SDRT and Multi-modal Situated Communication

Andy Lücking **Hannes Rieser** **Marc Staudacher**

Bielefeld University, CRC 360 "Situated Artifical Communicators", B3

{andy.luecking,hannes.rieser,marc.staudacher}@uni-bielefeld.de

Abstract

Classical SDRT (Asher and Lascarides, 2003) discussed essential features of dialogue like adjacency pairs or corrections and up-dating. Recent work in SDRT (Asher, 2002, 2005) aims at the description of natural dialogue. We use this work to model situated communication, *i.e.* dialogue, in which sub-sentential utterances and gestures (pointing and grasping) are used as conventional modes of communication. We show that in addition to cognitive modelling in SDRT, capturing mental states and speech-act related goals, special postulates are needed to extract meaning out of contexts. Gestural meaning anchors Discourse Referents in contextually given domains. Both sorts of meaning are fused with the meaning of fragments to get at fully developed dialogue moves. This task accomplished, the standard SDRT machinery, tagged SDRSs, rhetorical relations, the up-date mechanism, and the Maximize Discourse Coherence constraint generate coherent structures. In sum, meanings from different verbal and non-verbal sources are assembled using extended SDRT to form coherent wholes.

1 Introduction

Recently, the interest in retrieval and representation of *non-sentential speech* has been growing, as the collection (Elugardo and Stainton, 2005) shows. The debate on how to account properly for the phenomena is still ongoing. However, it emerges that it puts further constraints on how mainstream linguistics *should* be done. Non-sentential speech is an essential part of language.[1] Thus, notions such as grammaticality and coherence have to be applicable to it. In this paper, we are concerned more specifically with issues of the semantics/pragmatics-interface of non-sentential speech. We understand this kind of language use as being part of *situated communication* and propose a theory for it. Thus, we start by characterising situated communication. Consider the two examples (1) and (2).

(1) World economic growth slowed noticeably in 2005 from the strong expansion in 2004.

(2) In a two-person dialogue between I and C in a room with some bolts on a table:
   a. I: This bolt in the rear there (while I is pointing)
   b. C: This one? (while C is grasping some bolt)
   c. I: Yes

In opposition to (1), the kind of language use as in (2) is what we call *situated communication*.[2] Language use of this kind can be recognized by a couple of characteristics. First, utterances are typically sub-sentences and not "full-fledged sentences" in a grammatical sense. On a standard account, only sentences (and not parts) express propositions. Still, sub-sentences can be used to express propositions. For example, (2-a) says of a particular bolt on the table that it is the one to be grasped. So, after all, utterances of sub-sentences can express propositional content.

Secondly, such utterances are typically accompanied by linguistically relevant non-verbal behaviour such as pointing gestures or graspings. Deixis is typical for this kind of language use. In (2-b), for example, it is asked of a certain bolt on the table whether it is the one I meant in (2-a). To establish the reference to that bolt, C's grasping seems to be essential.

Thirdly, such utterances as in (2) can be used to perform speech acts. It can be meaningfully asked what the illocutionary role of such an utterance is (*e.g.* (2-b) is a `Check-back`) and which proposition is thereby expressed. However, it cannot be a property of the expression's content that makes it express a certain speech act or proposition. For example consider an utterance of 'scissors' in a sewing shop, in the rock-paper-scissors-game, or

---

[1] For data on its frequent use, see (Schlangen and Lascarides, 2003). In our corpus (see (Lücking and Stegmann, 2005, p. 15)) 50 instructor's requests were realized as definite NPs out of a total of 92 dialogue moves including acceptances and repairs.

[2] In contrast, the use of some fragments such as question-answer- or request-answer-pairs is determined by rules of grammar. We are interested in cases which are extremely context dependent and need inference for their resolution. These are cases calling for "resolution-via-inference" in the terms of the Schlangen and Lascarides (2003) approach.

on a shopping list. In each context, the utterance is used to express something different. While the first two can be taken to express a proposition, the inscription on the list cannot. It might merely be some mnemonic device to perform the shopping. Moreover, the different uses of 'scissors' seem to be governed by *conventions*. So, a special stock of conventions seems to regulate its interpretation. Being conventions each of them is mutually believed (in some dispositional sense). Together they allow agents the use of sub-sentential utterances and gestures to successfully communicate as (2) shows.

From these three properties of situated communication we derive the minimal requirements for a theory of situated communication. As a framework we are going to use SDRT. Given the use of non-sententials and nonverbal behaviour, the theory has to explain which sentential content a non-sentential utterance expresses and which dialogue move is performed. The explanation has to make use of a special stock of conventions. Moreover, discourse coherence should be explained.

For purposes of illustration we use discourse (3) as our main example:

(3)   a.   *I: Die rote $\searrow_a$ Holzscheibe*
            I: The red $\searrow_a$ wooden disc
      b.   *C: $\Downarrow_a$ Diese?*
            C: $\Downarrow_a$ This one?
      c.   *I: Ja*
            I: Yes

Some comment about (3) is in order. Dialogues like (3) are called *Object Identification Games* and have been examined in project B3 of the Collaborative Research Centre "Situated Artificial Communicators" (SFB 360)[3].

In (3-a) the symbol '$\searrow_a$' indicates, when the stroke of a pointing gesture occurs. The symbol is written after the word whose occurrence is immediately preceding in time. The index indicates the object a the pointing refers to. Likewise in (3-b), the symbol '$\Downarrow_a$' indicates the grasping of the object a. Two video-stills showing the pointing and the grasping in (3) are provided in Fig. 1(a) and 1(b), respectively.

(3) is a gloss for a corpus entry which has been built from the experimental data. Each corpus entry is a description of a dialogue which occurred in the experimental setting. The corpus annotation format features both verbal and non-verbal el-

ements in such a way that the role of pointing gestures can be studied theoretically. Fig. 1(c) shows a graphical representation of a corpus entry.

In Object Identification Games, two persons, the *instructor* (I) and the *constructor* (C), are involved in a coordination task. It is a two-player-game of spotting an object in a given situation. The instructor has the role of the "description-giver". The constructor has the role of the "object-identifier". The players interact by performing moves in the game. The game starts with the instructor's choosing a certain object out of the parts of a toy air plane spread on a table. She instructs the constructor to identify the object she has chosen by referring to it. The constructor then has to resolve the instructor's reference act and to give feedback. Thereby, reference has to be negotiated and established using a special kind of dialogue game. The game ends, if the constructor has identified the correct object on the table and the instructor has accepted it.

This paper is organised as follows. We first introduce Standard SDRT. Next, we discuss recent SDRT developments and introduce the concepts doing the explanatory work. We then apply the theory to our main example. In the remaining sections, related research is reviewed and ideas for linking SDRT to Logical Description Grammars are presented.

2 Exposition of Standard SDRT

As a dynamic discourse representation theory modelling the semantics/pragmatics-interface, SDRT is an apt framework for modelling situated communication. For our purposes it is important to note that "standard" SDRT as presented in the 2003 book (Asher and Lascarides, 2003) requires its input to be of a type corresponding to sentences in the grammar. To understand this point we illustrate SDRT's general architecture (Fig. 2) and its implicit notion of discourse construction using the sample dialogue (3).

Since SDRT provides no grammar, the *NL-input* is assumed to be available as *underspecified logical forms (ULFs)* constructed by a *parser*. The underspecification reflects the fact that, in general, the grammar does not determine a unique logical form but a set of possible forms corresponding to the interpretation licensed by the grammar alone. ULFs describe base logical forms, *i.e. SDRSs*.

So, SDRT's processing begins by assuming

---

[3]`http://www.sfb360.uni-bielefeld.de/`

(a) The pointing in (2a).   (b) The grasping in (2b).   (c) The annotation in TASX.
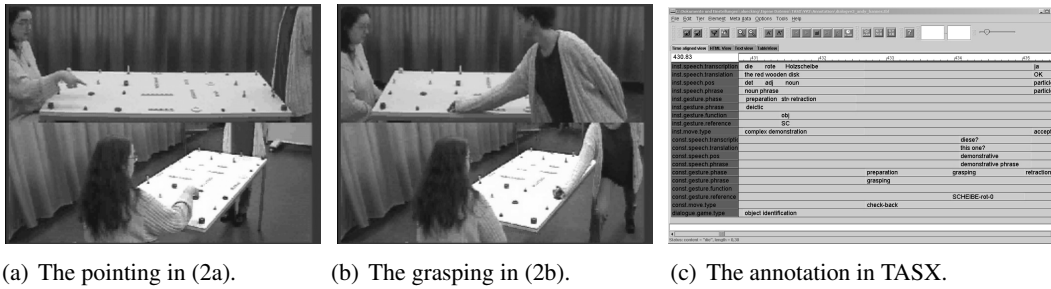
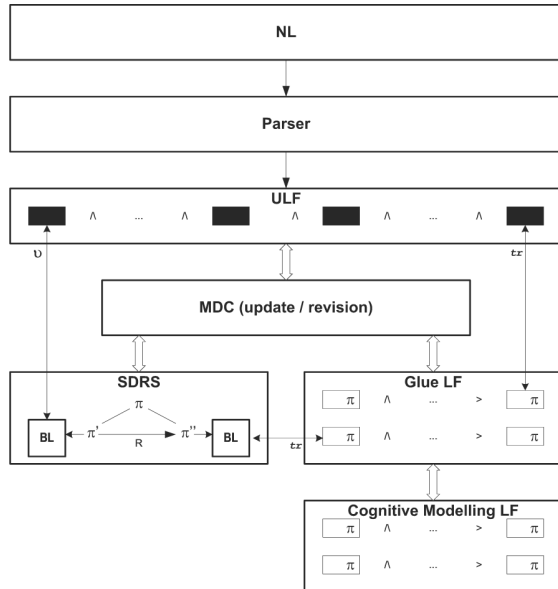Figure 1: Annotation of a natural dialogue in project B3 of SFB 360.

Figure 2: The general SDRT architecture.

some context *C* (a potentially empty description of SDRSs) and by assuming the ULF of the dialogue's first utterance (3-a) being part of it. In the next step, *C* is updated with the ULF of (3-b) yielding a new context *C′*. SDRT's *update mechanism* assumes that the new utterance is related to an available attachment point by means of an underspecified *discourse relation* $R(a,b)$. In this case, the most coherent one is (3-a). However, discourse relations relate only content having sentential satisfaction conditions.

At this point SDRT fails with sub-sentential utterances, if they just have their compositional meaning. The interpretations of (3-b) licensed by the grammar alone are not contents having sentential satisfaction conditions. Intuitively, the new context *C′* describes SDRSs in which *a* and *b* are related through some admissible resolution of *R*, e.g. $Q-Elab$. So, what $Q-Elab$ should relate is of the wrong semantic type. Something having

satisfaction conditions is required, however in the case of (3-b) an NP-denotation is present.

To make the illustration of the general architecture complete, let us assume that (3-b) had a sentential content. Then its ULF (*inter alia*) would be translated to the *Glue logic* and to the *Cognitive Modelling Logic* in order to resolve underspecification by pragmatic reasoning. The resolutions are translated back to the logic of ULF and added to the description. The update mechanism restricts the resolutions to those that are consistent, *i.e.* describe well-formed SDRSs.

3   Recent SDRT Developments

Recently progress has been made on how difficult NL dialogue data can be handled using SDRT's full theoretical power plus some additional assumptions. For our purposes three papers are of relevance: (Asher, 2002) on Deixis, Binding and Presupposition, (Asher, 2005), Bielefeld Lectures on SDRT, and (Schlangen and Lascarides, 2003)'s work on the Interpretation of Non-Sentential Utterances in Dialogue.

Asher (2002) deals with the following issues, relevant for our example: treatment of presuppositions, analysis of definite descriptions, especially their deictic uses, anchoring of definites in the non-linguistic context, the notion of internal and external anchors, the relation between anchoring and speech act related goals (SARGs), the cognitive effects of anchoring, the generation of mutual belief with regard to an object anchored. These concepts are briefly and somewhat fragmentarily introduced below.

As to presuppositions, Asher argues that the Heim-van der Sandt-Geurts account is incomplete and yields wrong predictions, the reason being that presupposition accommodation in the case of deictically used NPs is not always adequate. Definite descriptions introduce an underspecified relation, called *bridging relation*, between the referent and some other contextually given object, set

to identity by default.[4] Deictically used definites have to be anchored to some object in the non-linguistic context. As a consequence, anchoring involves a *de re* attitude towards the object, some sort of *knowing how* needed to solve the conversational goals (SARGs) of the speaker. SDRT uses, in opposition to specifying anchoring contextually as undertaken in Kaplan's Context Theory or Situation Semantics, DRT's external and internal anchors (Kamp, 1990). Anchoring requires linking an agent A's epistemic attitude to conversational goals. If an anchoring relation between the presupposition of a definite ψ and some element in the discourse context exists for the agent A, he is supposed to have a computable means of getting to the referent of ψ from the present non-linguistic context of utterance under some given purpose ϕ; to capture this, a notion of *path* is defined. If the anchoring function of a deictically used definite is accepted by the participants in dialogue, they are assumed to mutually believe that the definite picks out the same object for them. Hence, anchoring amounts to coordination or alignment.

Of similar importance as the discussion of definites, presupposition, binding and anchoring is the handling of fragments in dialogue, since, normally, natural dialogue does not come with utterances which can be mapped onto well-formed sentences in the theory of grammar sense. The idea in (Asher, 2005) is that fragments can be resolved iff the context in which the communication is situated provides us with two things: First, it must be mutual knowledge that a fragment with some meaning has been produced by an agent and secondly, it must be mutually believed that the fragment as produced expresses some more comprehensive content ϕ wrapped around the information reconstructed as a presupposition. In our example, the more comprehensive content ϕ is given by 'Grasp the red wooden disc!' and 'This one?', respectively. The status of these assumptions in the theoretical set up of SDRT is not yet clear, presumably, they belong to Cognitive Modelling, since mental states are involved.

Another approach to fragments is elaborated in (Schlangen and Lascarides, 2003). The idea is to assimilate sub-sentential utterances to sentences since such utterances express sentential content. Thereby, the problem discussed in the last section can be circumvented. From their point of

view such utterances have "holes" which need to be filled in in order to express the intended content. Schlangen and Lascarides understand hole-fillers as the resolution of semantically underspecified content (and as such these are not syntactic ellipses). *I.e.* the linguistic form of such utterances is of the category "sentence fragment" which in turn consists of the usual linguistic items such as an NP. The logical form is assumed to have a semantically underspecified relation linking its variables such that each resolution expresses sentential contents, among them the intended one. Schlangen and Lascarides' main thesis is that the resolution of such utterances can be modelled as a by-product of establishing coherence in discourse.

Schlangen and Lascarides found that their approach is more problematic with regard to sub-sentential utterances which need a "resolution-via-inference", *i.e.* a resolution that cannot use the immediate linguistic context containing a "copy" of the material needed (as in the case of short answers to wh-questions). The reason is simply that domain-specific knowledge is necessary.

We, following Asher (2002, 2005), propose a new direction for accounting for this class of utterances. We don't treat such utterances as sentences. Our thesis is that competent speakers have linguistic knowledge in form of situated conventions allowing the speakers to properly use and understand such utterances. Moreover, our original data shows that the role of gestures and graspings is central to correctly resolve newly introduced definites. Without a notion of external anchoring resolution cannot be explained correctly. As a by-product of the introduction of the notion of external anchoring resolution-via-inference becomes more tractable.

4 Coherence from the 2005 SDRT Perspective: A Giant Step for SDRT

SDRT's notion of coherence up to (Asher, 2005) rested on several mechanisms, the use of rhetorical relations and their semantics, especially the division into coordinating and subordinating relations, the use of SDRSs as part of context change potentials in the Kamp-Heim-tradition, the extended definition of up-date capturing revision in dialogue and, finally, the filter mechanism "Maximize Discourse Coherence" (MDC). All these notions were ultimately founded upon the notion of *complete* meaning, of whatever type and however expli-

---

[4]See (Asher and Lascarides, 1998) for more on bridging.

cated. These meanings in turn were conceived of as coming solely from verbal expressions using a construction algorithm in DRT fashion.

This picture fundamentally changes with 2005's SDRT: First of all, the information provided by the fragments of the description giver 'the red wooden disc' and the object identifier 'this one?', respectively, are not complete. The intuition is that the fragments combine with meanings from the context to give us complete meanings. Roughly, we want 'Grasp the red wooden disc!' on the description giver's and 'Do you mean this one?' on the object identifier's side. Once we arrive at complete meanings, the normal SDRT machinery can be put to work again. However, in order to get there, we have to use special postulates, which under specific conditions let agents in cooperative dialogue use these fragments as directives and clarification questions, respectively. (Of course, our account is not restricted to directives and clarification questions. Other postulates would allow other uses.) The missing information for the directive comes from the context at the beginning of the object identification game, in which the director of the experiment assigns the roles of description giver and object identifier, saying for example, 'you, A, tell the other one to grasp one of the objects in the domain' and 'you, B, identify the object described, pointed at etc. and indicate whether you have identified it'. These roles are preserved throughout the contexts developed, at least as a fall back option. In terms of SDRT: The director of the experiment fixes the type of the speech-act-related goals (SARGs) of the participating agents. Secondly, the dialogue is multi-modal as the example shows, the object introduced by the description is anchored to the context by the demonstration. Similarly, the pure demonstrative used in the clarification question is anchored to the context by the object identifier's grasping. Definiteness information is treated as presuppositional, entertaining the idea that presuppositions are locally bound.

On the whole, detailed context information plays a much greater role in the 2005 SDRT version as compared to the standard one, due to the fact that the meaning of the fragments has to be filled up using context information.

5 Tying Things Together

We now apply the theory to our main example (3) using a DRT-style notation. The application of the theory shows how demonstrations, discourse relations, a special stock of conventions and MDC interact in order to arrive at the intended interpretations. We assume that in the context of Object Identification Games a special stock of conventions holds which are represented as axioms of the following form:

$$K_{I,C}(\alpha(\pi_1) \wedge Ag(\pi_1) = I \wedge (MB_{I,C}(\alpha(\pi_1) \wedge$$
$$Ag(\pi_1) = I) \rightarrow Say_I(p_\phi))) \rightarrow \alpha(\pi_1) \text{ resolves to } \phi$$

Such conventions express linguistic knowledge which competent communicators in Object Identification Games are assumed to have. The formula can be read as follows: If both communicators I and C know ('$K_{I,C}$') that if I utters α and if it is mutually believed ('$MB_{I,C}$') that if I utters α she says that ϕ, then α resolves to ϕ.

In our example dialogue, we assume that the following convention holds: If I utters 'The red wooden disc' in (3-a), it resolves to the directive addressed to C that she should grasp the object referred to. Of course, not all NP-utterances are directives. So, the relevant convention has to be restricted to situations of a certain type.

Recall that SDRT distinguishes with regard to definites between *presupposed and asserted information*. Consequently, the utterance of (3-a) gives us the presupposed information π_{1p} in α and the asserted information π_{1a} in α. π_{1a} in α should be read as 'There is an SDRS but I don't know which one'. ϕ, in turn, expresses what the utterance α resolves to if the antecedent of the axiom holds.
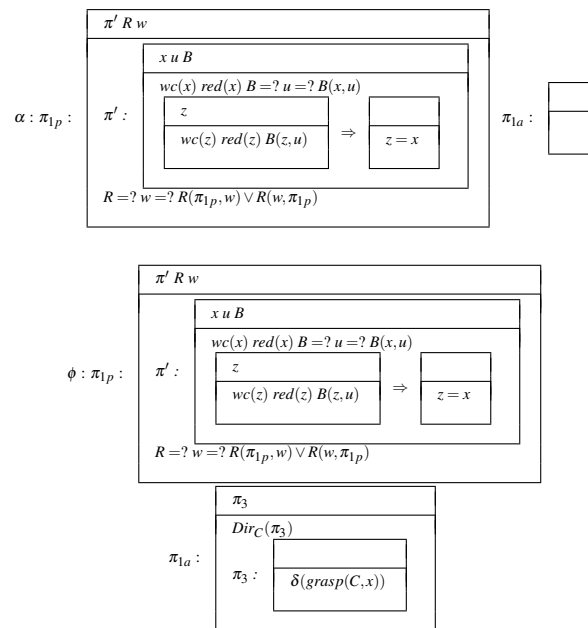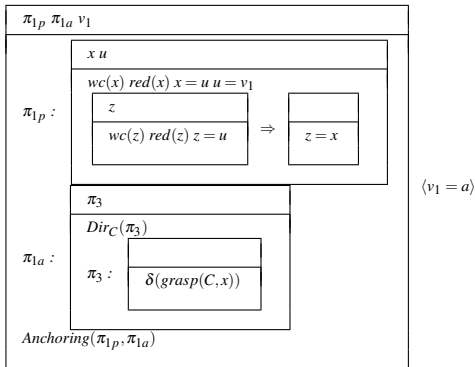
So, ϕ is what we get from the application of the linguistic information to the special convention. We assume a speech act theory style imperative semantics. Consequently, Dir_C is to be read as 'C is commanded that ...' and $\delta(grasp(C,x))$ in π_3 is the action commanded, namely that agent C grasp x. For the next step, we have to say how we represent gestural information. The pointing in (3-a) provides very little content. It merely relates some discourse referent to some external object:

$$\boxed{\begin{array}{c} v_1 \\ \hline \\ \end{array}} \quad \langle v_1 = a \rangle$$
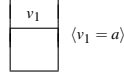
Combining the linguistic and gestural information, the result of an apt multi-modal integration strategy is:

$$\boxed{\begin{array}{l} \pi_{1p}\ \pi_{1a}\ v_1 \\ \pi_{1p}: \quad \boxed{\begin{array}{l} \pi'\ R\ w \\ \pi': \quad \boxed{\begin{array}{l} x\ u\ B \\ wc(x)\ red(x)\ B =?\ u =?\ B(x,u) \\ \boxed{z \quad wc(z)\ red(z)\ B(z,u)} \Rightarrow \boxed{z = x} \end{array}} \\ R =?\ w =?\ R(\pi_{1p},w) \vee R(w,\pi_{1p}) \end{array}} \\ \pi_{1a}: \quad \boxed{\begin{array}{l} \pi_3 \\ Dir_C(\pi_3) \\ \pi_3: \quad \boxed{\delta(grasp(C,x))} \end{array}} \end{array}} \quad \langle v_1 = a \rangle$$
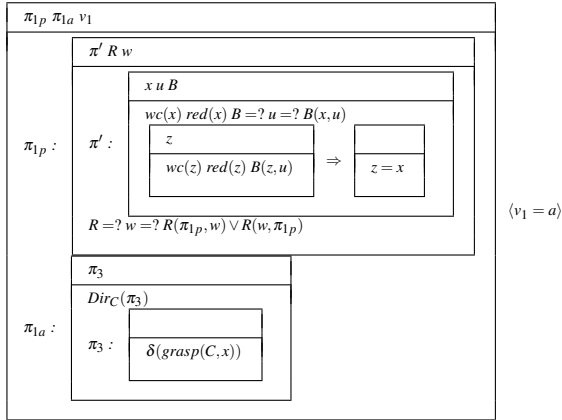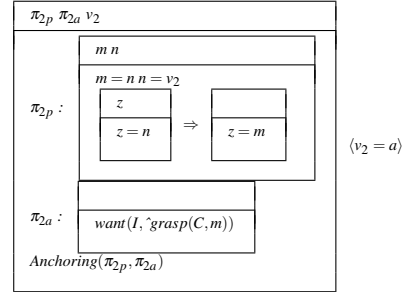
Now, underspecification can be resolved by using a tacit best-update-strategy. Thereby, we resolve the B-relation to identity ($\lambda x.\lambda y.x = y$), u to the externally anchored v_1, w to π_{1a} and R to *Anchoring*. Thus we get:

$$\boxed{\begin{array}{l} \pi_{1p}\ \pi_{1a}\ v_1 \\ \pi_{1p}: \quad \boxed{\begin{array}{l} x\ u \\ wc(x)\ red(x)\ x = u\ u = v_1 \\ \boxed{z \quad wc(z)\ red(z)\ z = u} \Rightarrow \boxed{z = x} \end{array}} \\ \pi_{1a}: \quad \boxed{\begin{array}{l} \pi_3 \\ Dir_C(\pi_3) \\ \pi_3: \quad \boxed{\delta(grasp(C,x))} \end{array}} \\ Anchoring(\pi_{1p},\pi_{1a}) \end{array}} \quad \langle v_1 = a \rangle$$

So, in the first turn I introduces a discourse referent v_1 which is externally anchored to the wooden disc a. The directive in π_{1a} presupposes that there is some object which can be grasped by C. The presupposition is satisfied through best-update's resolution of R to *Anchoring* in such a way that π_{1p} anchors π_{1a}.

The next turn is analysed similarly. There is, likewise, a special convention regulating the interpretation of (3-b) which says that when C utters the deictic 'Diese?' she thereby says that she wants to satisfy the directive. Combining the presupposed and the asserted content as before we get:

$$\boxed{\begin{array}{l} \pi_{2p}\ \pi_{2a}\ v_2 \\ \pi_{2p}: \quad \boxed{\begin{array}{l} m\ n \\ m = n\ n = v_2 \\ \boxed{z \quad z = n} \Rightarrow \boxed{z = m} \end{array}} \\ \pi_{2a}: \quad \boxed{want(I,\,\hat{}grasp(C,m))} \\ Anchoring(\pi_{2p},\pi_{2a}) \end{array}} \quad \langle v_2 = a \rangle$$

So, part of what the grasping does is that it externally anchors v_2. However, it seems that graspings have richer but underspecified content since they can be used to perform many things. We reflect this by assigning a highly underspecified content of type "action" to it:

$$\boxed{\begin{array}{l} v_2 \\ ?_{action}(C,v_2) \end{array}} \quad \langle v_2 = a \rangle$$
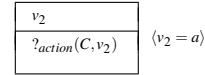
In our dialogue, the grasping presumably carries out the action demanded by I. This suggests that the grasping in (3-b) is used to satisfy I's request in π_{1a} and part of its SARG. Using best update, this amounts to saying that $?_{action}(C,v_2)$ resolves to $grasp(C,v_2)$ and that $Sat-Request(\pi_{2g},\pi_{1a})$ holds. Thus the grasping elaborates on π_{2a} yielding $Q-Elab(\pi_{2g},\pi_{2a})$. Usual reasoning additionally gives us $Q-Elab(\pi_{1a},\pi_{2a})$ and explains why 'This one?' in (3-b) is uttered. While the grasping satisfies the directive (see $Sat-Request$), it might not be *mutually believed* that it is satisfied. So, if $Q-Elab(\pi_{1a},\pi_{2a})$ holds, it also mutually believed that it does (using SDRT's axiom schemata Sincerity, Competence and Mutual Belief). Moreover, by SARG-transitivity, the SARG of π_{1a} is (part of) the SARG of π_{2a}. Thus by satisfying π_{2a}'s SARG the SARG of π_{1a} is satisfied. So, finally, we get the resulting SDRS in Fig. 3.

6 Related Research

Dealing with natural multi-modal dialogue in our paper, we touch on several research areas. Leaving out special SDRT literature here, the focus
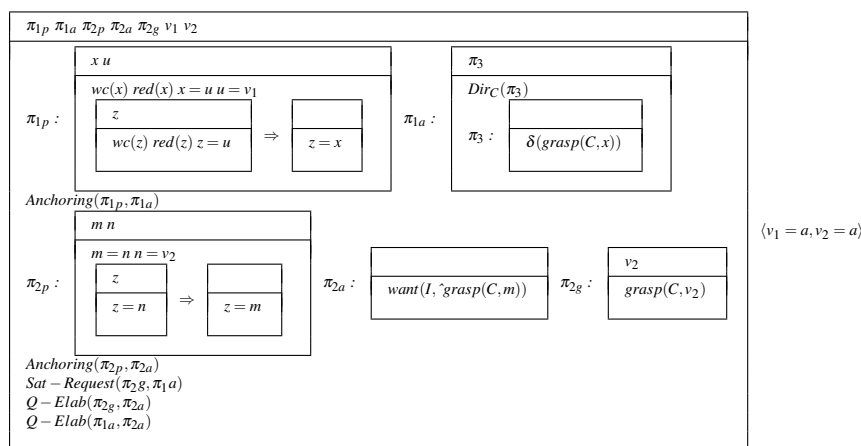
$\pi_{1p}\ \pi_{1a}\ \pi_{2p}\ \pi_{2a}\ \pi_{2g}\ v_1\ v_2$

$\pi_{1p}:$ | $x\ u$ — $wc(x)\ red(x)\ x=u\ u=v_1$ | z / $wc(z)\ red(z)\ z=u$ $\Rightarrow$ $z=x$

$\pi_{1a}:$ | π_3 — $Dir_C(\pi_3)$ | $\pi_3:$ $\delta(grasp(C,x))$

$Anchoring(\pi_{1p},\pi_{1a})$

$\pi_{2p}:$ | $m\ n$ — $m=n\ n=v_2$ | z / $z=n$ $\Rightarrow$ $z=m$

$\pi_{2a}:$ $want(I,{}^{\smallfrown}grasp(C,m))$

$\pi_{2g}:$ v_2 / $grasp(C,v_2)$

$Anchoring(\pi_{2p},\pi_{2a})$
$Sat-Request(\pi_{2g},\pi_1 a)$
$Q-Elab(\pi_{2g},\pi_{2a})$
$Q-Elab(\pi_{1a},\pi_{2a})$

$\langle v_1=a, v_2=a\rangle$

Figure 3: The resulting SDRS.

is on grammar-in-dialogue, description of fragments, and problems of integrating information from other channels.

The issue of syntax-in-dialogue was treated by Schegloff (1979) from the perspective of discourse analysis, mainly focussing on hesitations, restarts, turn construction, and repairs. Clark and Wilkes-Gibbs (1990) generalised the ethno-methodological approach and studied cooperation in syntax production, formulating principles of co-operative contributions for NPs-in-dialogue. A corpus investigation from the perspective of syntax cooperation is provided in (Skuplik, 1999). Fine tuned coordination on all grammatical levels, named 'alignment', forms the backbone of Pickering and Garrod (2004)'s theory, completions and fragments being their favourite examples for establishing implicit common ground. Based on (Skuplik, 1999) and hooking up to SDRT, change of speaker roles, completions and inference in task-oriented dialogue were studied in (Poncin and Rieser, 2000) using Von Wright's Practical Syllogism and Asher and Morreau's Default Inference. A reconstruction of completions and similar phenomena within PTT is undertaken in (Poesio and Rieser, 2006). Recently, even if restricted to sentences/propositions, the interest in retrieval and representation of fragmentary information has been growing, as the collection of articles in (Elugardo and Stainton, 2005) and their introduction to the volume shows. Above all, representation of ellipsis and fragmentary information has been investigated in the paradigm of Dynamic Syntax (Cann et al., 2005; Purver et al., 2005; Purver and Kempson, 2004) for some time, using advanced theory of grammar.

Since SDRT does not come with a worked out construction algorithm, it does not have a multi-modal interface. Its contribution to multi-modality issues lies therefore in applying the separation of presuppositional versus assertional information and especially in the notion of anchoring. Principles of interface construction and compositionality matters concerning speech and gesture integration are discussed in (Lücking et al., 2006), see also (Rieser, 2004, 2005), where one can see which problems have to be overcome. Once the mapping from verbal expressions to SDRSs is organised, SDRT could, in principle, be part of an MM interface.

7 Ideas for Linking SDRT Logical Description Grammars (LDGs)

Having sketched how non-sentential utterances can be accounted for from within SDRT, we now address the question how dialogue maps to ULF. In (Asher and Lascarides, 2003, p. 122) it is assumed that some syntax-semantics-interface maps verbal input into ULF, which, judging from the set-up of SDRT (p. 431), forms its bottom layer. ULFs have models in the logic of information content, represented as SDRSs of some sort. In the simple case, where we have no underspecification, we get only one model. In order to get the mapping from language to ULF going, we can start from Muskens' concept of Logical Description Grammars (Muskens, 2001). LDGs use a version of Lexicalized Tree Adjoining Grammar (LTAG) which can capture underspecification in a similar way as the Constraint Language for Lambda Structures (Egg et al., 2001) does, for example concerning PP-attachment, quantifier am-

biguity and polysemy. The semantic structures which we can use to tag LTAG-trees can be either type-logical formulae, as in (Muskens, 2001) or DRSs in the style of compositional DRT as in (Muskens, 1996). These we can take as substitutes for single SDRSs. Underspecification could arise due to syntactic structure or semantic ambiguity, *i.e.* we could get several SDRSs for one LTAG-formula. Once we reach this level, we seem to be done, since ULFs can be translated into Glue Logic, the place where the axioms substantiating admissible rhetorical relations are introduced.

We haven't yet tested this assumption in detail, we hope to report about it in the workshop. Observe that with respect to our example we have to face additional problems due to the fragments encountered. As a consequence, we would have to use additional axioms in our mapping process.

8 Conclusion and Further Work

We have given a first sketch of a theory of situated communication by means of SDRT plus special conventions used to determine the communicative meaning of non-sentential utterances. The theory relates gestures to NPs by way of presupposition representation, Anchoring and MDC. It remains to be seen whether alternative grasping representations are better suited to the project. Further work relates to statistical investigations concerning fragments of dialogue moves, coverage results and generalisability.

Acknowledgements

References

Nicholas Asher. Deixis, Binding and Presupposition. forthcoming in: Festschrift for Hans Kamp, 2002.

Nicholas Asher. Bielefeld Lectures on SDRT, 2005.

Nicholas Asher and Alex Lascarides. *Logics of Conversation*. Cambridge University Press, Cambridge, 2003.

Nicholas Asher and Alex Lascarides. Bridging. *Journal of Semantics*, 15(1):83–113, 1998.

Ronnie Cann, Ruth Kempson, and Lutz Martin. *The dynamics of language: an introduction*. Syntax and Semantics; 35. Elsevier, Amsterdam [a.o.], 2005.

Herbert H. Clark and Deanna Wilkes-Gibbs. Referring as a collaborative process. In Philip R. Cohen, Jerry Morgan,

and Martha E. Pollack, editors, *Intention in communication*, pages 463–493. MIT Press, Cambridge, MA, 1990.

Markus Egg, Alexander Koller, and Joachim Niehren. The constraint language for lambda structures. *Journal of Logic, Language, and Information*, 10(4):457–485, 2001.

Reinaldo Elugardo and Robert J. Stainton, editors. *Ellipsis and nonsentential speech*, volume 81 of *Studies in Linguistics and Philosophy*. Springer, Dordrecht [a.o.], 2005.

Hans Kamp. Prolegomena to a structural theory of belief and other attitudes. In C. Anthony Anderson and Joseph Owens, editors, *Propositional attitudes: the role of content in logic, language, and mind*. CSLI, Stanford, 1990.

Andy Lücking, Hannes Rieser, and Marc Staudacher. Multimodal integration. Brandial'06, to appear, 2006.

Andy Lücking and Jens Stegmann. Assessing reliability on annotations (2): Statistical results for the DEIKON scheme. Technical Report 3, SFB 360, Bielefeld University, 2005.

Reinhard Muskens. Montague semantics and discourse representation. *Linguistics and Philosophy*, 19:143–186, 1996.

Reinhard Muskens. Talking about trees and truth-conditions. *Journal of Logic, Language and Information*, 10(4):417–455, 2001.

Martin J. Pickering and Simon Garrod. Towards a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(2):169–190, 2004.

Massimo Poesio and Hannes Rieser. Prolegomena to a theory of completions, continuations, and coordination in dialogue. in prep., 2006.

Kristina Poncin and Hannes Rieser. Multi-speaker utterances and coordination in task-oriented dialogue. Report 2000/02, SFB 360, Bielefeld University, to appear in JoP 751. Technical Report 2000/06, SFB 360, Bielefeld University, 2000.

Matthew Purver and Ruth Kempson. Incremental parsing, or incremental grammar? In *Proceedings of the ACL Workshop on Incremental Parsing: Bringing Engineering and Cognition Together*, pages 74–81. Barcelona, 2004.

Matthew Purver, Ronnie Cann, and Ruth Kempson. Grammars as parsers: meeting the dialogue challenge. To appear in: Research on Language and Computation, 2005.

Hannes Rieser. Pointing in dialogue. In Jonathan Ginzburg and Enric Vallduví, editors, *Catalog '04. Proceedings of the Eighth Workshop on the Semantics and Pragmatics of Dialogue*, pages 93–101. Barcelona, 2004.

Hannes Rieser. Pointing and grasping in concert. In Manfred Stede, Christian Chiarcos, Michael Grabski, and Luuk Lagerwerf, editors, *Salience in discourse: multidisciplinary approaches to discourse*, pages 129–139. Nodus Publikationen, Münster, 2005.

Emmanuel A. Schegloff. The relevance of repair to syntax-for-conversation. In Talmy Givón, editor, *Syntax and semantics: Discourse and syntax*, volume 12, pages 261–286. Academic Press, New York, 1979.

David Schlangen and Alex Lascarides. The interpretation of non-sentential utterances in dialogue. In *Proceedings of the 4th SIGdial Workshop on Discourse and Dialogue*. Sapporo, Japan, 2003.

Kristina Skuplik. Satzkooperationen. Definition und empirische Untersuchung. Technical Report 1999/03, SFB 360, Bielefeld University, 1999.