Ja* and Inst's continuation. In short, seen from the Clark and Wilkes-Gibbs perspective, syntax production goes through three stages, presentation, repair by expansion and acceptance.

As is clear from the Clark and Wilkes-Gibbs approach, cooperatively produced utterances present a problem since they integrate production and parsing processes, normally treated as separate paradigms. Recent extensions of Dynamic Syntax (DS, Purver and Kempson (2004), Purver, Cann, and Kempson (2006)) have paved the way for modelling cooperative production of utterances by several agents in a theory of grammar framework achieving effects of the original Clark and Wilkes-Gibbs proposal. Purver *et al.* developed a tool which can toggle from parsing to generation states and *vice versa* and can this way also serve as a basis for modelling changes of speaker's activity.

At the heart of this account is the notion of a **parser state**  $P$ .  $P$  is a set of triples  $\langle T, W, A \rangle$ ,  $T$  being a semantic tree under construction,  $W$  a sequence of words, and  $A$  a set of lexical or computational actions used for building up trees or singling out words. The notion of a parser state is then used for the definition of a context  $C$  of a particular tree  $T$  in the set  $P$ .  $C$  consists of

- (a) a set of triples  $P' = \{ \dots, \langle T_i, W_i, A_i \rangle, \dots \}$  resulting from the previous sentence(s) and
- (b) the current triple  $\langle T, W, A \rangle$ .

At the start of the parsing process  $P'$  is empty and the context is identical to the current parser state  $P_0 = \langle T_0, 0, 0 \rangle$ . Here  $T_0$  is the basic axiom  $\langle ?Ty(t), \otimes \rangle$  with  $?Ty(t)$  a designated goal to be proved and  $\otimes$  serving as a pointer to the current stage of the parsing procedure. In  $P_0$  we have empty sequences for words and actions.

Defined in an analogous fashion, a **generator state** is a pair  $(T_g, X)$ .  $T_g$  is a goal tree and  $X$  a set of pairs  $(S, P)$  with  $S$  standing for a candidate string and  $P$  for the associated parser state, a set of  $\langle T, W, A \rangle$  triples. At the beginning of the discourse  $X$  will consist of an empty candidate string and the standard initial parser state  $(0, P_0)$ . The context  $C$  for generation is defined as for parsing, as the set of triples  $P' = \{ \dots, \langle T_i, W_i, A_i \rangle, \dots \}$  and the current triple  $\langle T, W, A \rangle$ .

Purver et al. (2005) provide a model of shared utterances capturing tightly fitting completions and the change of roles among speaker and hearer<sup>29</sup>. The model is given along the following lines. The structural description of the first part  $So^{30}$ , *jetzt nimmst du* which is incomplete by grammaticality standards can be used as an input for the parsing or the generation of the completion *eine Schraube*. Both productions together add up to a well-formed structure as in our example (1), ch. 1. Since one can switch from parsing to generation and *vice versa*, at least a necessary condition for role switches is met by means of the respective contributions.

The normal generation process starts with a generator state  $(T_g, \{0, P_0\})$ , where  $T_g$  is the goal tree, 0 the place holder for the candidate string and  $P_0$  the initial parser state  $\langle T_0, 0, 0 \rangle$ . As mentioned above,  $T_0$  stands for the basic axiom  $\langle ?Ty(t), \otimes \rangle$ . On condition that a suitable goal tree, basically representing the meaning structure of the candidate string, is available for generation, a continuation can be produced using the structure arrived at. This is achieved replacing  $P_0$  with the parser state (a set of triples  $\langle T, W, A \rangle$ ),  $P_t$  got from the structure produced, *called transition state*.

Now, assume speaker role A and the corresponding hearer role B. Hearer B parses up to  $P_t$ . If he has a suitable goal tree  $T_g$ , he can set up a transition generator state  $G_t = (T_g, \{(0, P_t)\})$  and continue. Thus  $G_t$  can be directly used for generation. In order that this may happen, a monotonicity constraint has to be satisfied: the goal tree  $T_g$  must be subsumed by one of the partial trees in  $P_t$ .

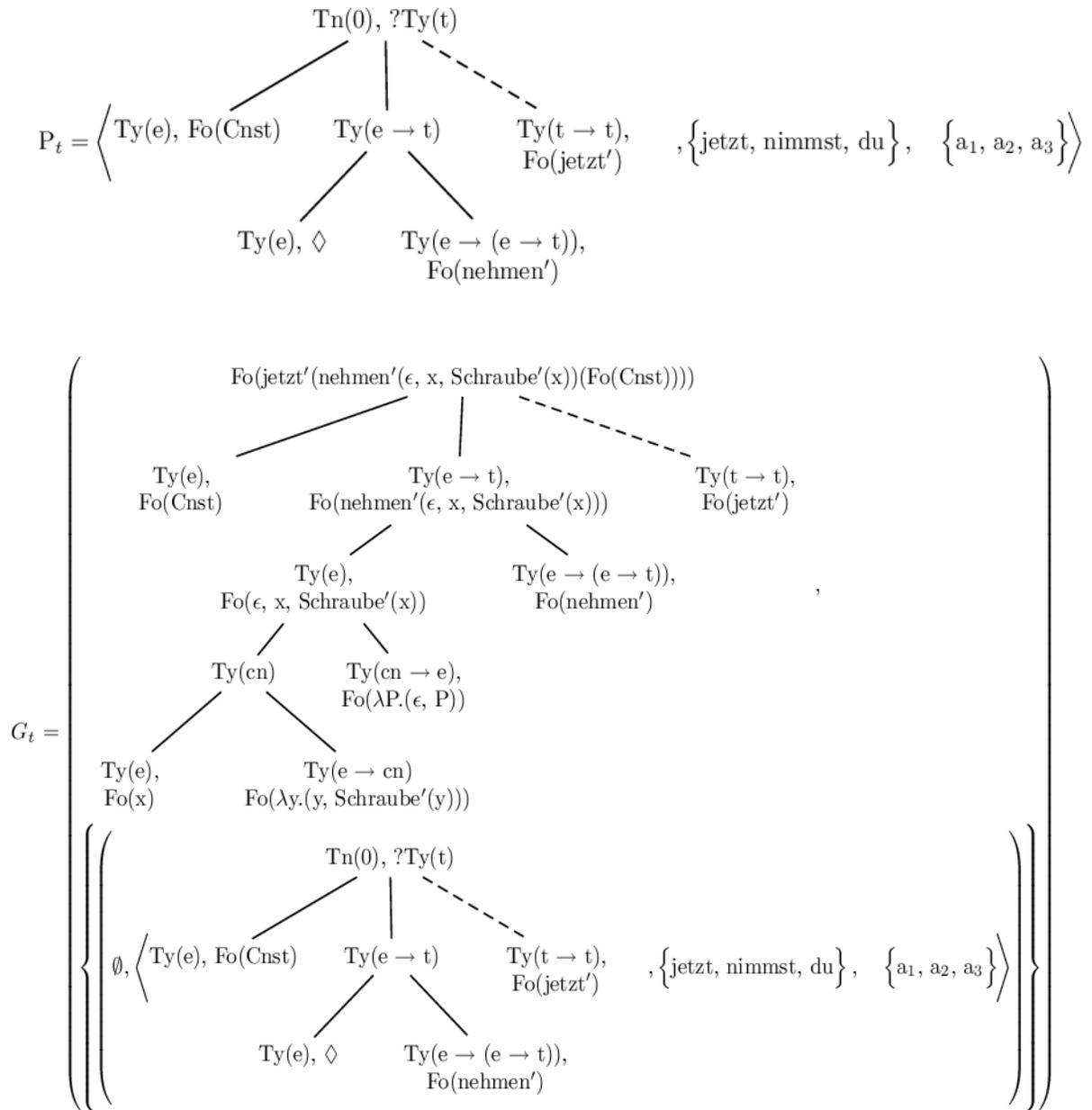
Considering both sides, A and B, we have the following: At the point of transition A's generator state  $G_t'$  contains the pair  $(S_t, P_t')$ .  $S_t$  is the string produced so far and  $P_t'$  the corresponding parser state, the transition state for B. A is assumed to interpret B's continuation in the context of his parser state  $P_t'$ , into which the structures extracted from B's productions will be integrated.

By way of illustration  $Cnst$ 's (= B's) transition from hearer to speaker role is represented in Fig. 7.3.1.  $P_t$  depicts his parsing procedure, the  $\otimes$  indicating that an object  $Ty(e)$  is expected due to the type  $Ty(e \otimes (e \otimes t))$  of *nehmen*. *jetzt* is treated as an unfixed sentence adverb.  $Cnst$  recognized the words [*so,*] *jetzt, nimmst, du* using actions  $a_1, a_2$ , and  $a_3$ . He has a suitable goal tree to set up a transition generator state  $G_t$ . The goal tree shows that the indexical must be changed. Hence we have the logical form  $Fo(jetzt'((nehmen'(\otimes, x, Schraube'(x)))(Fo(Cnst))))$ . The slot for the string is empty, and this is what one expects, given the example: the string produced is *eine Schraube*, it is fitted in as an object-NP, two new actions  $a_4$ , and  $a_5$  are needed to achieve that.

<sup>29</sup> In discussing the completion in the DS framework based on (Purver and Kempson (2004)) and (Purver et al. (2005)) we got support from Matthew Purver and Ruth Kempson, which we gratefully acknowledge. Remaining mistakes are ours, of course.

<sup>30</sup> For *So* we assume a functional interpretation along the lines of the discussion in 5.2; *So* will hence not turn up in the DS trees any more.

Fig. 1 only models the completion part. There is also a suggestion concerning the description of appositions in (Cann, Kempson, Marten (2005), ch. 7, p. 328, ch. 8, 8.4.1 and 8.67, p. 365), which could be used in order to graft the Instructor's *eine orangene mit einem Schlitz* onto *Schraube*. The apposition functions as an other-repair to Cnst's contribution. So far, repairs have not been treated in DS in a systematic way. The most explicit reconstruction of corrections known to us is in SDRT (Asher and Lascarides (2003), 345-373), which presupposes discourse relations but is not tied to incremental surface syntax.



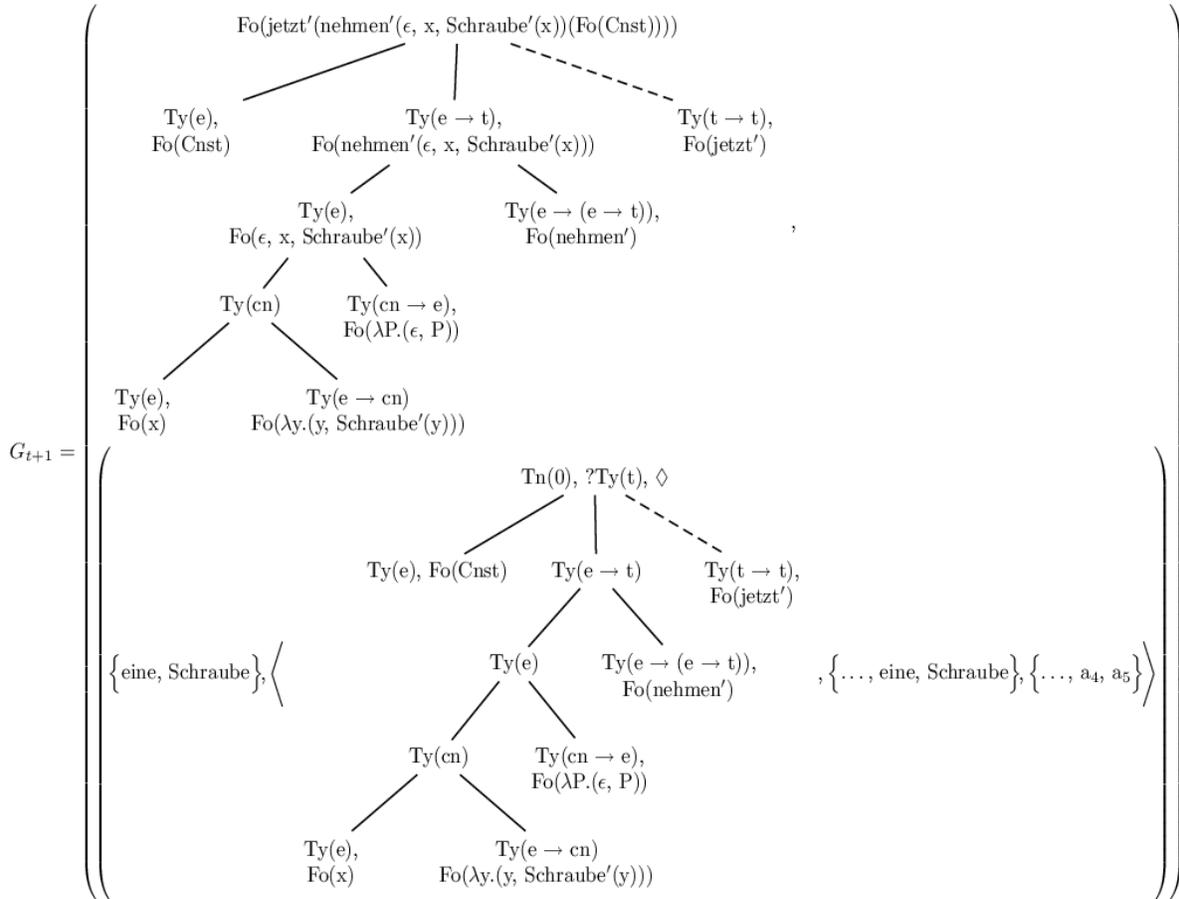


Fig. 7.3.1 Modelling of completion *jetzt nimmst du eine Schraube* involving role shift from hearer to speaker (simplified).

What is the difference between the DS model of completion and our PTT account? The difference can essentially be traced back to the DS-philosophy. By joining parsing and production DS provides a necessary condition for Cnst's production of some object NP but this will not necessarily be *eine Schraube*, it could be anything else, say *a car*, *a black hole* etc. The targeted production of language material is clearly plan-based. So, in order to make DS do that it would have to be tied to some theory of dialogue *via* a suitable interface.

## 8 Discussion

Theorists from Clark to Ginzburg to Pickering and Garrod all agree that perhaps *the* key test for a theory of dialogue is the extent to which it accounts for what is arguably the key difference between dialogue and monologue—the process by which the conversational participants achieve their communication goals by coordinating with each other—without assuming that this communication is always successful. This requirement has perhaps best been captured for information-state theories by Ginzburg in the following passage (Ginzburg 2008, p. 121):

... one could argue that the basic criterion of adequacy of a semantic theory of dialogue is the ability to characterize for any utterance type the update that emerges in the aftermath of successful grounding ... this is the early 21<sup>st</sup> century analogue of truth conditions

To our knowledge, what was presented in this paper is the first formal account of a crucial aspect of this process of coordination in dialogue: the fact that coordination can be achieved at the level of the single utterance. Most ideas have been developed elsewhere: our account builds, on the one hand, on Clark's proposals about coordination, as formalized by Tuomela for shared plans, and by Traum for grounding; on the other, on ideas about the dynamics of the common ground developed by Stalnaker, Lewis, and Kamp; last but not least, on the view of the common ground developed in Situation Semantics by Barwise, Perry, Cooper, and colleagues. As far as we know, however, this is the first attempt to put all these ideas together in a single account.

This being the case, it is almost inevitable that our account is not going to be the ultimate word on the argument. We mentioned through the text a number of issues to be addressed in further research, ranging from empirical questions such as whether alignment of representations à la Pickering&Garrod is epiphenomenal (i.e., resulting from more general coordination of MBs?) or a separate mechanism; to formal questions concerning the consistency of the underlying logic. Some issues will be addressed in a follow-up paper that will extend the analysis to the phenomena observed in the second part of the example dialogue, in particular, providing an incremental account of anaphora.

## References

- Abeillé, Anne & Rambow, Owen (eds) (2000). *Tree Adjoining Grammars*. CSLI Publ., Stanford, CA.
- Aijmer, Karin (1996). *Conversational Routines in English*. London: Longman.
- Allen, James F., Len Schubert, Peter Heeman, Chung-Hee Hwang, Marc Light, Brad Miller, Massimo Poesio and David Traum (1995). The TRAINS Project: A Case Study in Defining a Conversational Planning Agent, *Journal of Experimental and Theoretical AI*.
- Allen, James and Core, Mark (1997). Draft of DAMSL: Dialog Act Markup in Several Layers. October 18.
- Allwood, Jens, Joakim Nivre, and Elisabet Ahlsén (1993). On the Semantics and Pragmatics of Linguistic Feedback. *Journal of Semantics*, v. 9, n. 1,
- Asher, Nicholas (1993). *Reference to Abstract objects in Discourse*, Kluwer Academic Publishers
- Asher, Nicholas and Lascarides, Alex (2003). *Logics of Conversation*. CUP
- Austin, J. L. (1962). *How to Do Things with Words*. Oxford, OUP
- Bara, B. G., & Tirassa, M. (2000). Neuropragmatics: Brain and communication. *Brain and Language*, 71, 10-14
- Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M. Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, 42 (1), 1-22.
- Barwise, Jon & Perry, John (1983). *Situations and attitudes*. Cambridge, Mass.: MIT Press.
- Branigan, H.P., Pickering, M.J. & Cleland, A.A. (2000). Syntactic coordination in dialogue. *Cognition*, 75, B13-B25.
- Branigan, H.P., Pickering, M.J., Stewart, A.J. & McLean, J.F. (2000). Syntactic priming in spoken production: Linguistic and temporal interference. *Memory & Cognition*, 28, 1297-1302
- Bratman, M. (1992). Shared Cooperative Activity, *The Philos. Review* 101, 327-341
- Bratman, M. (1993). Shared Intention. In *Ethics* 104, pp. 97-113
- Brennan, S. E. (2005). How conversation is shaped by visual and spoken evidence. In J. Trueswell & M. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-action traditions* (pp. 95-129). Cambridge, MA: MIT Press.
- Brewka, Gerhard (1989). Nonmonotonic reasoning: from theoretical foundation towards efficient computation, 1989 . - Hamburg, Univ., Diss., 1989
- Brewka, Gerhard (1989). Preferred Subtheories: An Extended Logical Framework for Default Reasoning. IJCAI 1989: 1043-1048
- Brewka, Gerhard (1990). Handling Incomplete Knowledge in Artificial Intelligence. IS/KI : 11-29
- Bunt, Harry (1995). 'Dialogue control functions and interaction design'. In: R.J. Beun, M. Baker & M. Reiner (eds.) *Dialogue in Instruction*. Springer Verlag, Heidelberg, 1995, pp. 197 -- 214.
- Cann, R., Kempson, R., Marten, L. (2005), *The Dynamics of Language*. Amsterdam: Elsevier
- Carletta, J., et al. (1997). The Reliability of a Dialogue Structure Coding Scheme. *Computational Linguistics*, 23(1), 13-32.
- Carpenter, Bob (1994). Quantification and Scoping: A deductive account. In *Proceedings of the 13th West Coast Conference on Formal Linguistics*. San Diego. CSLI Press.
- Chater, N.J., Pickering, M.J., & Milward, D. (1995). What is incremental interpretation? In D. Milward & P. Sturt (eds.), *Edinburgh working papers in cognitive science 11: Incremental interpretation* (pp. 1-22). Centre for Cognitive Science, University of Edinburgh
- Clark, Herbert H. (1992). *Arenas of Language Use*. The Univ. of Chicago Press and CSLI .
- Clark, Herbert H. (1996). *Using Language*. Cambridge: Cambridge University Press.

- Clark, Herbert H. and Deanna Wilkes-Gibbs (1990). Referring as a collaborative process. In Cohen, Ph. R., J. Morgan, and M.E. Pollack, eds., *Intentions in Communication*, MIT Press, 1990: 461-493.
- Clark, Herbert H., and Brennan, S. A. (1991). Grounding in Communication. In L. B. Resnik, J. M. Levine, and S. D. Teasley (Eds.), *Perspectives on socially shared cognition*, Washington, DC: APA Books, pp. 127-149
- Clark, Herbert H., and Schaefer, E. F. (1987). Collaborating on contributions to conversations. *Language and Cognitive Processes*, 2 (1), 19-41
- Clark, Herbert H., and Schaefer, E. F. (1989). Contributing to Discourse. *Cognitive Science*, 13, 259-294
- Clark, Herbert, H. and Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22, 1-39
- Cohen, Ph. R. and Levesque, H. J. (1990a). Persistence, Intention, and Commitment. In Cohen, Ph. et al. (eds.), *Intentions in Communication*, pp. 34-69
- Cohen, Ph. R. and Levesque, H. J. (1990b). Rational Interaction as a Basis for Communication. In Cohen, Ph. et al. (eds.), *Intentions in Communication*, pp. 221-225
- Cohen, P.R., Morgan, J., Pollack, M.E. (Eds.), *Intentions in Communication*, MIT Press, Cambridge, MA, 1990.
- Cole, Peter, ed. (1978). *Syntax and Semantics, 9: Pragmatics*. New York: Academic Press.
- Cooper, Robin (1992). A Working Person's Guide to Situation Theory 1992, in *Topics in Semantic Interpretation*, ed. by Steffen Leo Hansen and Finn Soerensen, Samfundslitteratur, Frederiksberg, Denmark
- Cooper, Robin and Poesio, Massimo (1994). Situation Theory. In The FRACAS Consortium (eds.), *FRACAS Deliverable D8*, Centre for Cognitive Science, Edinburgh.
- Cooper, Robin (1996). The role of situations in generalized quantifiers. In Lappin, S. (ed), *Handbook of Contemporary Semantic Theory*, Blackwell.
- Cooper, Robin, Stina Ericsson, Staffan Larsson, Ian Lewin (2003): An information state update approach to collaborative negotiation. In Kuhnlein, Rieser and Zeevat (eds.): *Perspectives on Dialogue in the new Millennium*, John Benjamins Publishing Company. (pp. 271-287, 16 pages)
- Cooper, Robin (2005). *Type Theory with Records*. University of Gothenburg, Ms.
- Coulthard, Malcolm (1977). *An introduction to Discourse Analysis*. Longman.
- Dale, R. and Reiter, E (1995) Computational Interpretations of the Gricean Maxims in the Generation of Referring Expressions. *Cognitive Science*, 19, April-June 1995.
- deMena Travis, Lisa (1992). Parameters of Phrase Structure and Verb-Second Phenomena. In Freidin, Robert (ed.) (XXXX). *Principles and Parameters in Comparative Grammar*. Cambridge, Mass: The MIT Press, pp. 339-365
- Fagin, R., Halpern J.Y., Moses, Y., Vardi, M.Y. (1995). Reasoning about Knowledge. The MIT Press: Cambr., Mass.
- Farkas, D. and de Swart, H. (2003). *The Semantics of Incorporation*. CSLI.
- Ferguson, George, James Allen, and Brad Miller (1996). "TRAINS-95: Towards a Mixed-Initiative Planning Assistant," in *Proceedings of the Third Conference on Artificial Intelligence Planning Systems (AIPS-96)*, Edinburgh, Scotland, 29-31 May, pp. 70-77.
- Fernandez, Raquel (2006). *Non sentential utterances in dialogue: classification, resolution and use*. PhD thesis, King's College, London.
- Ferreira, F., Ferraro, V., & Bailey, K.G.D. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11, 11-15
- Ferreira, F., Lau, E.F., & Bailey, K.G.D. (2004). Disfluencies, parsing, and tree-adjointing grammars. *Cognitive Science*, 28, 721-749.

- Fletcher, C. (1994). Levels of representation in memory for discourse. In Gernsbacher, M. A. (eds), *Handbook of Psycholinguistics*, 1<sup>st</sup> edition, Academic Press.
- Frank, Robert (2002). *Phrase Structure Composition and Syntactic Dependencies*. Cambridge, Mass: The MIT Press
- Frazier, Lyn (1987). *Sentence Processing: A Tutorial Review*. In: Max Coltheart (ed.), *The Psychology of Reading*. Hove: Erlbaum. 559-586.
- Gargett, A., Gregoromichelaki, E., Howes, Ch., Sato, Y. (2008). Dialogue-Grammar Correspondence in Dynamic Syntax. In Ginzburg, J., Healey, P., Sato, Y. (eds.), *Proceedings of the 12<sup>th</sup> Workshop on the Semantics and Pragmatics of Dialogue (LONDIAL '08)*. 2-4 June, London 43-51
- Garnham, A. (2001). *Mental Models and the Interpretation of Anaphora*, Psychology Press, Hove, UK
- Garrod, Simon and Anderson, Anne (1987). Saying what you mean in dialogue: A study in conceptual and semantic coordination. *Cognition* 27, pp. 181-218
- Geurts, Bart (1995). *Presupposing*. PhD dissertation, University of Stuttgart.
- Ginzburg, Jonathan (1997). "On some Semantic Consequences of Turn Taking". In: P. Dekker et al (eds.) *Proceedings of the 11th Amsterdam Colloquium*.
- Ginzburg, J. (To appear). Situation Semantics: from indexicality to metacommunicative interaction. To appear in .... CHECK REFS
- Ginzburg, Jonathan and Cooper, Robin (2004). "Clarification, Ellipsis, and the Nature of Contextual Updates". *Linguistics and Philosophy*, 27:3, pp. 297-366
- Ginzburg, Jonathan and Fernandez, Raquel (2005). 'Scaling up to multilogue: some benchmarks and principles', *Proceedings of the 43rd meeting of the Association for Computational Linguistics, Ann Arbor, Michigan*.
- Ginzburg, Jonathan, Sag, Ivan. (2000). *Interrogative Investigations*. CSLI Publications
- Ginzburg, Jonathan. (2008). *Semantics for Conversation*. CSLI Publications
- Goldman, Alvin J. (1970). *A Theory of Human Action*. Prentice-Hall.
- Grice, H. P. (1969). "Utterer's Meaning and Intention." *Philosophical Review*, 68, 147-77.
- Grice, H. P. (1975). "Logic and Conversation." In Cole and Morgan (1975), pp. 41-58.
- Grice, H. P. (1978). "Further Notes on Logic and Conversation." In Cole (1978), pp. 113-27.
- Grice, H. P. (1991). Utterer's Meaning, Sentence-Meaning, and Word-Meaning. In: Davis St. (ed.) *Pragmatics*. New York, Oxford: OUP, pp. 65-76
- Groenendijk, J. and M. Stokhof (1991). Dynamic Predicate Logic. *Linguistics and Philosophy* 14: 39-100.
- Gross, Derek, James Allen and David Traum (1993). The TRAINS 91 Dialogues. Technical Report, University of Rochester, Department of Computer Science, TRAINS Technical Note 92-1.
- Grosz, Barbara J. and Sidner, Candace L. (1986). Attention, intention, and the structure of discourse. *Computational Linguistics* 12: 175-204
- Grosz, Barbara J. and Sidner, Candace L. (1990). Plans for Discourse. Cohen *et al.* eds., 417-445
- Grosz, Barbara and Kraus, S. (1996). Collaborative plans for complex group action. *Artificial Intelligenc.* 86, pp. 269-357
- Heeman, Peter and James Allen (1995). The TRAINS 93 Dialogues. Technical Report, University of Rochester, Department of Computer Science, TRAINS Technical Note 94-2.
- Hobbs, Jerry R., Douglas E. Appelt, John Bear, Mabry Tyson, and David Magerman (1992) "Robust Processing of Real-World Natural-Language Texts", in *Text-Based Intelligent Systems: Current Research and Practice in Information Extraction and Retrieval*, P. Jacobs, editor, Lawrence Erlbaum Associates, Hillsdale, New Jersey, pp. 13-33.
- Homer, Bruce D. and Catherine S. Tamis Le Monda, editors (2005). *The development of social cognition and communication*. Psychology Press.

- Hwang, C.H. and Schubert, L. K. (1993)"Meeting the interlocking needs of LF-computation, deindexing, and inference: An organic approach to general NLU." In *Proc. 13th Int. Joint Conf. on Artificial Intelligence*, August, 1993.
- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.
- Joshi, A. K. (2004). 2003 Rumelhart Prize Special Cognitive Science Issue Honoring Aravind K. Joshi, Vol, 28, Number 5, September/October 2004
- Jurafsky, Daniel (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science* 20.137–194.
- Kamp, Hans and Uwe Reyle (1993). *From Discourse to Logic*. Dordrecht: Kluwer.
- Kautz, H. A. (1987). *A Formal Theory of Plan Recognition*. PhD thesis, University of Rochester, Rochester, N.Y.
- Kempson, R., Meyer-Viol, W. and Gabbay, D. (2001) *Dynamic Syntax*. Oxford: Blackwell
- Kranstedt, A., Lücking, A., Pfeiffer, Th., Rieser, H., Wachsmuth, I. (2006). Deictic object reference in task-oriented dialogue. In: Rickheit, G. and Wachsmuth, I. (eds.), *Situated Communication*. Mouton de Gruyter: Berlin, New York, pp. 156-209
- Kroch, Anthony, S. and Santorini, Beatrice (1992). The Derived Constituent Structure of the West Germanic Verb-Raising Construction. In Freidin, Robert (ed.), pp. 269-339
- Larsson, Staffan and Traum, David (2000). Information state and dialogue management in the TRINDI Dialogue Move Engine Toolkit. *Natural Language Engineering*, Special Issue on Best Practice in Spoken Language Dialogue Systems Engineering. v.6, pp. 323-340.
- Lerner, G. H. (1987). *Collaborative turn sequences : Sentence construction and social interaction*. Ph. D. dissertation, University of California, Irvine
- Levelt, J.M. (1989) *Speaking: From Intention to Articulation*. Cambridge Massachusetts: MIT Press.
- Levin, James A. and Moore, James A. (1977). Dialogue Games: Meta-communication Structures for Natural Language Interaction. *ISI/RR-77-53*, Information Sciences Institute, Univ. of Southern California
- Levinson, Stephen, C. (1983). *Pragmatics*. CUP
- Litman, Diane, J. and Allen, James, F., (1987). A plan recognition model for subdialogues in conversation. *Cognitive Science* 11: 163-200
- Litman, Diane, J. and Allen, James, F., (1990). Discourse Processing and Commonsense Plans. Cohen *et al.* eds., 365-389
- Mann, William C. (1988). Dialogue Games: Conventions of Human Interaction. *Argumentation* 2 : 511-532
- Matheson, Colin, Massimo Poesio, and Traum, David (2000). "Modeling Grounding and discourse obligations using update rules", *Proc. Of the NAACL*.
- McNeill, D. 1992. *Hand and Mind: what gestures reveal about thought*. Chicago: University of Chicago Press
- Morton J. Interaction of information in word recognition. *Psychological Review*. 1969;76:165–178
- Muskens, Reinhard (1995). Tense and the Logic of Change. In Urs Egli, E.P. Pause, Christoph Schwarze, Arnim von Stechow, and G. Wienold, editors, *Lexical Knowledge in the Organization of Language*, pages 147-183. John Benjamins, Amsterdam.
- Muskens, Reinhard (1996). Combining Montague Semantics and Discourse Representation. *Linguistics and Philosophy* 19, pp. 143-186
- Muskens, Reinhard (2001). Talking about Trees and Truth-conditions. *Journal of Logic, Language and Information*, 10(4):417-455, 2001.
- Neumann, G.: (1998), Interleaving natural language parsing and generation through uniform processing. In: *Artificial Intelligence* 99 (1998), pp. 121-163.

- Parsons, Terence (1990). *Events in the Semantics of English: A Study in Subatomic Semantics*. Current Studies in Linguistics Series, Vol. 19. Cambridge: MIT Press
- Pereira, Fernando C. N. (1990). Categorical semantics and scoping. *Computational Linguistics* 16(1), pp. 1-10.
- Perrault, Ray (1990). An application of default logic to speech-act theory. In Cohen, Morgan and Pollack (eds), *Intentions in Communication*, pages 161–187, 1990.
- Perrault, C. Raymond & Allen, James F (1980). A Plan-Based Analysis of Indirect Speech Acts. *American Journal of Computational Linguistics* 6 (3-4): 167-182 (1980)
- Pickering, Martin, J. & Garrod Simon (2004). Toward a Mechanistic Psychology of Dialogue. *Behavioral and Brain Sciences* 27, pp. 169-226
- Platzack, Christer (1986). COMP, INFL, and Germanic Word Order. In Hellan, Lars and Christi Koch Christensen (eds.) *Topics in Scandinavian Syntax*. Dordrecht: D. Reidel
- Poesio, Massimo (1993). "Definite Descriptions and the Dynamics of Mental States," AAAI Spring Symposium on Reasoning About Mental States, March, 1993.
- Poesio, Massimo (1994). *Discourse Interpretation and the Scope of Operators*. PhD dissertation, University of Rochester, May 1994
- Poesio, Massimo (1995). "A Model of Conversation Processing Based on Micro Conversational Events," In *Proceedings of the 17th Annual Meeting of the Cognitive Science Society*, July 1995.
- Poesio, Massimo (1996). "Underspecification and the Semantics / Pragmatics Interface," In *Proceedings of the Workshop on the Cognitive Science of Natural Language Processing*, September 96.
- Poesio, Massimo (2001). "What psycholinguistics tells us about the semantics / pragmatics interface: the case of pronouns ", *Amsterdam Colloquium*, Amsterdam, December 2001.
- Poesio, Massimo and Muskens, Reinhard (1997). "The Dynamics of Discourse Situations", *Proc. Of the 11<sup>th</sup> Amsterdam Colloquium*, 1997, 247-252.
- Poesio, Massimo and Traum, David (1997). "Conversational Actions and Discourse Situations", *Computational Intelligence*, v. 13, n.3, 1997, pp.1- 45
- Poesio, Massimo and Traum, David (1998). "Towards an Axiomatization of Dialogue Acts", *Proc. Of TWENDIAL*, 1998, pp. 1-7
- Poesio, Massimo and Traum, David, R., eds., (2000). *Proceedings of GÖTALOG 2000*. Fourth Workshop on the Semantics and Pragmatics of Dialogue, Göteborg Univ.
- Pollack, M. (1986), *Inferring Domain Plans in Question-Answering*. University of Pennsylvania Doctoral Thesis, Philadelphia.
- Poncin, K. and Rieser, H. (2006). Multi-speaker utterances and co-ordination in task-oriented dialogue. *Journal of Pragmatics* 38 , pp. 718-744. Revised version of Rieser and Skuplik, 2000.
- Portner, Paul (2007). Imperatives and Modals. *Natural Language Semantics*.
- Purver, M., Cann, R. and Kempson, R. (2006), Grammars as Parsers: Meeting the Dialogue Challenge. *Research on Language and Computation*, Vol. 4, Numbers 2-3, pp. 289-326
- Purver, M. and Kempson, Ruth (2004). Incremental Parsing, or Incremental Grammar? In *Proceedings of the ACL Workshop on Incremental Parsing*, pages 74-81, Barcelona, July 2004.
- Purver, Matthew and Ginzburg. Jonathan (2004). Clarifying Noun Phrase Semantics. *Journal of Semantics* 21(3), pages 283-339, 2004.
- Purver, Matthew (2004). *The theory and use of clarification in dialogue*. PhD thesis, King's College, London.
- Reiter, R. (1980). A logic for default reasoning. *Artificial Intelligence*, 13:81-132.
- Rickheit, G. & Wachsmuth, I. (2008). Alignment in communication - Collaborative Research Center 673 at Bielefeld University. *Künstliche Intelligenz*, Heft 2-2008, 62-65

- Rieser, Hannes (2004). Pointing in Dialogue. In Ginzburg, J. and Vallduvi, E. (eds.) *Catalog '04 . Proceedings of the Eighth Workshop on the Semantics and Pragmatics of Dialogue*. Department of Translation and Philology, Universitat Pompeu Fabra, Barcelona.
- Rieser, Hannes. (2008). Aligned Iconic Gesture in Different Strata of MM-Route-description Dialogue. In: *Proceedings of LONdial 2008*, pp. 167-174
- Rieser, H., Kopp, St. and Wachsmuth, I. (2007). Speech-Gesture Alignment. In: *ISGS Abstracts, Integrating Gestures*. Northwestern University, Evanston, Chicago, pp. 25-27
- Rieser, Hannes and Kristina Skuplik, (2000). Multi-speaker Utterances and Coordination in Task-oriented Dialogue. In Poesio, M. and Traum, D. (eds.), pp. 143-151
- Sacks, Harvey (1967). 'The Search For Help: No One To Turn To', *Essays in Self-Destruction*, E. S. Shneidman (ed.). New York: Science House
- Sacks, Harvey, Schegloff, Emanuel and Jefferson, Gail (1974): "A Simplest Systematics for the Organization of Turn-Taking for Conversation," *Language*, 50, 4 (1974) 696-735
- Sadek, M.D. (1992). A Study in the Logic of Intention. In: B. Nebel, Ch. Rich, W. Swartout (eds.): *Principles of Knowledge Representation and Reasoning*. Proceedings of the Third International Conference. Morgan & Kaufmann Publishers: San Mateo, CA, pp. 462-475
- Sadek, D. (1994). Communication theory = rationality principles + communicative act models. In *Proc. Of the AAAI 94 Workshop on Planning for Interagent Communication*, 1994.
- Sanford, A.J. and Garrod, S. (1981) *Understanding Written Language*. Chichester, John Wiley.
- Schabes, Y., Abeillé, A., & Joshi, A.K. (1988). New parsing strategies for tree adjoining grammars. In *Proceedings of the 12th International Conference in Computational Linguistics*, pp. 578– 583. Budapest, August.
- Schegloff, E.A., Jefferson, G., and Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, 53(2), 361-382.
- Schlangen, D. and A. Lascarides (2003). The Interpretation of Non-Sentential Utterances in Dialogue, in *Proceedings of the 4th SIGdial Workshop on Discourse and Dialogue*, Sapporo, Japan.
- Searle, John. R. (1969). *Speech Acts*. CUP
- Searle, John. R. and Daniel Vanderveken (1985). *Foundations of Illocutionary Logic*. Cambridge: Cambridge University Press.
- Schabes Y. (1990). *Computational and Mathematical Properties of Lexicalized Grammars*, PhD Dissertation, Univ. of Pennsylvania, Philadelphia.
- Skuplik, Kristina (1999). Satzkooperationen. Definition und empirische Untersuchung. Bielefeld Univ., Report 1999/03 of SFB 360
- Stalnaker, Robert (1978) "Assertion." *Syntax and Semantics* 9 (1978): 315-332. (Reprinted in Steven Davis, *Pragmatics: A reader*, New York and Oxford: Oxford University Press, 1991.)
- Stone, Matthew (2004) Intention, interpretation and the computational structure of language. *Cognitive Science* 28(5):781-809.
- Sturt, P. and Crocker, M. (1996). Monotonic syntactic processing: A cross-linguistic study of attachment and reanalysis. *Language and Cognitive Processes* 11, pp. 449–494.
- Swinney, D (1979). Lexical access during sentence comprehension: (Re) consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, 18, 645-659
- Tanenhaus M.S., Leiman, J.L., and Seidenberg, M.S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior*, 18, 427-440.
- Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M. & Sedivy, J.E. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632-1634.
- Traum, David (1994). *A Computational Theory of Grounding in Natural Language Conversation*, TR 545 and Ph.D. Thesis, Computer Science Dept., U. Rochester, December 1994.

- Traum, David R. (1999). 20 Questions on Dialogue Act Taxonomies. [Amstelogue'99].
- Traum, David R., Elizabeth A. Hinkelman (1992): Conversation Acts in Task-Oriented Spoken Dialogue. *Computational Intelligence* 8: 575-599 (1992)
- David R. Traum Computational Models of Grounding in Collaborative Systems, in working notes of AAAI Fall Symposium on Psychological Models of Communication, p. 124-131, November, 1999.
- Traum, David R. and Nakatani, Christine H. (1999). A Two-level Approach to Coding Dialogue for Discourse Structure: Activities of the 1998 Working Group on Higher-level Structures, in proceedings of the ACL'99 Workshop *Towards Standards and Tools for Discourse Tagging*, pp 101--108, June.
- Traum, David and Larsson, Staffan (2003): The Information State Approach to Dialogue Management. In Smith and Kuppevelt (eds.): *Current and New Directions in Discourse & Dialogue*, Kluwer Academic Publishers. pp. 325-353
- Tuomela, R. (2000). *Cooperation*. Kluwer Academic Publishers
- Walker, M. (1993). *Informational redundancy and resource bounds in dialogue*. PhD dissertation, University of Pennsylvania.
- Wilkes-Gibbs, D. (1995). Coherence in collaboration: Some examples from conversation. In: Gernsbacher, M. A. and Anderson, A. (eds.): *Coherence in spontaneous text*. Amsterdam: Benjamins, 239-267
- Wolpert DM, Doya K and Kazato (2003). *M. Phil Trans R Soc Lond B Biol Sci* 358, 593–602

## APPENDIX A: FULL TRANSCRIPT OF THE DIALOGUE PRIOR TO THE EXCERPT

### Versuchspersonen-Paar 21

"so, jetzt nimmst du - eine Schraube"

Instrukteur (I) m; Konstrukteur (K) f

Bedingung: Sicht blockiert, Vorlage Modell, Beschreibung/Konstruktion simultan

Gesamtdauer 12: 45

Anmerkung: I erwähnt 'Flugzeug' zunächst nicht; K kommt von selbst drauf.

Konstruktionsresultat I:

Aa=10.3Ab=08.1Ba=10.1Bb=08.2C=02.1Da=12.2Db=01.1Dc=08.3Ea=12.1  
Eb=01.2Ec=07 Ed=05 Ee=13 Ha=11.3Hb=02.2Hc=04 Hd=03.1He=11.1  
La=14.1Lb=15.1Lc=09.1Ld=16.1Le=17.1Ra=14.2Rb=15.2Rc=09.2Rd=16.2  
Re=17.2Va=11.2Vb=03.2Vc=06Vd=03.3Ve=03.4Vf=10.2Muttern=VL

Konstruktionsresultat K:

Aa=10.3Ab=08.1Ba=10.1Bb=08.2C=02.1Da=12.2Db=01.1Dc=08.3Ea=12.1  
Eb=01.2Ec=07 Ed=05 Ee=13 Ha=11.3Hb=02.2Hc=04 Hd=03.1He=11.1  
La=14.1Lb=15.1Lc=09.2Ld=16.1Le=17.1Ra=14.2Rb=15.2Rc=09.1Rd=16.2  
Re=17.2Va=11.2Vb=03.2Vc=06Vd=03.3Ve=03.4Vf=10.2Muttern=VR  
Montiert Hb parallel zu C bugwärts zeigend.

21I001 mit den sieben Löchern.  
21K001 ja. <sil: 2>  
21I002 da hast du zwei Stück von <noise: klappern>  
21K002 ja.  
21I003 <noise> die kannst du erstmal zur Seite legen. dann ist eins mit eins,  
zwei, drei, vier, fünf Löchern </noise: klappern>  
21K003 {ja}<noise: klappern>  
21I004 <noise> <hum: atmen> und <sil: 1> eins mit drei Löchern </noise:  
klappern>  
21K004 {ja}<noise: klappern>  
21I005 <noise> so, <--> jetzt nimm das äh mit den fünf Löchern mal in die  
linke Hand </noise: klappern>  
21K005 mhm.  
21I006 das andere in die rechte <noise: klappern> <-> und leg(e) sie so  
übereinander, daß sich zwei Löcher überschneiden.  
21K006 also die zwei mit den fünf Löchern? <sil: 1>  
21I007 ja.  
21K007 das sie sich wo  
21I008 das kleine ist dann untendrunter so  
21K008 hä? also das mit den drei Löchern?  
21I009 ja.  
21K009 untendrunter.  
21I010 ja. <attrib> oh, das ist aber schwierig </attrib: leise> <-> jetzt  
überschneiden sich ja zwei Löcher, ne? <sil: 2>  
21K010 zwei?  
21I011 bei dir auch?  
21K011 wie wann? <-> <hum: lachen> ich soll das in der Mitte zusammenlegen  
<par> überkreuzen </par: 1>  
21I012 <par> nein, nicht in der Mitte so, </par: 1> daß es also wieder quasi  
eine <--> eine <par> (ei)n Teil so </par: 2>  
21K012 <par> ah ja, Verlängerung </par: 2> mit zwei Löcher.  
21I013 mhm.  
21K013 <quest: ?> überschneiden.

21I014 also wenn du jetzt von oben draufguckst, hast du sechs Löcher.  
21K014 ja.  
21I015 ja?  
21K015 mhm.  
21I016 zwei zwei überschneiden sich, ist das richtig?  
21K016 ja.  
21I017 so und jetzt nimmst du die beiden <--> m die mit den äh <-> sieben  
{Löchern}<spk: K, mhm> ja?  
21K017 ja.  
21I018 die die legst du so drauf <-> auf diese beiden, die sich überschneiden,  
daß sich das (ei)n Kreu/, daß es sich ein Kreuz ergibt so. <-> weißt  
du? <-> legst du auf den Tisch am besten.  
21K018 <noise> das mit den drei Löchern immer noch unten und dann leg(e) ich  
was? <hum: atmen> </noise: klappern>  
21I019 ja.  
21K019 <hum: stöhnen> (ei)n Kreuz.  
21I020 ja, daß es so <sil: 2> jetzt überschneiden sich quasi so die ähm <hum:  
>?> eben haben sich ja nur die <--> beiden überschritten, ja?  
21K020 ja.  
21I021 und jetzt überschneiden die sich dreimal. <-> verstehst du? <hum:  
lachen>  
21K021 nee.  
21I022 m, dann machen wir das mal anders, warte mal <noise: klappern> <hum:  
lachen> <--> m, leg(e) das mal so auf den Tisch mit den, daß die sich  
die bei, wie wir es eben hatten, daß wir sechs Löcher so haben.  
21K022 ja.  
21I023 legst du so auf den Tisch.  
21K023 hab(e) ich schon.  
21I024 und da ist noch ein anderes mit dr/, nimmst du noch ein anderes mit  
drei Löchern,  
21K024 {mhm}<noise: klappern>  
21I025 und das schiebst du vor <-> von <noise> links <--> drunter, </noise:  
klappern> <par> daß sich auf der anderen Seite </par: 3> <cont>  
21K025 <par> auch nur das zwei überschn/ </par: 3>  
21I026 </cont> zwei überschneiden.  
21K026 ah ja.  
21I027 ja?  
21K027 dann habe ich sieben Löcher.  
21I028 ja, ja, so können wir es liegen lassen.  
21K028 mhm.  
21I029 jetzt nimmst du so einen roten Würfel.  
21K029 <attrib> was war mit dem {sieben}<noise: klappern> Löchern, den wir  
</attrib: leise> ja gut.  
21I030 ja?  
21K030 ja.  
21I031 und den legst du ganz links <--> drauf.  
21K031 ganz links?  
21I032 ja, auf das äußerste Loch.  
21K032 mhm.  
21I033 so <-> jetzt nimmst du <-> eine <-> rote Schraube <sil: 2>  
21K033 mhm.  
21I034 und schraubst es von unten fest. kannst du so, mußt du so hochnehmen  
dafür.  
21K034 (ei)ne kleine rote?  
21I035 ja.  
21K035 eckige?  
21I036 ja, ja.  
21K036 gut. <noise> <-> dann ist jetzt nur der Dreier fest <--> erstmal  
</noise: klappern>  
21I037 ja, genau, mhm. <-> kannst es aber wieder so hinlegen. ja, paß auf und  
dann nimmst du jetzt erstmal noch so(\_ei)n Teil mit fünf Löchern.  
21K037 ja.

21I038 und <-> schraubst es oben auf den Würfel drauf mit einer runden Schraube mit (ei)nem Schlitz drin.

21K038 <noise> mit der roten runden? </noise: klappern>

21I039 ja. <sil: 7> hast du?

21K039 ja.

21I040 so, jetzt nimmst du <-> (ei)ne gelbe runde

21K040 mhm.

21I041 und so(\_ei)n <sil: 2> ähm orangenes Teil, so(\_ei)n Gewinde, mit Gewinde.

21K041 ja.

21I042 ja? und schraubst das äh <-> Dreier, die Dreierstange an die Fünfer.

21K042 also so wie gehabt ne, diese untere? <sil: 2> oder was?

21I043 ja, ja, also die Dreier wird von unten an die Fünfer geschraubt.

21K043 mhm.

21I044 verstehst du?

21K044 ja. <sil: 2> und wie rum ich das mache ist egal, von unten die gelbe rein?

21I045 nee, von oben die gelbe.

21K045 und wo <-> wo von den beiden? <sil: 3> ich habe ja zwei Löcher, wo ich es reinstecken kann.

21I046 also die gelbe Schraube ist jetzt <--> <hum: räuspern> rechts von dem roten Würfel.

21K046 direkt <par> dran. </par: 4>

21I047 <par> oben. </par: 4> <sil: 1> ja, ja.

21K047 <noise> ja. <sil: 3> </noise: klappern>

21I048 <noise> und das ist quasi in dem ersten Loch von dem Fünfer, ne </noise: klappern>

21K048 mhm.

21I049 in dem ganz links äußersten und <par> mittig </par: 5>

21K049 <par> und dem mittleren vom Dreier </par: 5>

21I050 mittig von den dreien. genau. und jetzt nimmst du die <-> eine eckige gelbe Schraube und machst sie in das rechte <-> Loch daneben.

21K050 ah ja.

21I051 und unten wieder so(\_ei)ne Mutter vor. <sil: 2>

21K051 mhm.

21I052 na. <sil: 4> <hum: atmen> <noise: klappern> jetzt wird es schwieriger. <hum: lachen> <noise: klappern> <-> paß auf, jetzt nimmst du <--> jetzt legst du den anderen, dies Dreierteil, ne dies andere

21K052 ja.

21I053 von rechts, also das ist genau symmetrisch und mit dem <-> mit dem linken Dreier, was du schon festgeschraubt hast.

21K053 ja.

21I054 so <--> jetzt nimmst du eine lange Stange mit sieben Löchern drin,

21K054 ja.

21I055 die legst du jetzt auf <--> die Fünfer.

21K055 auf die {obere}<spk: I, ja.> auf die untere? auf die obere?

21I056 quasi <sil: 2> {ja}<attrib: leise> <noise> <sil: 4> wart(e) ähm m </noise: klappern>

21K056 was war denn jetzt mit der unteren Dreier, das habe ich noch nicht.

21I057 die kommt dadrunter, also legst du einfach, hältst du mal fest so <noise: klappern>

21K057 ja.

21I058 wenn die jetzt festgeschraubt wär(e), wär(e) die genauso wie die andere.

21K058 ja.

21I059 ja?

21K059 <noise: klappern> das ist klar.

21I060 jetzt nimmst du in die linke Ha/, also das hast du jetzt in der rechten Hand, halt das mal so fest, <noise: rascheln> oder legst es auf den Tisch oder irgendwie so.

21K060 ja <noise: klappern>

21I061 ist <quest: (ei)ne> jetzt ähm, legst du <--> das Teil mit den <--> sieben Löchern

21K061 ja.  
21I062 so <sil: 2> <attrib> mhm so dadrauf, warte mal, wie beschreib(e) ich dir das jetzt, mhm </attrib: leise>  
21K062 der obere oder der untere Fünfer, wo soll ich es hintun jetzt?  
21I063 oben auf den <-> auf den Fünfer.  
21K063 ja.  
21I064 also nicht der auf den roten Würfel ist. <sil: 4> da ist ja auch einer mit fünf Löchern auf dem roten Würfel.  
21K064 ja, also auf dem {unteren}<spk: I, ja.> Fünfer.  
21I065 wo links die beiden gelben Schrauben drin sind.  
21K065 ja.  
21I066 dann kommt eine, erst ein Loch, das läßt du frei  
21K066 mhm.  
21I067 und dann kommt <noise> (ei)n Loch, <-> wo du es draufpackst </noise: klappern>  
21K067 ach so, <par> also zwei </par: 6> <cont>  
21I068 <par> da legst du es drauf </par: 6>  
21K068 </cont> überschneiden sich wieder.  
21I069 genau, die überschneiden sich, genau wie (ei)n Kreuz legst du es drauf und zwar mittig, ne <hum: schmatzen>  
21K069 wie (ei)n Kreuz nicht.  
21I070 doch <-> ja so, was heißt (ei)n Kreuz, eine Kreuzung, wie eine Kreuzung.  
21K070 hä, welches überschneidet sich denn dann?  
21I071 unten dieses ähm mit den fünf Löchern, wo die zwei gelben Schrauben drin sind.  
21K071 das zweite von rechts bei dem? <sil: 3>  
21I072 ja, genau, zweite von rechts.  
21K072 und dann leg(e) ich das mittlere von dem Siebener drüber.  
21I073 genau  
21K073 ah ja.  
21I074 so machst du es und  
21K074 ah ja gut  
21I075 dadrunter liegt wiederum dies mit dem <-> drei Löchern.  
21K075 genau, das äußerste linke Loch vom Dreier.  
21I076 {mhm}<spk: K, gut> also die beiden von dem Dreier, der da noch drunterliegt äh, da schneiden jetzt zwei <-> ja?  
21K076 ja, das ist klar, das haben wir.

## APPENDIX B: A GRAMMAR FRAGMENT FOR THE EXAMPLE DIALOGUE

In this Appendix we introduce the elementary trees associated with the words used in the fragment and discuss some of the more complex linguistic phenomena that occur in it, abstracting away from micro-conversational events. Our LTAG treatment is based on (Abeille and Rambow, 2000; Joshi 2004) and on Frank's view that elementary trees represent extended projections (Frank, 2002). Our semantics is based on Compositional DRT as presented in (Muskens, 1996). Some ideas from Muskens' (2001) *Logical Description Grammar*, an integration of LTAG with CDRT, were also taken into account.

### B.1 A Brief Overview of Compositional DRT

The aim of Muskens in developing Compositional DRT (Muskens, 1996) was to develop a framework in which the dynamics of DRT could be combined with compositional methods of meaning specification without developing an entirely new logic as done, e.g., in Heim's File Change Semantics (Heim 1983) or Groenendijk and Stokhof's Dynamic Predicate Logic (Groenendijk and Stokhof, 1989). Muskens uses the term *grafting* to describe his approach: extend a type logic like those standardly used in Montagovian approaches with the types necessary to provide a reconstruction of the DRT constructs and an (axiomatic) formalization of their properties.

Specifically, the logic proposed by Muskens includes two new types: the type of **discourse referents**  $\pi$  and the type of **states**  $s$ . Discourse referents are used to model the dynamics of context in the same way as they are used in DRT, i.e., in the sense that each noun phrase introduces a new discourse referent. States are used to model contexts themselves, and the way they are modified by natural language sentences; they are the object-language equivalent of the assignments used to formalize the semantics of DRSs in (Kamp and Reyle, 1993). This dynamics is mediated by DRSs, viewed as relations between states. A function  $v: \pi \rightarrow (s \rightarrow e)$  provides the mapping from discourse referents and states to entities, in the sense that  $v(x)(i)$  specifies the 'value' of discourse referent  $x$  at state  $i$ .

Muskens specifies a translation for all DRS constructs in terms of this type logic. The two most important translations are shown in (B.1.1)

- (B.1.1) a.  $R\{x_1 \dots x_n\}$  is short for  $\lambda i. R(v(x_1)(i), \dots v(x_n)(i))$   
 b.  $[x_1 \dots x_n \mid \phi_1 \dots \phi_m]$  is defined as  $\lambda i \lambda j i [x_1 \dots x_n] j \wedge \phi_1(j) \dots \wedge \phi_m(j)$

where  $i [x_1 \dots x_n] j$  is short for  $i$  and  $j$  differ at most over  $[x_1 \dots x_n]$ . For example, the translation of the DRS in (3.1.2) (repeated below as (B.1.2)(a)) is shown in (B.1.2)(b).

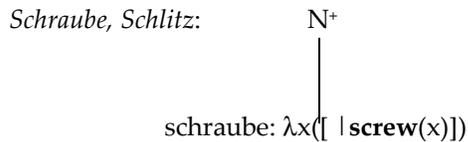
- (B.1.2) a.  $[x, w, y, z, s, s' \mid \mathbf{engine}(x), \mathbf{Avon}(w), s: \mathbf{at}(x, w), \mathbf{boxcar}(y), s': \mathbf{hooked-to}(z, y), z \text{ is } x]$   
 b.  $\lambda i \lambda j i [x, w, y, z, s, s'] j \wedge [\mathbf{engine}(x)](j) \wedge [\mathbf{Avon}(w)](j) \wedge [s: \mathbf{at}(x, w)](j) \wedge [\mathbf{boxcar}(y)](j) \wedge [s': \mathbf{hooked-to}(z, y)](j) \wedge v(z)(j) = v(x)(j)$

In Compositional DRT, DRSs are the translations of sentences—the propositions. Hence, type  $\langle s \rangle$  is the type of propositions, and most translations of NL expressions are chains of functions whose ultimate value is an object of this type: e.g., common nouns have translations of type  $\langle \pi, \langle s, \langle s, t \rangle \rangle \rangle$ . Muskens introduces an abbreviated notation in which  $[t_1 \dots t_n]$  stands for  $\langle t_1, \langle t_2, \dots \langle t_n, \langle s, \langle s, t \rangle \rangle \rangle \rangle$ : in this notation,  $[\ ]$  stands for the type  $\langle s, \langle s, t \rangle \rangle$  of DRSs,  $[\pi]$  is the type  $\langle \pi, \langle s, \langle s, t \rangle \rangle \rangle$  of common nouns,  $[[\pi][\pi]]$  is the type  $\langle \langle \pi, \langle s, \langle s, t \rangle \rangle \rangle, \langle \pi, \langle s, \langle s, t \rangle \rangle \rangle \rangle$  of determiners,  $[[\pi]]$  is the type  $\langle \langle \pi, \langle s, \langle s, t \rangle \rangle \rangle, \langle s, \langle s, t \rangle \rangle \rangle$  of quantifiers, etc. We will use this abbreviated notation in what follows.

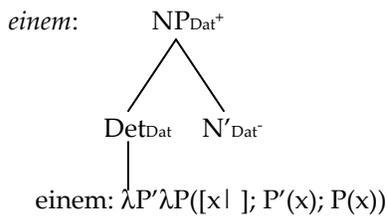
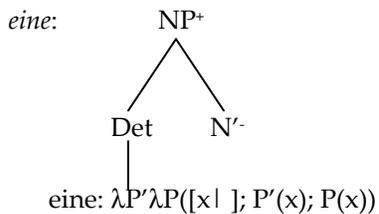
## B.2 LTAG: Basic Trees

### B.2.1 Nominal Phrases

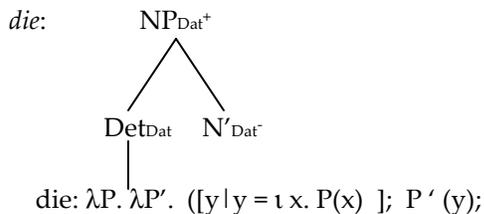
As standard in LTAG, we assume that nominal phrases (NPs) are the projection of determiners rather than of nouns. The elementary tree for noun does not therefore include a projection. We also assume the standard semantic treatment of nouns in CDRT, where nouns are translated as functions of type  $[\pi]$  (i.e.,  $\langle \pi, \langle s, \langle s, t \rangle \rangle$  - see B.1).



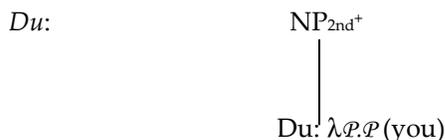
Our treatment of determiners is also derived without modification from their treatment in LTAG and Compositional DRT. Although determiners are not viewed in LTAG as heads of separate Determiner Phrases, but as heads of NPs, the elementary trees associated with determiners do introduce an expectation of an  $N'$ , to which the elementary trees associated with nouns can be attached by substitution. Semantically, determiners are given the standard CDRT type  $[[\pi], [\pi]]$ . Both indefinite and definite determiners introduce new discourse referents, but the latter have a presuppositional component.



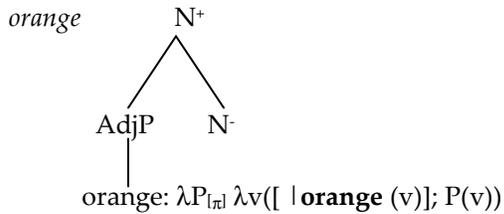
We assume a 'Loebnerian' treatment of definites (see (Poesio and Rieser, in preparation)):



Deictic and anaphoric pronouns introduce their own NPs.



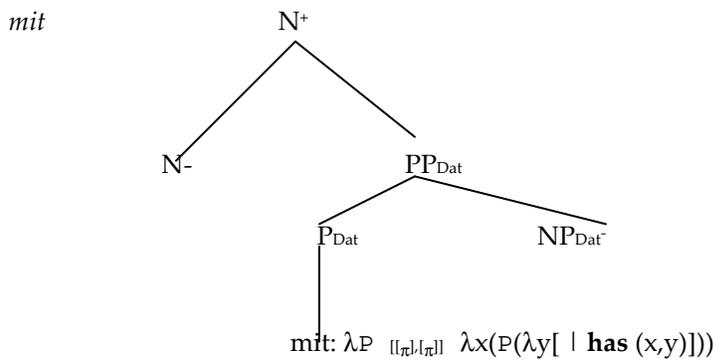
Syntactically, adjectives in LTAG introduce auxiliary (adjunct) trees, incorporated into NPs by adjunction. Semantically, they are treated as functions from predicates to predicates, of type  $[[\pi], \pi]$  ( $\langle\langle\pi, \langle s \langle s \ t \rangle\rangle\rangle, \langle\pi, \langle s \langle s, t \rangle\rangle\rangle$ ).



In German, adjectives can be nominalized, as in the case of *orangene* in 1.3. We treat these nominalized adjectives syntactically and semantically as nouns.

### B.2.2 Prepositional Phrases

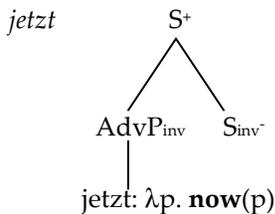
As discussed in Section 3.4, prepositional phrases are treated in LTAG as adjunct trees, incorporated into NPs and VPs by adjunction. Semantically, prepositions are assumed to be functions from quantifiers to predicate modifiers, i.e., to have type  $[[[\pi]\pi]\pi]$ .



### B.2.3 Verbal Phrases and Sentences

In LTAG, sentences are viewed as projections of verbs, just as in HPSG. We'll discuss the lexical interpretation of verbs in the example dialogue later, as they both require some discussion of aspects of German grammar.

Adverbs are interpreted as VP and S modifiers; we show here the elementary (auxiliary) tree involved in the interpretation of *jetzt* discussed in Section 5.2.

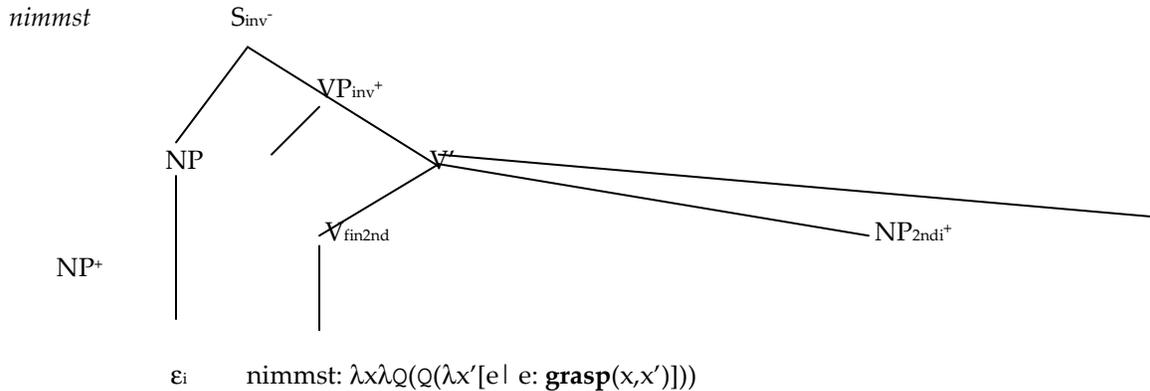


### B.3 Constructions occurring in the first turn: 1.1-1.3

The first turn of the example dialogue contains two linguistic phenomena whose treatment needs some discussion: *inversion* (the fact that the subject of *nimmst* occurs after the verb) and *apposition*. We discuss each in turn.

#### B.3.1 Inversion

Our syntactic treatment of subject-verb inversion is based on the standard treatment of inversions in German adopted, e.g., by most papers in (Freidin, 1992), and is specifically modelled after the proposal by deMena Travis (1992). The elementary tree for the verb *nimmst* shown below includes an empty element in subject position, coindexed with the postverbal NP.



#### B.3.2 Appositions

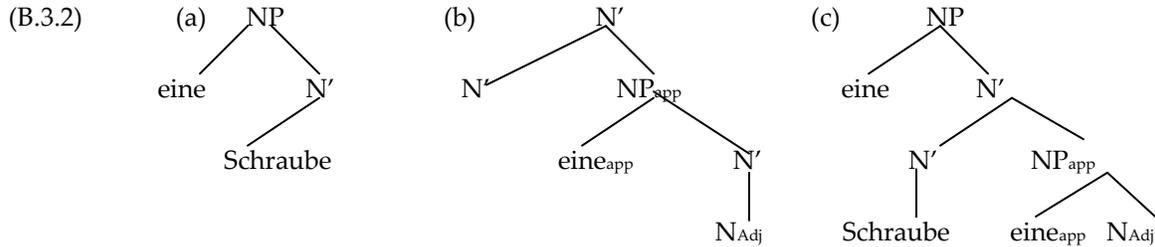
The utterance of a noun phrase in 1.3 would generally be classified as an apposition, a form of parenthetical expressed with a noun phrase. Appositives are extremely common in dialogue, reflecting the fact that utterance generation is incremental, as well; hence an account of this construction is essential for a theory like PTT. Providing a unified account is however quite difficult, as nominal parentheticals have a wide range of forms and uses in addition to those illustrated by the example dialogue, as shown by the examples in (B.3.1). (See Meyer (1992) for an extensive, if purely descriptive, investigation of appositions.)

- (B.3.1) Westinghouse Electric Corp. is wooing Michael H. Jordan,
- a. former head of PepsiCo's international operations / 44 / from Lawrenceville, NJ
  - b. the former head of Pepsico's international operations / a 36-year company veteran to be its new chief executive officer.

As accounting for all these types of nominal parentheticals would be quite a challenge in itself, we only concentrate here on the class of appositions illustrated in the example dialogue, in which the apposition plays a restrictive role. Our treatment of such constructions is consistent with the central claim of Meyer (1992) that syntactically and semantically, such appositions are a type of nominal modification; we formalize this claim by proposing that semantically appositions are of type  $[[\pi] [\pi]]$ , like PPs and other nominal modifiers. (A reminder that nominal predicates, the type  $\langle e, t \rangle$  of Montague grammar, in CDRT are functions from discourse referents to DRs,  $\langle \pi, \langle s, \langle s, t \rangle \rangle$ , abbreviated  $[\pi]$ .) Our main objective here is to explain how the apposition is incrementally integrated with the interpretation of the previous utterances, both from a syntactic perspective and from a compositional semantic point of view.

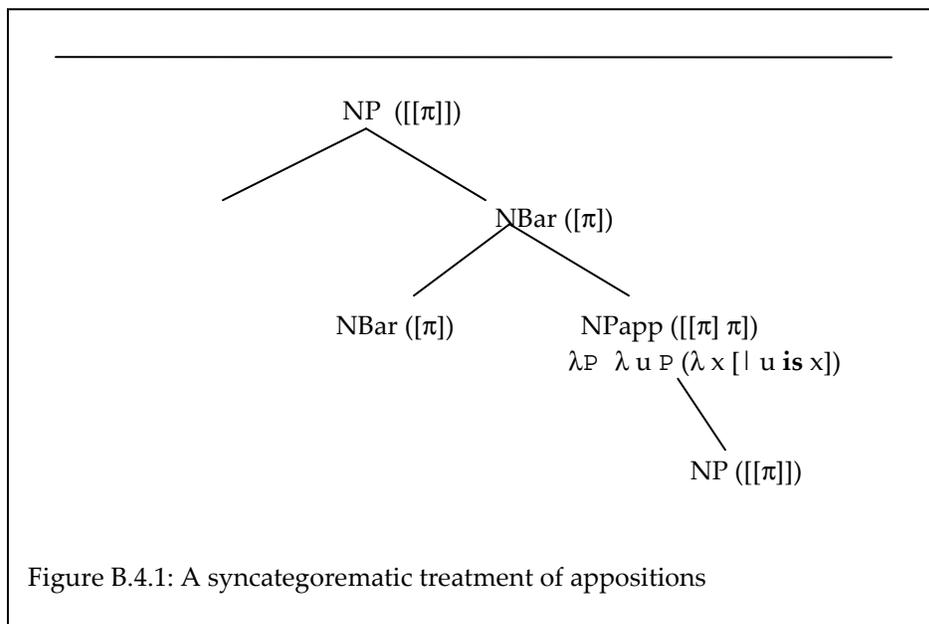
This need to explain how appositions are incorporated from a syntactic perspective is one of our main reasons for adopting a syntactic formalism like LTAG which allows for adjunction operations. The basic idea of our proposed LTAG derivation for *eine Schraube ... eine*

*orangene* is that the final interpretation is obtained by adjoining tree (b) for the apposition to tree (a) for the NP, as shown schematically in (B.3.2): we assume that the syntactic interpretation for 1.3 (shown in (b)—we assume *orangene* is an adjective nominalization, as discussed below) is adjoined to the syntactic interpretation for 1.2 (in (a)), obtaining (c). (The PP *mit einem Schlitz* is omitted for brevity.)



As said above, we follow Meyer in hypothesizing that the class of appositions we are concerned with semantically behave as nominal modifiers, of type  $[[\pi], [\pi]]$ . This still leaves open the question of how certain types of NPs—particularly indefinites and definite NPs—can serve as predicate modifiers. We take these to be cases of NPs whose type has been lowered to that of predicates (Farkas & de Swart, 2003), and exploit the properties of Muskens' Compositional DRT to explain how the restriction expressed by the apposition is grafted onto the appropriate discourse entity.

However, the implementation of this intuition is subject to a few constraints determined by LTAG. LTAG restricts the possible semantic analyses of NPs used predicatively, because only lexical categories are allowed, and the only semantic operation which is allowed is application. Such restrictions prevent treatments in which the predicative meaning of NPs in appositions can be derived syncategorematically—say, by stipulating a structure as in Figure B.4.1, with a semantic operation associated with  $NP_{app}$  that 'lowers' the NP-type interpretation for *eine orangene, mit einem Schlitz* to one of type  $[[\pi], [\pi]]$ .



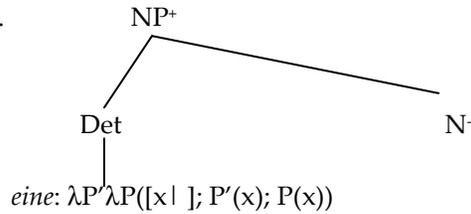
In LTAG, one has to stipulate instead that German *eine* is ambiguous between a normal determiner interpretation, shown in (B.3.3a), and the interpretation used in appositions, of type  $[[\pi], [[\pi], [\pi]]]$  (the CDRT translation of  $\langle\langle e,t\rangle, \langle\langle e,t\rangle, \langle e,t\rangle\rangle\rangle$ ) shown in (B.3.3b).

- (B.3.3) a.  $eine: \lambda P' \lambda P([\lambda x | ]; P'(x); P(x))$   
 b.  $eine_{app}: \lambda P' \lambda P \lambda x([\lambda v | ]; P(x); P'(v); [ | v \text{ is } x])$

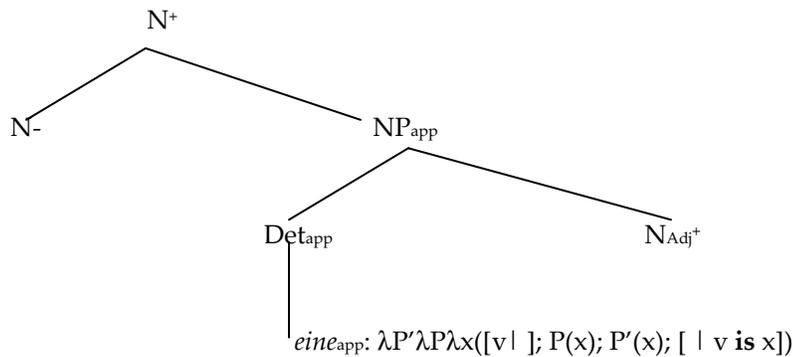
The two interpretations are also associated with the distinct elementary trees in (B.3.4a) and (B.3.4b):

(B.3.4)

a.

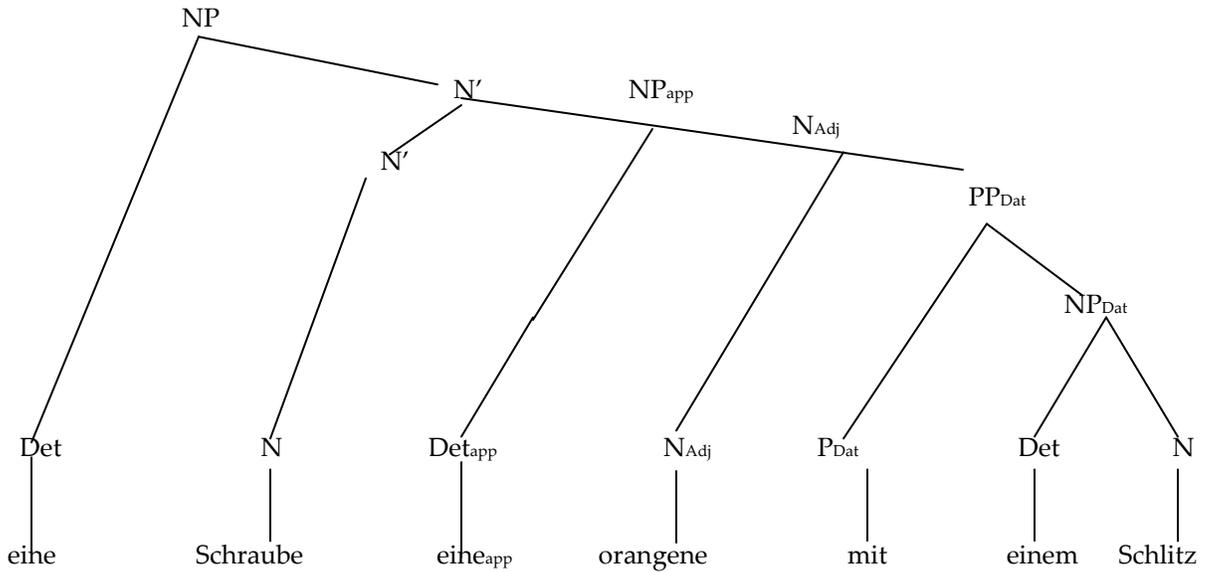


b.



With these assumptions, and assuming standard elementary trees and semantic interpretations for the other words in utterance 1.3, we get the LTAG analysis for *eine Schraube, eine orangene mit einem Schlitz* in (B.3.5).

(B.3.5)



The interpretation of the NP in (B.3.5) is arrived at through the derivation in (B.3.6).

(B.3.6)

$eine_{app}: \lambda P' \lambda P \lambda x ([y | ]; P(x); P'(y); [ | y \text{ is } x])$   
 $orangene: \lambda P \lambda y ([ | \text{orange}(y)]; P(y))$   
 $einem: \lambda P' \lambda P ([x | ]; P'(x); P(x))$   
 $Schlitz: \lambda y ([ | \text{slit}(y)])$   
 $einem(Schlitz): \lambda P' \lambda P ([x | ]; P'(x); P(x)) (\lambda y ([ | \text{slit}(y)])) =$   
 $\lambda P ([x | ]; [ | \text{slit}(x)] ; P(x))$   
 $mit: \lambda P \lambda x (P(\lambda y [ | \text{has}(x,y)]))$   
 $mit'(einem'(Schlitz')): \lambda P \lambda x (P(\lambda y [ | \text{has}(x,y)])) (\lambda P ([z | ]; [ | \text{slit}(z)] ; P(z))) =$   
 $\lambda x (\lambda P ([z | ]; [ | \text{slit}(z)]; P(z)) (\lambda y [ | \text{has}(x,y)])) =$   
 $\lambda x ([z | ]; [ | \text{slit}(z)]; [ | \text{has}(x,z)]) \quad /*renaming of variables$   
 $= \lambda x ([y | ]; [ | \text{slit}(y)]; [ | \text{has}(x,y)])$   
 $Orangene'(mit'(einem'(Schlitz'))):$   
 $\lambda P \lambda z ([ | \text{orange}(z)]; P(z)) (\lambda x ([y | ]; [ | \text{slit}(y)]; [ | \text{has}(x,y)]))$   
 $\lambda z ([ | \text{orange}(z)]; ([y | ]; [ | \text{slit}(y)]; [ | \text{has}(z,y)]))$   
 $eine_{app} (orangene'(mit'(einem'(Schlitz')))):$   
 $\lambda P \lambda x ([z | ]; P(x); [ | \text{orange}(z)]; ([y | ]; [ | \text{slit}(y)]; [ | \text{has}(z,y)]); [ | x \text{ is } z])$   
 $Schraube'(eine_{app}((orangene'(mit'(einem'(Schlitz')))))):$   
 $\lambda x ([z | ]; [ | \text{screw}(x)]; [ | \text{orange}(z)]; x = z; [y | ]; [ | \text{slit}(y)];$   
 $[ | \text{has}(x,y)])$   
 $eine': \lambda P' \lambda P ([x | ]; P'(x); P(x))$   
 $eine'(Schraube'(eine_{app}((orangene'(mit'(einem'(Schlitz'))))))):$   
 $\lambda P' \lambda P ([x | ]; P'(x); P(x)) (\lambda z ([w | ]; [ | \text{screw}(z)]; [ | \text{orange}(w)]; [ | w \text{ is } z];$   
 $[y | ]; [ | \text{slit}(y)]; [ | \text{has}(z,y)]))$   
 $= \lambda P ([x | ]; [v | ]; [ | \text{screw}(x)]; [ | \text{orange}(v)]; [ | v \text{ is } x];$   
 $[y | ]; [ | \text{slit}(y)]; [ | \text{has}(x,y)]; P(x))$



common ground). And of course it could also be argued that assuming the same turn-taking translation for *well*, *now*, and *umm* is another big simplification.

*Grounding acts Ja* in 1.4 is an explicit acknowledgment. Again, it is assumed in PTT that at least some such interpretations are lexically encoded. The interpretation used in 1.4 is shown in (B.3.8).

(B.3.8)  $ja \rightarrow [u, ce, DU \mid u: \mathbf{utter}(A, "ja"), ce: \mathbf{ack}(A, DU), \mathbf{generate}(u, ce)]$