Information State Revision and implicit Dialogue Moves in Instructional Texts

Annie Zaenen Staffan Larsson Agnes Sandor
David Traum

Distribution: PUBLIC

Task Oriented Instructional Dialogue
LE4-8314

Deliverable D3.3

February 2000

LE4-8314 TRINDI

Task Oriented Instructional Dialogue

Gothenburg University
Department of Linguistics

University of Edinburgh
Centre for Cognitive Science and Language Technology Group, Human Communication Research Centre

Universität des Saarlandes
Department of Computational Linguistics

SRI Cambridge

Xerox Research Centre Europe

For copies of reports, updates on project activities and other TRINDI-related information, contact:

The TRINDI Project Administrator
Department of Linguistics
Göteborg University
Box 200
S-405 30 Gothenburg, Sweden
trindi@ling.gu.se

Copies of reports and other material can also be accessed from the project's homepage, http://www.ling.gu.se/research/projects/trindi.

Responsibility for the authorship is divided as follows. Annie Zaenen was the overall editor. Staffan Larsson contributed the ImDis system described in Chapter 4.2. David Traum contributed to the introduction and the material on adapting PTT to instructional dialogue and text in Section 4.1. Agnes Sandor was a co-author with Zaenen of much of the remaining material. The authors also wish to thank Anette Frank and Massimo Poesio for pointers, discussions, and suggestions. The authors solely are responsible for errors and omissions.

Contents

Chapter 1

Introduction: Purpose of the Deliverable

Discourse analysis and dialogue modeling come from different traditions. The one takes as its object of study mainly written text whereas the other looks at spoken interaction among two or more participants. In a world where content and its expression are inseparable, this sharp distinction can be attacked on theoretical grounds: most texts are written with a reader in mind[1] and they take the shape they have because the writer imagines a certain reader. In the electronic world different types of relations between content and expression are possible and it is useful from a practical point of view to think about content and its expression as potentially separate. One would like to be able to present the same information in a different form depending on the needs of the intended addressee and the transmission channel. One of the types of variation in presentation would be to construct monologues and dialogues from a same basis.

To investigate the problems that are inherent in the realization of this aim one can start from different perspectives. We can ignore the traditional ways of presenting content and develop new methods to generate documents of all types starting from an abstract representation of the content that is the object of the communication. Instead of taking this route, we approach the problem by analyzing what exist and by investigating what it would take to transform material that is presented as text into an interactive presentation. We have chosen that route because at this moment abstract representations of content are in most cases not available whereas texts currently are. We limit our attention to a specific genre of text: instruction manuals because it is not possible to tackle this problem in its entirety (it would include an investigation of how novels are turned into operas and

[1] This can of course be contested and certain forms of written text can be conceived as private. Even then one could claim that they address themselves to the writer at a later moment in time. In any case, the type of material under consideration here is definitely written with a reader/user in mind.

the like) and instructional material is the kind of material for which such a transformation might be useful even if the interactivity achieved is limited. At the end of our investigation, we have to conclude that three types of information are necessary to construct (the textual parts of) instructional material: knowledge about the task structure, knowledge about the interaction possibilities of instructors and learners and linguistic knowledge. Once a text is constructed it embodies implicit knowledge about the task and the interaction possibilities but it is not likely that this explicit knowledge can be made explicit on the basis of the text alone. However, some textual clues can be used to guide the transformation and some textual structure can be kept constant.

We restrict our attention here to one kind of instructional material: user and repair manuals, with examples taken out of user manuals because they tend to be simpler. There are, of course, other kinds of situations in which instructions are given. Much of the work in the TRINDI project is related to *transaction* tasks, in which a 'system' provides some service for a 'user', often requiring more specific information from the user about some aspects of the desired service. The type under consideration here involves one party with information to issue *instructions* to another who has the ability to act, to complete a task. In the case of the user manuals, the instructions are subordinated to task performance, and questions and explanations about the instructions are subordinated to aspects of the particular situation, rather than a more general understanding of the class of problems for which the particular situation is one instance. In repair manuals the purpose is more ambiguous, the successful achievement of a particular action is important, but the *training* of one participant in tasks of this type, with the objective of the trainee subsequently being able to perform such tasks on her own is also important. In this type of manual it is important to impart to the user a model of the task as well as instructing her to execute the task.

These two types of instruction manuals are similar in that both are concerned with instructions, yet different in the context for the instructions as well as the overall goals. In the former case, the instructor can ignore possible situations of application that do not conform to the current case; if task performance is the objective, then the trainee knowing about these possible but not actual contingencies does not lead directly to enhanced task satisfaction, and thus are inefficient given the purpose of simply solving the task. On the other hand, if the purpose is to allow the trainee to act independently, it can be important to describe not just what is needed for task completion in the current situation, but also what is needed for similar situations, as well as how to tell which actions are applicable in which situations. The repair manuals are not pure examples of *training* material, however, and we ignore the difference between the two subgenres here. The different types of manuals point to different user needs which is another factor that can influence the differences in presentation of the same material. In this deliverable we only look at channel differences not at differences in users needs.

We start with a chapter that first sketches the hierarchical view on discourse structure and

illustrates it in more detail with the description of one theory, the LDM. In the second part of the chapter we relate this view on discourse structure to the the theory of information states and updates through the view of discourse as question and answering sequences as developed by Van Kuppevelt and compare this to some notions used in the TRINDI view of dialogue.

In Chapter 3, we analyze briefly the type of textual material that is the object of this deliverable, instructional manuals, and point out some of the underlying assumptions that are specific to the genre.

Chapter 4 is heart of the deliverable: we first argue that question-answering sequences are not necessarily dialogues and investigate what is needed to get to a minimal dialogue situation.

Chapter 2

Hierarchical Discourse Structure and Information State updates.

Work on discourse structure seeks to characterize text units that are larger than a sentence. The enterprise aims at the discovery of subunits and at the characterization of the relations that the subunits have to each other. This characterization is needed to explicate what constitutes the difference between coherent and incoherent discourse. Traditionally discourse theory looks at relations between sentences. However, the relations between different clauses within a complete sentence can be viewed as very similar in nature and it is a rather natural move to expand the approach to a description of intra-sentential clausal relations[1].

It has become traditional to divide the approaches to discourse coherence into two schools: the *informational* and the *intentional*. According to the informational approach, coherence can be characterized in terms of semantic relations between the information conveyed by the successive units. The intentional approach holds that the intentions of the writer need to be recognized. This approach draws on work in planning where recently the need for the recognition of shared intentions between writer and reader has been stressed, bringing this strain of work on discourse closer to the concerns evident in work on dialogue. After a very brief mention of some of the discourse theories currently on the market, we start here with an informational linguistically based schema and introduce more intentional notions in the subsequent discussion.

[1]Some frameworks exploit the similarity between intra- and interclausal structures explicitly, see for instance [Webber and Joshi(1998)]

2.1 Discourse Hierarchies

In various papers Hobbs e.g. [Hobbs(1993)] develops a theory of discourse relations that in its broad categories goes back to Hume's types of connections among ideas. Kehler [Kehler(1995)] adapts the proposal in distinguishing *resemblance* relations (e.g. parallel, contrast, exemplification, generalization and elaboration), *cause-effect* relations (e.g. result, explanation, denial of prevention, violated expectation) and narration or *contiguity* relations. While the work of Hobbs is seminal in the area, some other approaches have recently attracted more attention, mainly Rhetorical Structure Theory and Discourse Representation Theory.

2.1.1 Rhetorical Structure Theory

Mann and Thompson, e.g.[Mann and Thompson(1988)] propose a hierarchical organization of text spans. Each span is either a *nucleus* or a *satellite* of one of a set of discourse relations. The relations hold either between two or more nuclei or between one nucleus and one satellite. The number and nature of the discourse relations are a matter of lively debate among the proponents of RST. The relations recognized by RST are both informational and intentional. Examples of intentional relations are JUSTIFY, CONCESSION, EVIDENCE. Examples of informational relations are ELABORATION, SUMMARY, CONDITION. An extreme view of which informational relations need to be considered part of RST will of course lead to the recognition of all semantic relations, e.g. temporal, causal relations as part of RST. An attempt to come up with a practically useful set of relations can be found in the work of Knott and Dale[Knott and Dale(1993)].

Within RST there are proponents of parsimonious and proliferate approaches to the question of how many and which type of relations should be recognized. Hovy, e.g.[Hovy(1991)] tries to reconcile the two views by allowing a hierarchical organization of the relations themselves, e.g. a TIME relation is part of a CIRCUMSTANCE relation. This, however, does not solve the problem of the relation between informational and intentional relations: the two same text units can be related by both an intentional and an informational relation. Moore and Pollack [Moore and Pollack(1992)] who discuss this problem, give as an example "George Bush supports big business: he is sure to veto House Bill 1711", where the relation between the two clauses can be seen as informational (CAUSE) or intentional (EVIDENCE). Moreover, as shown in More and Paris [Moore and Paris(1993)], informational relations are in a many-to-many relation to intentional ones. This state of affairs undermines the attempts of RST to structure texts in one hierarchy.

2.1.2 Discourse Representation Theory

DRT is squarely on the informational side of the spectrum of discourse theories. It was first conceived to model intersentential anaphora and temporal phenomena but is being developed further to model various other semantic and pragmatic phenomena. The work of Lascarides and Asher e.g. [Lascarides and Asher(1991)], [Asher and Lascarides(1993)], explicitly incorporates rhetorical relations in DRT. Their work on Segmented DRT, (SDRT) has concentrated mainly on the analysis of temporal relations and phenomena of modal subordination. Its contribution is an adaptation and an explicit formalization of some of the proposals of RST combined with DRT.

2.1.3 The Linguistic Discourse Model

The Linguistic Discourse Model (LDM) developed i.a. in [Scha and Polanyi(1988)], [Prust *et al.*(1994)], [van den Berg(1996)] is also an informational approach. It stresses the incremental way in which the discourse structure is built. This focus on the incremental build-up of the discourse is an important feature in facilitating the understanding of discourse as a limit case of dialogue. It is one of the reasons we have chosen to discuss the LDM in a bit more detail.

LDM recognizes two main types of discourse relations: *coordination* and *subordination*. Coordination can be further subdivided in narratives, lists, and the like; subordination into elaboration, etc. There is at this point no stable list of relations and no completely stable terminology but the theory attempts to keep the number of basic relations small. When it is felt to be useful, a basic relation can be annotated with a more precise label. A category 'binary' is used for cases that do not fall comfortably in either of the two big categories. One hopes, of course, either to do away with this in the near future or to define it better.

The relations build up complex DCU (discourse constituent units) out of basic ones. A basic unit is typically a clause (a proposition). The system assigns some sentence internal structure as the same relations often hold between sentences and clauses. One of the most important differences between intersentential and intrasentential relations is that the intrasentential relations are most of the time clearly signaled by embedding complementisers or the like. The discourse function of these elements is that of *operators*. Operators combined with the clause they operate on are represented as modification relations. This is a kind of subordinating relation but it is represented in a special way to stay close the sentence syntax.

More concretely,

- In simple clauses with phrasal complements
  - VPs with independent objects are coordinated BDUs
  - VP's with the same objects are treated as one DCU
- In complex clauses
  - relative clauses form a separate BCU
  - infinitivals that express a separate activity form a separate BCU
  - embedded tensed clauses are either elaborations or modifications

As said above discourse modification corresponds to the syntactic notion of adverbial clause. To this category belong, for instance, temporal clauses, because clauses, etc. A special case of modification are preposed adverbial phrases and verbless adjectival clauses which have a special status in LDM as they are felt to be stage-setting elements that differ in function from syntactically similar elements occurring elsewhere in the sentence. Modifications are built up out of DCU's combined with an operator, typically a 'discourse marker'. Examples of operators are: conjunctions, coordinators, relative and interrogative pronouns ("who, when"), "now, ok", interjections, if/then, phrasal modals (it is possible that...). Typically operators combine with BDU's to form *modifiers*, which then combine with another DCU , basic or not, to form a new one. We distinguish, left and right operators.

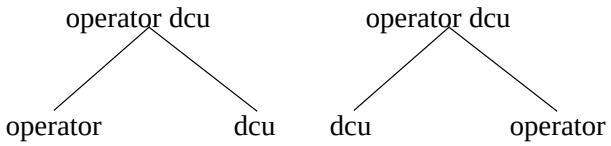


Figure 2.1: Operator

Modifiers also come in a left and right variety.

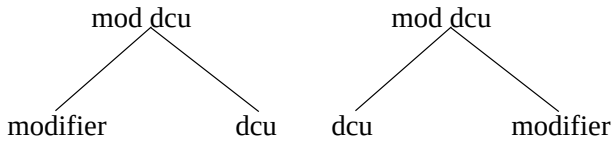


Figure 2.2: Modifier

Note, however, that left operators do not necessarily correspond to left modifiers: right operators are rare in English but modifiers can occur to the left or the right of their 'main clauses'. The basic DCU's can also be combined directly via the discourse relations. In a language like English, the complex DCU's within sentences are typically built up with operators and enter in the modification relation. Across sentences, however, the discourse

relation is often not overtly expressed and has to be inferred. There is no operator and no modifier.

Consider for instance the following two texts:

- John fell. Bill pushed him.
- John fell because Bill pushed him.

In the first case we have a subordination relation between the two clauses, with 'John fell' as the 'main' clause, in the second case we have a modification relation. The distinction is made because LDM wants to stay close to the surface syntax and hence wants to encode the difference in sentence structure between the two cases.

The analysis of the two texts above would be (a) and (b) respectively:

(a)

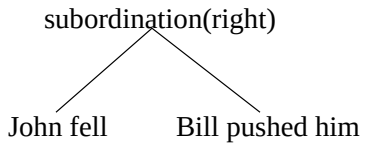


Figure 2.3: Subordination

(b)

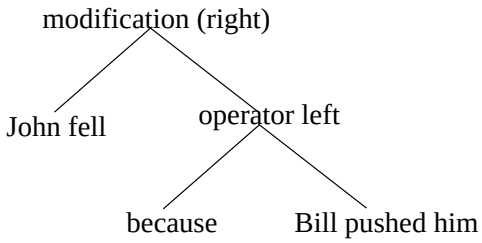


Figure 2.4: Modification

In what follows we will ignore the operator-DCU structure. Notice that the modifier-DCU structure will be represented as immediately above: the discourse relation between two daughters is used as the name of the mother node. The << indicates that the modifying element is to the right of the modified element. Similarly >> will indicate that the modifying

element is to the left[2].

The incremental parsing algorithm insures that at any point in the process, only the right nodes in the discourse parse tree are open for expansion. This is diagrammed below, from [Prust *et al.*(1994)][3].

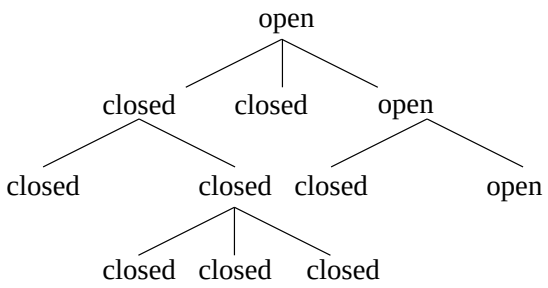


Figure 2.5: Closed and open constituents

Trees are built up via *construction* and *extension* rules[4]. When a construction rule is used the node at which the construction takes place is replaced with the left-hand constituent of the rule. The original node becomes the leftmost daughter and the rest of the right hand constituents become the other daughters of the node corresponding the left hand constituent. For example if we have a rule A $\rightarrow$ N B, the construction can be diagrammed as follows.

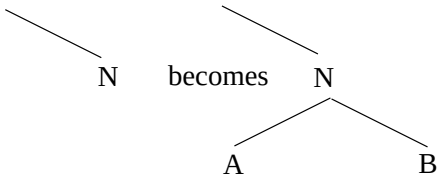


Figure 2.6: Construction

The other possible way to expand a node is through extension. Extension comes down to adding a new daughter to the dcu. E.g. n $\rightarrow$ n nj creates the second tree from the first one in the diagram below.

These rules give the way to put together discourse units. Together with the constituent tree a *semantic representation* is built up for the unit. To calculate the semantics a form of dynamic logic is used and to represent it a notation very similar to DRT is used.

[2]The labeling does not correspond directly to that of the syntax tree that most theories evolving from generative PSG's would assume. Most PS based approaches would label the 'because' node as the operator and the mother node would indicate the relation between that node and it sister, rather than the relation between its two daughters.

[3]A similar constraint is proposed in [Webber and Joshi(1998)]

[4]In fact one can reduce the tree construction algorithm to construction only, see [van den Berg(1996)]

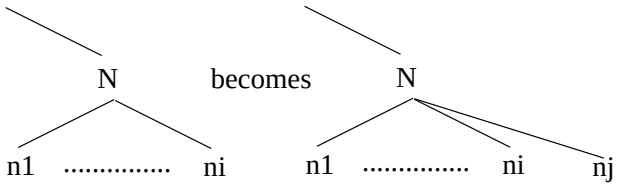


Figure 2.7: Extension

A worked-out version of the semantic interpretation proposed in LDM and SDRT should be very similar in coverage but at this moment, the two approaches have concentrated on different aspects and are hence not comparable.

Examples

The following two examples are extracted from chapter 6 of a user manual for the HomeCentre, a low end Xerox MultiFunctional Device.

Example 1: begin of the section "Replacing the print head".

- Replacing the print head
- You need to remove the print head to:
  - Replace it with a new one
  - Protect it when moving the HomeCentre to another location
  - Switch from a black-only cartridge to color ink cartridges and vice versa.
  - Store the print head in a safe place when you aren't using it or moving it.
- .....

As we said above, mother node labels encode the relation between daughter nodes. This type of encoding loses information about the direction of the relation if it is not given explicitly. For this reason we add 'left' or 'right' to the subordination relation. Left means that the subordinated material is to the left of the 'main' DCU, whereas 'right' indicates that it is to the right. Figure 2.8 diagrams the analysis of this short passage.

The information encoded in this tree is the following: the passage is composed of two main units that are in a binary relation as title and body. The body part is composed of a main part and a subordinated one (an aside). The main part of the body is the clause 'you need to remove the print head which is modified by an enumeration of four elements: "replace it...", "protect it...", "switch from...". The element "protect it" is modified by "when moving...". The aside, "store the print head..." is modified by "when you aren't using...."

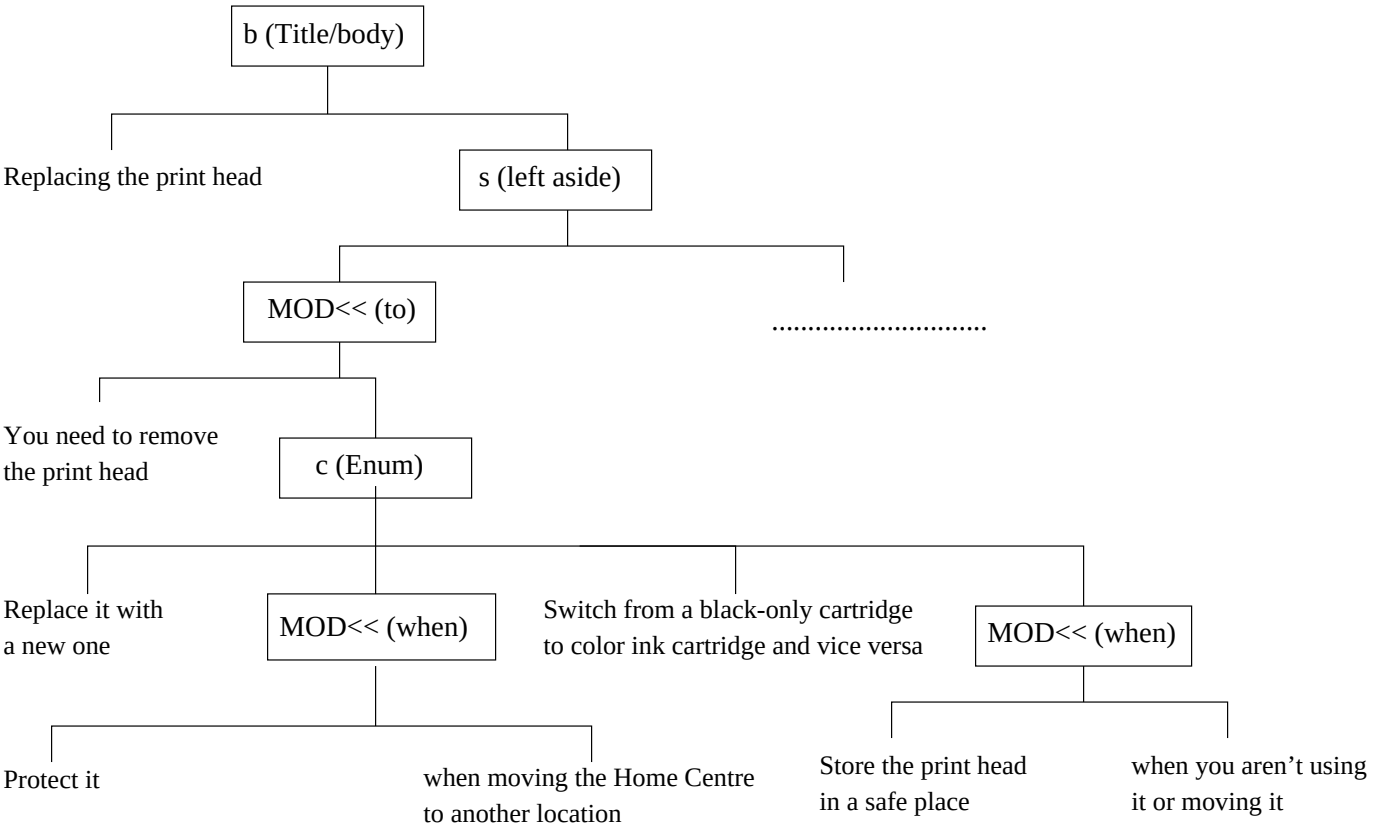


Figure 2.8: Replacing the printhead

Some lowly embedded distinctions have not been marked explicitly, e.g. the clause "when you aren't using it or moving it' should be split further into two coordinating parts. This is not done because this degree of detail would make the structure more difficult to read.

Example 2: Reinstalling the print head.

- Reinstalling the print head
- Caution: Make sure that the green carriage lock lever is STILL moved all the way forward before you reinstall the print head.
- 1. Line up the hole in the print head with the green post on the printer carriage.
- Lower the print head down gently into position. item 2. Gently push the green cartridge lock lever up until it snaps into place.
- This secures the print head.
- 3. Close the top cover and re-attach the scanner.
- 4. Press and release the yellow LED button.
- The printer will prepare the cartridge for printing.

- Note: If the carriage does not move from the center position after you press the cartridge change button, remove and reinstall the print head.

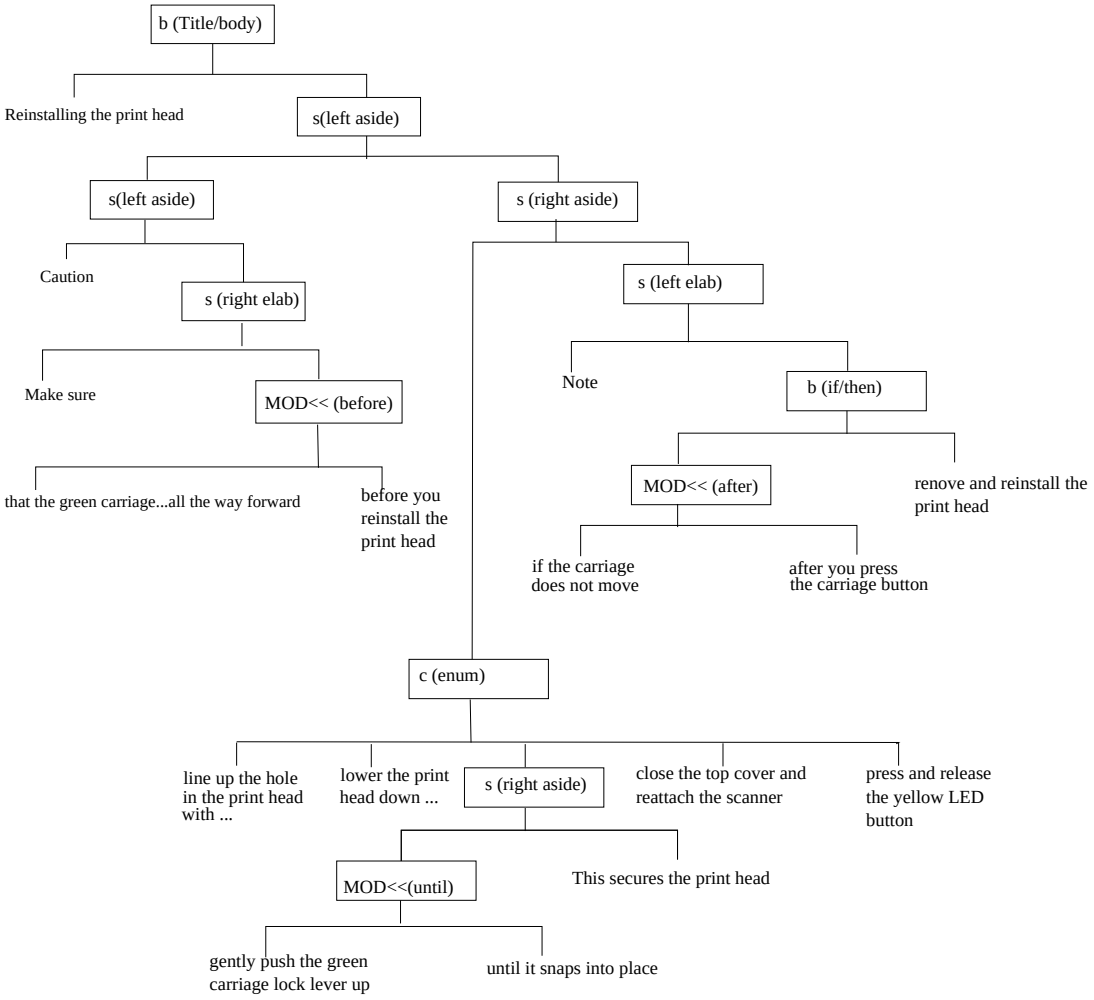


Figure 2.9: Reinstalling the print head

See Figure 2.9 for an analysis of this passage.

The hierarchical ways of analyzing discourse conceive of a text as a communicative unit: there is an initial goal that gets realized though the succession of sentences. There is of course an implicit representation of an interlocutor: the text assumes that the reader knows a certain number of things, has the capability to reason, and has a certain intention reading the text. These aspects are not represented explicitly because they don't change during the processing of the text. Of course one thing is not constant: what the reader knows about the content of the text itself. The clearest indication of this is that through the progression of the discourse different referents become available as part of the assumed shared understanding between the writer and the reader.

2.2 Information States and their Updates

2.2.1 Common Ground

The various views on discourse structure all take for granted that a coherence assumption guides our understanding of a text. And that the various discourse relations spell out various subcases of coherence. What is meant by coherence is, however, not always stated very explicitly. A different tradition (but one that has influenced DRT) has spelled out one important view on coherence: new information updates the information state of the discourse participants, so for discourse to be coherent these updates have to be possible within a set of assumptions about what discourse game the participants are involved in. Over the years there has been a shift in opinion in part of the community about how explicitly the information state of the various discourse participants needs to be represented to model updates adequately.

In the 70ties work in pragmatics led to the proposal that discourse has to be interpreted against a common ground of shared presuppositions. The aim of communication is supposed to be to come to agreement about how things are in the world. The common ground can be thought of as a set of propositions that the interlocutors hold to be true. At each moment of the discourse this common ground defines a context set. In Stalnaker's [Stalnaker(1979)] view, one of the most influential versions of this approach, 'each assertion change this context set by adding the content of what is asserted to what is presupposed' (p.323). So, Stalnaker's view of updates is rather simple: assertions can be accepted or rejected and if they are accepted the context set is updated with their content.

The general idea of information update has found its way into linguistic theories in various ways. One development is not concerned with multiple participants but looks at the way text or monologue is updated. This view is found in the work of Heim [Heim(1982)], most of DRT and dynamic semantics in general.

These views have been refined in different directions. Carlson [Carlson(1983)] who following Wittgenstein and Hintikka, analyses discourse as a game, distinguishes two types of moves that can be made, *set-up* moves or questions and *pay-off* moves or assertions. The contributions of the speakers are constrained by two types of rules, *conventional* ones and *conversational* ones. The conventional rules include the linguistic rules of various types whereas the conversational ones follow from assumptions about rational behavior (e.g. Gricean maxims).

This work has influenced a strain of the work in discourse analysis which sees discourse, even monologues, as analyzable in terms of questions and answers.

2.2.2 Discourse as Questions-Answers Sequences

Most recently, this approach has been developed further in several publications of Jan van Kuppevelt e.g. [Kuppevelt(1995b)]. Under this proposal, a passage like the one given on p.18 would come out as follows:

- How does one reinstall the print head?
- What should one make sure of before reinstalling the print head?
- Make sure that the green carriage lock lever is STILL moved all the way forward before you reinstall the print head.
- What should one do first to reinstall the print head?
- Line up the hole in the print head with the green post on the printer carriage.
- And then?
- Lower the print head down gently into position.
- And then?
- Gently push the green cartridge lock lever up until it snaps into place.
- Why should one do this?
- This secures the print head.
- And then?
- Close the top cover and re-attach the scanner.
- And then?
- Press and release the yellow LED button.
- What will happen now?
- The printer will prepare the cartridge for printing.
- What if the carriage does not move from the center position after one presses the cartridge change button?
- If the carriage does not move from the center position after you press the cartridge change button, remove and reinstall the print head.

Another example was given in [Zaenen and Larsson(1999)]. We repeat here the question/answer part:

- How is paper managed?
- What should one do to insure the right paper management?
- When should paper rolls be changed?
- When a paper roll is finished, an 'End of Paper' alarm will be generated and the print engine will stop automatically. A new roll should be mounted.
- In some cases you might want to change rolls before the mounted roll is finished

- In which cases might we want to change the rolls before the mounted roll is finished?
- Change of paper grade
- As a premature roll change results in some extra paper loss and non-productive time, it is advisable to group the printing of jobs as much as possible on their paper needs

The question-answer technique as such gives a less direct picture of the hierarchical structure of text. It is not immediately clear that some questions are subquestions with respect to previous ones and that some answers are at the same time an answer to the immediately preceding question and the completion of the answer to a higher question. To remedy this, Van Kuppevelt [Kuppevelt(1995b)], [Kuppevelt(1995a)] distinguishes *topic-constituting questions, subtopic-constituting subquestions* and *topic digressions*[5]. Subquestions can be further subdivided in *qualitative* elaborations and *quantitative* elaborations. Quantitative elaborations extend the answer given to the previous question, whereas qualitative elaborations provide support for what was said before. They correspond roughly to the elaborations and explanations of LDM. The topic digressions correspond to asides in LDM: they give information that is not part of the main topic but are raised by the previous answer.

Van Kuppevelt introduces a *dynamic principle of topic termination*(DPTT), which has a similar effect to that of the right frontier principle in LDM. With these additions to the pure question-answer technique, his proposal has a similar hierarchical structure as that proposed in LDM. He also adds a *Principle of Recency* (PR) stating that 'in a well-formed discourse every explicit or implicit subquestion is asked as the result of an answer A, which is the most recent unsatisfactory answer to a preceding question'. This gives a special status to context that is string local, if we think in representational terms as string locality corresponds to temporal adjacency in the model.

The question-answer technique has the advantage that is identifies more explicitly what the contribution of each unit is than is the case in the hierarchical structured approaches sketched in the previous section. As we said above, the relation between sentences is often implicit. Seeing the sentence as an answer to a question makes explicit what the relation is, for instance in the first passage above the four points are clearly marked as a temporal sequence by the 'and then?' questions. The sentence 'this secures the print head' is clearly marked as an explanatory remark. The last sentence comes out as a topic digression. The first sentence is, however, a bit difficult to classify. It is also a topic digression of a kind but it cannot be linked to a previous answer that would bring it up. We will discuss this further.

Apart from making explicit some of the discourse relations among the various textual units, the question-answer transformation also clarifies the topic-focus structure. This too contributes to showing more clearly what the contribution of each unit is. In this way the

[5]He has also a category of non-associated topic shifts but these are not relevant to our discussion here.

representation helps us to identify the information state update intended by each textual unit.

But of course, the exercise remains intuitive as long as there is no algorithm to produce the questions. The only discourse representation theory for which there exists an explicit algorithm is a version of RST implemented by Marcu [Marcu(1997)]. Current work by Van Kuppevelt addresses this problem from a theoretical side but not from the computational one.

2.2.3 Information States and Dialogue

Along side with the analysis of monologues as a succession of questions and answers, interest in common ground issues has also stimulated work on dialogue within the linguistic community. Recent work, e.g. by Ginzburg [Ginzburg(1996)] and Cooper [Cooper and Larsson(1999)], starts from Stalnaker's view on common ground and some notions of situation theory but elaborates these ideas in various ways.[6] Important contributions of this recent work are the notions of *questions under discussion* and of the *latest move*. In the Stalnakerian approach, the context is an unordered set of propositions, in the recent view it is a (partially) ordered set of questions under discussion QUD, and information about the LATEST MOVE. In the view of Cooper, the QUD corresponds broadly to the DPTT of van Kuppevelt, which as we saw is similar to the right frontier constraint of LDM. In Ginzburg's view, it is not specified how the partial order on the QUD comes about. The LATEST MOVE addresses a problem that is also addressed by van Kuppevelt's PR: even when there is more than one question under discussion, the more local structure tends to be more immediately relevant for the next piece of discourse than earlier ones even if they are still under discussion.

The main important difference, however, between the proposals of Ginzburg and Cooper and the ones discussed above, is that they model the different *discourse participants* in the discourse and their *information states* explicitly. They have this in common with another strand of research coming from AI. This approach, the BDI model (where B stands for *Beliefs*, D for *Desires*, and I for *Intentions*) for dialogue analysis, was mainly developed by in papers such as [Cohen and Perrault(1979), Allen and Perrault(1980), Pollack(1990), Cohen and Levesque(1990), Sadek(1991)]. Under this model, speech acts, which we will assimilate here to dialogue moves, make the desire for information of the participants mutually known. For instance the assertion of P has as an intended effect to establish P as a shared belief. The BDI model adds something that is not stressed in the informational approaches, we have been looking at in the previous section, namely the notion of *goals* and intentions. The assumption is that goals too are hierarchically ordered. As in the previous proposals this leads to a hierarchical structure: questions can be subquestions to larger

[6] It is not clear that Stalnaker would agree with these elaborations.

ones and answers can fulfill a small goal but at the same time also a higher one. Another approach that stresses the importance of intentions, is found in the work of Grosz and Sidner and their followers [Grosz and Sidner(1986)]. Here we get back to the problem that was raised in RST but now the structure of goals and intentions is linked to the discourse participants instead of being a general property of the text per se. This seems to be a more insightful way to state the problem.

The introduction of discourse participants and their information states as first class citizens also allows us to be more precise about notions that in a theory like that of van Kuppevelt tend to be vague such as the notion of 'unsatisfactory' answer. As van Kuppevelt discusses, answers can be unsatisfactory for various reasons: because they are incomplete or because they have 'not yet been accepted as a true answer by the addressee' ([Kuppevelt(1995b)], p. 817). This idea is given a more explicit modeling in the notion of *grounding*. Before the content of a dialogue move becomes part of the common ground, it needs to be accepted by the various participants in the discourse. Under this view it is necessary to make a distinction between the private context of each participant and the shared context. Propositional content gets form the one to the other after it has been grounded for all participants. Grounding has been studied by Clark [Clark and Schaefer(1989)] and in a computational context by Traum [Traum(1994)]. Grounding implies some kind of acknowledgment that the content is understood. It can take different forms, be *explicit* or *implicit*. In dialogues in which the participants do not see each other, grounding tends to involve a linguistic act through a new utterance that would not make sense if the previous one had not be accepted. However, grounding does not need to be linguistic: in the context of instructions, the execution of an instruction can be seen as implying that the instruction has been understood (or at least that the participants share the belief that it has been understood.)

In the next chapter, we turn to the main topic of this report: how these views can illuminate the structure of instructional text and the transformations necessary to present its content in an interactive way.

Chapter 3

Are all question/answering sequences dialogue?

Chapter 2 gave an overview of some view on discourse structure getting us from a purely informational view to a view that was potentially enriched with information about updates of information states and of information states that contain the goals and intentions of the writer and reader. In this chapter we describe a first version that is minimally different from the view developed in [Zaenen and Larsson(1999)]. We show how it can be modeled within the TRINDI approach and how it relates to assumptions about instruction manuals as a genre. We question, however, its status as bona fide dialogue.

3.1 From Questions and Answers to Information States

Let's start with the example discussed in [Zaenen and Larsson(1999)]. There we illustrated how a piece of discursive text could be turned into a sequence of questions and answers and gave the following example:

- How is paper managed?
- What should be done to insure the right paper management?
- When should paper rolls be changed?
- When a paper roll is finished, an 'End of Paper' alarm will be generated and the print engine will stop automatically. A new roll should be mounted.
- In some cases you might want to change rolls before the mounted roll is finished
- In which cases might we want to change the rolls before the mounted roll is finished?

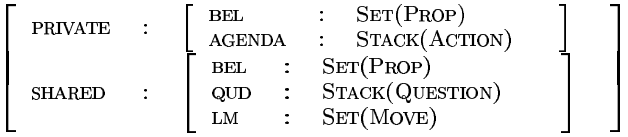


Figure 3.1: Cooper/Larsson Information state type

- Change of paper grade.
- As a premature roll change results in some extra paper loss and non-productive time, it is advisable to group the printing of jobs as much as possible on their paper needs

As said in[Zaenen and Larsson(1999)], in this monological text there is no feedback and hence no overt grounding process. As we will discuss further in the next section, however, there is an assumption that is part of the genre 'manual' that the reader has understood what she has read before. In other words, the manual assumes that the text is immediately grounded when it is read. We will thus maintain the distinction between private and shared information. The strategy used is the one that is described as 'optimism' in [Cooper and Larsson(1999)]. According to this strategy, participants assume that their contributions have been understood and entered as parts of the participants' common beliefs or QUD's as soon as proffered. When discussing discourse structure we saw that texts are not written sentence by sentence but that there is a progression from topic to subtopic. This idea is captured in the LDM analysis in the form of a discourse tree. In the TRINDI formalism we can capture this hierarchical structure in the form of a plan.

The set of moves that need to be postulated at this point is rather small and is a simplified version of the optimist strategy. Under the optimism strategy, the flow of information between the agenda and the QUD is as follows (see [Cooper and Larsson(1999)].), where DP is a dialogue participant, q is a question, and the information state looks something like that in Figure 3.1.

- When DP raises(q) on the agenda, DP expresses q and pushes it on the QUD;
- When q is on the QUD , the DP whose turn it is pushes respond on her agenda
- When respond is on the top of a DP's agenda, DP produces an appropriate response to q and pops the agenda and the QUD and adds the response to the common beliefs.
- When a DP notices that question q as been responded to with p, this DP is integrated into the beliefs of the DP

The one thing that needs to be added to this overall strategy is the consultation of a discourse *plan*. In the case at hand, it is the plan that contains moves such as INFORM and

RAISE, as the answer to one question might raise new ones, which in turn might trigger new plans. Finally after a satisfying answer is found, a question or a set of questions are removed from the QUD and the agenda.

In the record based implementation this leads to the sequence of dialogue moves proposed in [Zaenen and Larsson(1999)].

As shown in this example, the plan imposes a hierarchy on the questions that are asked and addressed. But where does this plan come from? It is not enough to have a database of facts about a device, one also needs an idea about how this information has to be presented. In the case at hand this information is embodied in the existing service manual.

3.2 Instruction Manuals as a Genre and Manuals as a Discourse Participant

Even a cursory inspection of instruction manuals of various types (end user manuals, maintenance manuals, repair manuals) shows that they have a rather fixed structure. Like other documents of a certain size they have a table of contents and several chapters. The first chapter tends to be a parts list, the next one a general overview and the remainder contain sections with specific instructions. The grouping of these instructions tends to determined by the functions of the device that they describe (user manuals), or by its parts (maintenance and repair manuals). The sections contain specific instructions, namely information about how to execute a specific subtask that can be thought of as achieving a goal that can be achieved through a set of actions that are adjacent in time. They are often divided in two parts, a description giving the rational for the action as well as a general overview of what is required and the stepwise procedure itself. In what follows we will call such a section containing an optional description and a procedure an *instructional unit*. We will also assume that instructional units are what corresponds the closest to 'dialogues'. They are the object of our analysis in this deliverable. We will refer to the optional first part as the *description*, to the procedural part as a whole as the *procedure* and to each item in a procedure as an *instruction*.

We will assume that the parts list and the introductory chapter are considered to be background at the moment these instructional units are consulted. The rest of the manual is, however, not considered as background: when reading a chapter 5 about 'paper paths' in a Xerox repair manual, one is not assumed to know everything about humidity and condensation even if these are treated in an earlier chapter.

Apart from having a structure, manuals[1] also make some assumptions about their users and about the way they will be consulted. These assumptions can be thought of as motivating the conversational rules of the genre. They are rather commonsensical. They are, however, different from what is in general assumed in the literature: most of the literature assumes that the aim of the discourse game is to arrive at the truth. Here the aim is, to paraphrase Marx, not to understand the device but to change it. Conversational maxims should be adapted to this task. Maxims like 'don't say what you know to be untrue', have to be replaced by 'don't give an instruction that you know cannot be executed'. In a context where actions take place, it is also important to prevent accidents. In such an environment, the maxim of quantity needs to be specialized to 'always give enough information to prevent the user from doing harm to herself or to the machine and anticipate frequently occurring failures'. Another rule that is assumed is that the instructions are normally given in the order that corresponds to the temporal order of their execution. This is of course a rule that has wider applications than just in instructional texts.

Whereas the manual has a rather precise idea of the aim for which it will be used, it has only very general knowledge of its users and no way to adjust it through interaction with them. But it is not without a view of its user: an interesting feature of instructions is that few of them are given without any rational or indication of their purpose and often, overt indications of what would happen if the instruction were not followed are added[2]. These explanations seem to be geared to human users: if the repair were done by a (simple) robot, they would be unnecessary. They seem to follow from a constraint on human interaction that one should not give commands for which the rational is not clear.

Of course a manual has also no good idea about the state the device is in at the moment the operation will take place.[3]

These characteristics of a manual, its aims and its awareness of the limits of information that it has about the situation in which it will be consulted, structure the plan that one can associate with instructional units. The most straightforward influence is that of the ordering constraint. The use of NOTE and CAUTION follow from the requirement not to hurt the user or destroy the device in a context where the manual cannot check directly whether the user is paying attention or knows about the danger. We analyze them as

[1] As we want to come to an understanding of instructional text as dialogue, we will anthropomorphise the 'manual' here as a discourse participant.

[2] As said in the introduction, different types of manuals differ in how far they go in this. Simple user manuals tend to have minimal information of this type.

[3] It would be useful to have a worked out set of maxims that guide cooperations when the cooperation is not geared to the discovery of the truth but to get to achieve more mundane tasks. This would allow us to make a better distinction between what is common to the task and what is imposed by the form of interaction. Given that this strain of literature has mainly be developed by philosophers, there is, however, very little. We are only aware of some work by Jens Allwood, e.g. [Allwood(1997)], who doesn't exactly address this issue. Straight planning literature addresses these issues extensively but in a too detailed form to be useful for the exploration at hand.

'asides' in the discourse analysis of our text but they have a rather specific function in a manual.

The very fact that it are assumptions about the manual genre that help us understand how a question answering sequence corresponding to a section of it is structured, indicates that we don't have to do with 'real' dialogue[4]. In the next chapter we will look at the differences between instruction manuals and real dialogues in more detail.

[4]We apologize for using the term 'dialogue' in a rather confusing way in this text. Our problem is that we want to treat the instruction manual as much as possible as a dialogue system in order to bring out both the similarities and the differences. But as there are differences, we need another notion of dialogue that refers to what are normally called dialogues

Chapter 4

From Question and Answer Sequences to Dialogue

The genre constraints that we briefly described in the previous chapter are quite different from the ones that are found in most real dialogue. Let's illustrate this and look at one of the examples given in Chapter 2, repeated here in part for convenience.

- How does one reinstall the print head?
- What should one make sure of before reinstalling the print head?
- Make sure that the green carriage lock lever is STILL moved all the way forward before you reinstall the print head.
- What should one do first to reinstall the print head?
- Line up the hole in the print head with the green post on the printer carriage.
- And then?
- .....
- And then?
- Press and release the yellow LED button.
- What will happen now?
- The printer will prepare the cartridge for printing.
- What if the carriage does not move from the center position after one presses the cartridge change button?
- If the carriage does not move from the center position after you press the cartridge change button, remove and reinstall the print head.

It is very unlikely that in a real dialogue, a user would follow up the first question with the second. As we saw, in Van Kuppevelt's analysis this should be a topic digression but topic digressions should be anchored to immediately previous answers, whereas here there is no answer available. In an interactive situation, the system might ask: 'Which state is the green carriage lock lever in?' and then, depending on the user's answer, either instruct the user to put the lever in the right position or to go to the first instruction. Similarly at the end of this passage, the manual spells out a case in which the procedure hasn't had the expected result. In an interactive system, this would only be necessary if there had been an explicit question from the user[1].

But how do we model these observations? The TRINDI approach points to an answer: the differences lie in what is shared in the information state of the DP's. In a non-interactive presentation, the manual can only execute its plan while making a number of basic unchangeable assumptions about the user. In an interactive setting, the manual can ask questions and on the basis of the answer, its assumptions about the user will change and the plan can be adapted.

These observations highlight the importance of the TRINDI approach: it gives us a direct way to model the difference between a traditional manual and a full-fledged dialogue system. The difference lies in the INFORMATION STATES of the DISCOURSE PARTICIPANTS and, more precisely, in the difference in information that they have about each other's IS. What we need to model, then, is on the one hand a constant goal of the information exchange and on the other sets of IS's that vary according to the assumptions that are made about what the different DP's know about each other.

To do this formally we would first need to develop a TRINDI model for monologue. This would involve the extension of the various TRINDI proposals because now they include an assumption that communicative behavior is part of the dialogue situation, in which each party has the ability to take appropriately timed communicative turns, and e.g., in the Poesio-Traum theory of information state, has obligations and commitments relating to this participation, as well as a developing model of common ground. This assumption is obviously violated in a situation of producing off-line instructional text, such as a manual.

In the following sections we will push this further and show how we can use the assumption that instructional text is some kind of degenerate dialogue to clarify the similarities and the differences between the presentation of instructional material in both modes. First, we discuss the Poesio-Traum formalization of information states in dialogue and show how these might be extended to monologue and different types of restricted dialogue systems. In Section 4.2, we turn to a system that can produce both monologues and dialogues.

[1]We can even imagine an interaction mode in which the manual might have direct information about the state of the device and hence can skip these questions or warnings in its interaction with the user all together. We will not consider this case explicitly but it should be kept in mind that the discussion in the text is only illustrative and that many other forms of interaction can be envisioned.

4.1 A TRINDI Model for Monologues

As background, we first present a brief overview of the Poesio-Traum theory of information states. For more details, consult [Traum *et al.*(1999b)], chapter 5, or [Poesio and Traum(1997), Poesio and Traum(1998), Matheson *et al.*(2000)].

4.1.1 The PTT Dialogue Model

PTT partitions the information state of each participant into public, semi-public, and shared aspects. The public part, the assumed *common ground*, called `G` contains a dialogue history, containing the sequence of dialogue acts that have been performed, a set of obligations of the dialogue participants to perform actions, and a set of social commitments to propositions (SCP) of the dialogue participants.[2] The semi-public component consists of a stack of *ungrounded Discourse Units* `UDUs`, each DU consisting of the same type of information as G. Discourse Units represent bundles of coherent information that can be entered into the common ground together, through the performance of a single grounding act. Information is put into a DU that is part of UDUs, given an initial utterance contribution, and can be added to `G`, when an acknowledgment is produced by the other dialogue participant. The private information consists of the beliefs and intentions (and goals or desires) of the individual participants.

A rich set of dialogue moves are used to provide the pragmatic effects of utterances on the information state. The set of dialogue moves used[3] and their effects used is shown in Figure 4.1. Some of the important effects are as follows: acknowledgments will move the contents of the DU argument to G, thus introducing commitments and obligations to the common ground (and relieving obligations to perform actions that have been performed in the DUs). Directives introduce obligations to address the directive: an acceptance obliges the speaker to actually perform the directed action. Assertions introduce an SCP to their propositional content.

[2]Also included are sometimes *options*, actions that participants are allowed (but not obliged) to do, and *conditionals*, representing obligations and SCPs that will be established given the performance of particular dialogue acts.

[3]based on the DRI dialogue acts [Discourse Resource Initiative(1997)], though using the grounding acts from [Traum and Hinkelman(1992)]

act	ID:2, **accept**(DP,ID2)
effect	*accomplished via rule resolution*
act	ID:2, **ack**(DP,DU1)
effect	peRec(w.Gnd,w.pdu.tognd)
effect	remove(DU1,UDUS)
act	ID:2, **agree**(DP,ID2)
effect	push(SCP,**scp**(DP,**P**(ID2)))
act	ID:2, **answer**(DP,ID2,ID3)
effect	push(SCP,**ans**(DP,**Q**(ID2),**P**(ID2)))
act	ID:2, **assert**(DP,PROP)
effect	push(SCP,**scp**(DP,PROP))
effect	push(COND,**accept**(**o**(DP),ID)→ **scp**(**o**(DP),PROP))
act	ID:1, **assert**(DP,PROP)
effect	push(COND,**accept**(**o**(DP),ID)→ **scp**(**o**(DP),PROP))
act	ID:2, **check**(DP,PROP)
effect	push(OBL,**address**(**o**(DP),ID))
effect	push(COND,**agree**(**o**(DP),ID) → **scp**(DP,PROP))
act	ID:2, **direct**(DP,Act)
effect	push(OBL,**address**(**o**(DP),ID))
effect	push(COND,**accept**(**o**(DP),ID) → **obl**(**o**(DP),Act))
act	ID:2, **info_request**(DP,Q)
effect	push(OBL,**address**(**o**(DP),ID))

Figure 4.1: Dialogue acts and effects in PTT

4.1.2 From Dialogue to Monologue

There are several ways one could change dialogue models to account for this sort of situation. A first approach is to recognize the different communicative setting, extrapolating from low-bandwidth communication, such as mail, in which one has rather large and self-contained contributions rather than relying on rapid feedback to deal with potential grounding problems. Here, we would treat all communications as part of the same continuing discourse unit or move and we would posit a higher grounding criterion (since all explanatory information must be given in advance, rather than on demand), and different resources (e.g., the ability to read over the material more than once), to produce differently structured communications. Thus, we might not correlate a simple dialogue act such as *direct* with an action in the domain plan, and of course, we could not use an interactive dialogue act, such as a question/answer exchange, but we might instead provide rationale for each action or likely situations of applicability, as well as instructions of what to do when things go wrong. The plans may then resemble in some ways a tutorial dialogue, however, there will still be differences based on the lack of feedback ability.

A second possibility is to actually remove some of the basic dialogue features of the theory for such a setting. E.g., avoid the use of grounding acts, and just have each dialogue act have its effects directly on the common ground. This does have the advantage of a simpler theory which is, to some degree orthogonal to the dialogue case (without positing extra implicit actions), although it does make the theory somewhat different.

A third possibility is to treat the situation as one of dialogue, in which the addressee is making *implicit* rather than overt moves. Thus, after each suitably sized utterance contribution, the system should postulate acknowledgments, acceptances (which will lead to commitments and obligations on the part of the addressee), and the performance of obliged actions. An issue here is that of the granularity of dialogue acts and contents. Is a complex instruction one directive act, with a very complex action as its content, or a sequence of simpler action directives? In the dialogue case, one can finesse this question to some degree, by looking at the natural chunks in which the complex action is broken into by the interactive (and grounding) phenomena. However, in text, one does not have such short cuts.

From a theoretical point of view this last approach might be objectionable: clearly a text does not give an explicit indication of how to divide it up into moves. When one looks at instructional manuals in practice, however, one gets the impression that, at least for the procedures themselves, some natural units are assumed. They correspond to items on lists with sometimes notes or cautions attached to them. For the more explanatory parts of a manual, the situation is not so clear but the annotation exercise that we did in [Zaenen and Larsson(1999)] showed that whether we approached a text with a discourse structure aim in mind or with the idea to find a plan for an odd kind of dialogue we can

up with the same structure.

4.1.3 Moves, Information States, Dialogue and Task Plans

Whatever the mode of interaction assumed or chosen, the main goal of instructional material remains the same[4]: get the user to want a device to be in a certain state. We can assume that the manual has the same overall goal and that both also share the assumption that a good way to reach that state is for the manual to instruct the user to execute a given procedure (a sequence of instructions) that will lead to the desired state. We will first discuss different assumptions lead to different moves and information states and then what the use of such an analysis might be for the organisation of instructional databases.

Moves and Information States

Let's look at three different cases of user-manual interaction and see how the assumptions about interactions change the available information states and moves[5]

- case 1: traditional manual: no overt interaction (except for the fact that the user opens the manual, or clicks on a particular section)
- case 2: the manual can ask yes/no questions and understand 2 types of user responses:
  - yes/no responses
  - 'done' or 'don't understand' response after the execution of an instruction
- case 3: the user can ask indicate that she already knows certain procedures or subprocedures

These types of restricted systems can also be seen in relation to the types of restricted systems considered in [Traum *et al.*(1999a)]. In this case, however, the restrictions are not mainly concerned with the types of input, output, and interpretation and generation limitations, but rather the amount of interactivity allowed, and the types of dialogue moves used.

[4]Modulo the differences in instructional material we have discussed in brief in Chapter 1.

[5]These are of course only illustrations. There is an infinite set of possibilities.

Overt and Covert Moves in the limit cases

In the first case, we can describe the procedural and instructional part of a manual with two types of overt moves for the manual and one overt type of move for the user. The overt moves are DIRECT and ASSERT for the manual and EXECUTE for the user. Things are, however, more interesting if we also look at the covert moves and the updates that have to be assumed for the whole interaction to take place.

Given that there are no overt interactions between the two participants the dialogue can only be fruitful if a clear structure of obligations and updates is assumed from the onset on: the manual is assumed not to give instructions that cannot be executed, to give only instructions that lead to the correct result, etc. and the user is assumed to be able to execute the instructions and to do so. When an instruction has been given, it is assumed that it is executed before the next one is taken into account. These assumptions seem to be a rather essential part of the way the manual game is defined. We will model them using a simplified version of PTT.

Given the following abbreviations,

- U : user
- M : manual
- a : action
- p : proposition
- i : instruction
- D : device
- s : state
- du : discourse unit

we can describe the dialogue moves used as follows:

- DIRECT(M,U,a)
- ASSERT(M,U,p)
- EXECUTE(U,i)
- ACKNOWLEDGE(U,du)

In this description, however, we go beyond the current abilities of the EDIS system [Matheson *et al.*(2000)], which implements PTT, e.g., by recording the performance of

non-dialogue acts, so as to release the obligations to perform them. We can adapt the framework in two ways in the current information state: putting them in the same field with dialogue acts, or creating another field of non-linguistic actions. One complication is that such non-linguistic acts may be communicative or not. I.e., the non-performer may or may not be able to sense that such actions have been performed. In the former case, the situation is precisely identical with communicative acts: when the action is first performed, it is placed in a DU that is part of UDUs, signifying semi-public status. When the addressee acknowledges having observed the act, this DU's contents becomes part of G, and thus the act's performance is part of common ground (and any obligations in G to perform such an act are relieved).

The situation is slightly different when actions can not be directly observed. Here, even though the performer will (privately) believe that an obligation has been lifted, the action itself, and the lifting of the obligation will not be part of common ground. In this case, the actor can assert that he has performed the action. This will thus introduce in a DU an SCP for this agent that the action has been performed. If this DU is acknowledged, then it will be common ground that the agent has such an SCP. This is not enough, however, to release the obligation. In addition to acknowledging the DU, the responder must actually *accept* the claim, thus committing herself to this claim as well. This mutual commitment of the action having been performed is thus enough to release the obligation.[6]

In addition to directing a dialogue partner to perform an action, it is also often useful to issue directions concerning a state of affairs. Here, the effect will be on the addressee's future action rather than the speaker, but concerning a proposition rather than an action as the complement. There are still a number of different kinds of action that could be relevant to this proposition and the types of commitment and obligation involved. Rather than complicate the ontology of dialogue acts, we will instead introduce a set of *meta-actions* that will be the complement of the directive act, and will also be the obligation of the addressee to perform, given an acceptance of such a directive. We use s and a to refer to the states and the actions in the world that corresponds to propositions or instructions in the dialogue or monologue.

- ACHIEVE(U,s) - U brings about the case in which s pertains, or more formally: $\exists$ a :DONE(a,U) $\wedge$ CAUSE(a,Holds(s))
- VERIFY(U,?([],s)) - A determines whether or not s pertains, e.g., KNOWIF(U,hods(s))
- CERTIFY(A,s) - A compound of the previous two, in which U verifies s, and if s does not pertain, U achieves s.

[6]Achieving common ground that the obligations have been released, can thus lead to a need for task-confirmations, in a manner similar to Cohen and Levesque's joint-intentions [Cohen and Levesque(1991)].

- If(U,s,x,y) - U performs x or y, depending on whether s is pertains or not. Either x or y could be the null action or a compound action involving another conditional. In the case in which x and y are distinct, performing x involves an SCP that s pertains, while performing y involves an SCP that s does not pertain.

If an agent has an obligation to perform one of these acts, this is equivalent to a commitment to the holding in the future of the effects of these acts. Likewise, an agent reporting on having done the acts will hold commitments to the effects holding.

Given these considerations, ASSERT-moves by the manual lead to an update that SCP the user to p and to an obligation of ACHIEVE(U,s). A DIRECT-move of the manual also leads to an update that obliges the user to ACHIEVE(s). If desired, we can also use the above action ontology to specialize the *direct* move into subclasses, e.g., we can define For DIRECT-CHECK(M,U,s) which is defined as DIRECT(M,U,CERTIFY(M,s)). In this case the user is only obliged to bring the state of affairs about it it does not yet pertain. We assume that an EXECUTE by the user leads to ACHIEVE(U,s) and a covert acknowledge update on the side of the manual: the rest of the text is written under the assumption that the instruction has been executed.

We could also model some very limited failures. For instance we could define a REJECT-move which leads to a return to the state before the previous instruction.

The instructional part of the manual can now be modeled as a dialogue composed of a sequence of alternations between DIRECT moves and EXECUTE-moves which can be assumed to include a covert ACKNOWLEDGE-move. The only other possibility is that instead of an INSTRUCT-EXECUTE-move the manual makes an INSTRUCT-CHECK-move. This can be followed by a move of the user where both the acknowledgment and the execute move are covert, in the sense that after checking the state of the device, the user has nothing to do but to pass to the next instruction. [7]

The descriptive part of the manual consists basically of ASSERT moves by the manual which impose on the user only a very vague obligation: to evaluate of how much use they are for him with respect to the goal with which she consults the manual.

When we move to the second case, the inventory of moves gets extended by yes/no questions and overt ACCEPT- or REJECT- moves. Given that the manual is now 'aware' of some information it didn't have before about the state of the device, it has the obligation to adapt the information it gives to the pertaining state of user+device. It is only in this case, where a minimal adaptation of the moves of one DP to the IS of the other is possible, that the term 'dialogue' as used normally becomes applicable.

[7] This is a simplification: sometimes there is an implicit assumption that the user will create the desired state even if it doesn't yet pertain.

The difference in available moves and obligations changes the plan that is needed in this situation. To take just the situation in which the manual asks the user to check the state of the device (in general to see whether a post- or a pre-condition holds), it is now necessary to present only information about the case that pertains whereas before the manual was obliged to give information for all the possible cases given it could not obtain information about the actual situation.

The third case, of course, changes the game rather drastically: now the manual has to adapt its ASSERT moves to the user's knowledge. This means that the plan has to include ways to access this information. This change implies that the user can impose obligations on the system by asking questions.

In the next section we describe the way in which the GoDiS system has been adapted to model the relation between a traditional manual and a limited dialogue. In future work, we also expect to extend the EDIS system to fully formalize the relations described above in the PTT model.

4.2 From Manual Text to Instructional Dialogue: adding Task Plans and extending the set of Moves

We use the formal representation of dialogue information states that has been developed in GoDis with the addition of a subfield SHARED.ACTIONS whose value is a stack of actions which the system has instructed the user to perform but whose performance has not yet been confirmed by the user. We call this variant of GoDiS, IMDiS for Instructional Monologue and Dialogue System. The IMDiS information state is shown in (1).

(1)
$$
\text{IS}: \left[\begin{array}{lll}
\text{PRIVATE} & : & \left[\begin{array}{lll}
\text{PLAN} & : & \text{StackSet(Action)} \\
\text{AGENDA} & : & \text{Stack(Action)} \\
\text{BEL} & : & \text{Set(Prop)} \\
\text{TMP} & : & \left[\begin{array}{lll}
\text{BEL} & : & \text{Set(Prop)} \\
\text{QUD} & : & \text{Stack(Question)} \\
\text{ACTIONS} & : & \text{Stack(Action)} \\
\text{LU} & : & \text{Utterance}
\end{array}\right]
\end{array}\right] \\
\text{SHARED} & : & \left[\begin{array}{lll}
\text{BEL} & : & \text{Set(Prop)} \\
\text{QUD} & : & \text{StackSet(Question)} \\
\text{ACTIONS} & : & \text{Stack(Action)} \\
\text{LU} & : & \text{Utterance}
\end{array}\right]
\end{array}\right]
$$

$$
\text{Utterance}: \left[\begin{array}{lll}
\text{SPEAKER} & : & \text{Participant} \\
\text{MOVES} & : & \text{assocSet(Move,Bool)}
\end{array}\right]
$$

The main division in the information state is between information which is private to the agent and that which is shared between the dialogue participants. The private part of the

information state contains a PLAN field holding a dialogue plan, i.e. is a list of dialogue actions that the agent wishes to carry out. The plan can be changed during the course of the conversation. The AGENDA field, on the other hand, contains the short term goals or obligations that the agent has, i.e. what the agent is going to do next. We have included a field TMP that mirrors the shared fields. This field keeps track of shared information that has not yet been grounded, i.e. confirmed as having been understood by the other dialogue participant. The SHARED field is divided into four subfields. One subfield is a set of propositions which the agent assumes for the sake of the conversation. The second subfield is for a stack of questions under discussion (QUD). These are questions that have been raised and are currently under discussion in the dialogue. The ACTIONS field is a stack of (domain) actions which the user has been instructed to perform but has not yet performed.The LU field contains information about the latest utterance.

The monologue mode uses only 3 moves, plain instructions (**InstructExec**, instructions to check preconditions **InstructCheck** and **Inform**). Since there is no user to confirm that actions have been performed, all actions are automatically confirmed using the update rule **autoConfirm**.

(2) RULE: **autoConfirm**
CLASS: **integrate**
PRE: { fst#rec(shared.actions, A)
EFF: { pop#rec(shared.actions)
add#rec(shared.bel, done(A))

The dialogue version uses 10 move types, basically the 6 used in GoDiS (**Ask**, **Answer**, **Repeat**,**ReqRep**, **Greet**, **Quit**) plus **InstructCheck**, **InstructExec**), confirmations (**Confirm**) and **Inform**. Confirmations are integrated by assuming that the current top-most action in SHARED.ACTIONS has been performed, as seen in the update rule in (3).

(3) RULE: **integrateUsrConfirm**
CLASS: **integrate**
PRE: { val#rec(shared.lu.speaker, usr)
assoc#rec(shared.lu.moves, confirm, false)
fst#rec(shared.actions, A)
EFF: { set_assoc#rec(shared.lu.moves, confirm, true)
pop#rec(shared.actions)
add#rec(shared.bel, done(A))

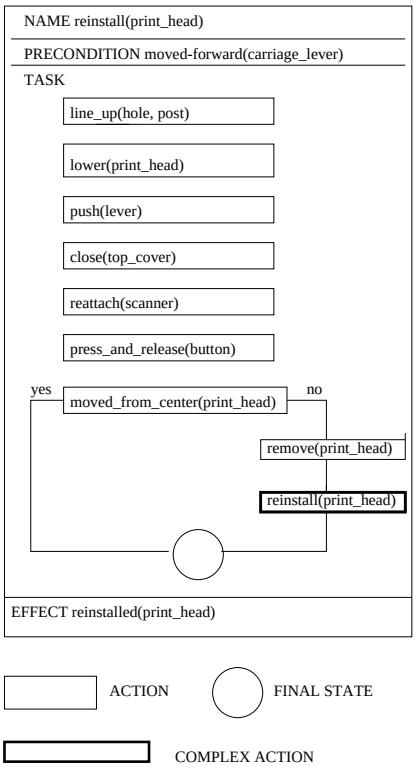
4.2.1 Domain task, manuals and dialogues

Let's see how a monologue and a dialogue version of the same task are related. Below we have again the example from the user manual for the HomeCentre.

- Reinstalling the print head
- Caution: Make sure that the green carriage lock lever is STILL moved all the way forward before you reinstall the print head.
- 1. Line up the hole in the print head with the green post on the printer carriage.
- Lower the print head down gently into position.
- 2. Gently push the green cartridge lock lever up until it snaps into place.
- This secures the print head.
- 3. Close the top cover and re-attach the scanner.
- 4. Press and release the yellow LED button.
- The printer will prepare the cartridge for printing.
- Note: If the carriage does not move from the center position after you press the cartridge change button, remove and reinstall the print head.

From this text, one can (re)construct a task plan for reinstalling the print head. Such a plan may be represented as in (4). Note that this is a conditional plan, i.e. it contains branching conditions.

(4)



From this plan, IMDiS generates two plans: a monologue plan and a dialogue plan. This is done using the "translation schema" in (5)

(5)

DOMAIN	MONOLOGUE	DIALOGUE
precondition P	**InstructCheck**(P)	**InstructCheck**(P)
action A	**InstructExec**(A)	**InstructExec**(A)
`if_then_else`(C, A, A')	**Inform**(`if_then_else`(C, A, A'))	`findout`(C); `if_then_else`(C, **InstructExec**(A), **InstructExec**(A'))
effect E	**Inform**(E)	**Inform**(E)

The difference between the text plan and the dialogue plan is in the way that conditionals in the domain plan are interpreted. In the monologue plan, they correspond to simply informing the user of the conditional. In dialogue mode, however, the system raises the question whether the condition holds. When the system finds out if the condition holds, it will instruct the user to execute the appropriate guarded action.

Conditionals are treated as follows by the system in dialogue mode: When the system has found out what the user's task is, it will load the appropriate dialogue plan into the PRIVATE.PLAN field of the information state. It will then execute the actions in the appropriate order by moving them to the agenda and generating appropriate utterances. When a conditional statement is topmost on the plan, IMDiS will check whether it has been established that the condition holds (by checking the SHARED.BEL field). Since the system has previously asked the user and the user has answered, either the condition or its negation will be in the set of established propositions. If the condition or its negation holds, the conditional will be popped off the plan and replaced by the first or second guarded action (respectively).

4.2.2 Monologue and Dialogue Behavior

In the monologue mode in IMDiS, the control module does not call the input and interpretation modules. The text is output "move by move" as a sequence of utterances from the system.

(6)
```
S: Reinstalling the print head.
S: Make sure that the green carriage lock
lever is STILL moved all the way forward
before you install the print head.
S: Line up the hole in the print head with the
green post on the printer carriage
...
```

Compared to the monologue mode, even a very restricted dialogue mode offers several advantages:

User attention and control The user can direct her attention to the machine and does not have to look at the manual. This means that the user does not have to keep track of the current point in the manual. Also, the user decides when to move on to the next action, by confirming that the previous action is done.

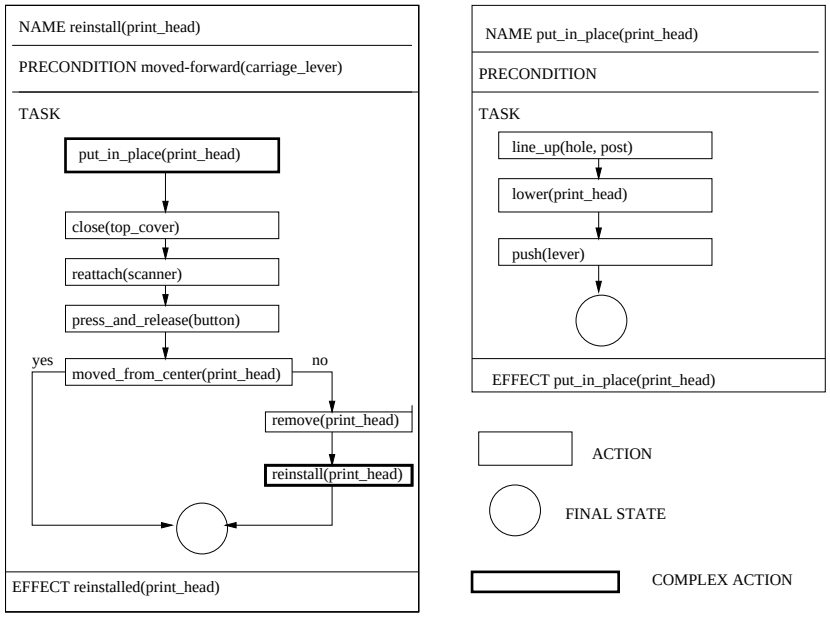
Avoid irrelevant information When the action to be taken depends on a condition, the system does not give irrelevant information.

(7)
```
S: Has the carriage moved from the center
position?
U: yes
S: The print head is now installed
```

4.2.3 More complex task plans

In the example above we illustrated how a simple task plan can give rise to a dialogue and a monologue rendering. In that case the overall structure of the monologue and the dialogue stay rather similar and we can see that some linguistic clues in the monologue versions like 'caution', 'make sure that', 'note', and conditionals (i.e. the clues that lead us in the monologue discourse to postulate 'asides') give us hints to what the dialogue plan will be. We can, however, get some added flexibility by giving more structure to the the task plan. For instance in the example above, one can argue that the reinstallation proper of the print head is described point 1 to 2 and that 3 and 4 describe termination conditions. To reflect this we can revise the task plan as follows:

(8)



With this structure the user can control the level of detail of the instructions given (case 3). If the user does not know how to perform a substep, she can ask the system for more detailed instructions.

(9)
```
S: put the print head in place
U: how?
S: Line up the hole in the print head with the green
post on the printer carriage
U: right, ok
S: Lower the print head down gently into position
U: ok
S: Gently push the green cartridge lock lever up until
it snaps into place
U: ok
S: The print head is now securely in place
```

On the other hand, if the user already knows how to perform a substep, the system moves on to the next step.

(10)
```
S: put the print head in place
U: done, and now?
S: Close the top cover
```

If the user manages to complete the whole action sequence without instructions, she can tell the system this and the system proceeds to the next relevant action.

(11)
```
S: put the print head in place
U: how?
S: Line up the hole in the print head with the green
post on the printer carriage
U: right, done
S: Lower the print head down gently into position
U: done, I remember now (pause) the print head is put
in place
S: Ok.  Close the top cover
```

Note that we can also allow for some explicit grounding and question accommodation.

Grounding If the user does not hear or understand a system utterance, she can ask the system to repeat it.

(12)
```
S: Has the carriage moved from the center position?
U: I didn't understand
S: Has the carriage moved from the center position?
```

Question accommodation A minimum of accommodation is also provided: for example, the user does not have to wait for the system to ask what task the user wants to perform.

(13)
```
S: Hello and welcome to the IMDiS homecentre
assistant
U: i want to reinstall the print head
S: Make sure that the green carriage lock lever is
still moved all the way forward before you install the
print head.
```

In (14) we show a sample dialogue which exhibits these features. Utterances in italics could have been skipped (U3-S7 by not asking "how?"; U11-S12 by not asking "what?").

(14) S1: Hello and welcome to the IMDiS homecentre assistant
U1: i want to reinstall the print head
S2: Make sure that the green carriage lock lever is still moved all the way forward before you install the print head.
U2: ok, done
S3: secure the print head
U3: how?
S4: Line up the hole in the print head with the green post on the printer carriage
U4: right, ok
S5: Lower the print head down gently into position
U5: ok
S6: Gently push the green cartridge lock lever up until it snaps into place
U6: ok
S7: The print head is now secured
U7: ok, and now?
S8: Close the top cover
U8: done
S9: Reattach the scanner
U9: ok
S10: Press and release the yellow LED button
U10: done
S11: Has the carriage moved from the center position?
U11: what ?
S12: Has the carriage moved from the center position?
U12: yes
S13: The print head is now installed
U13: ok
S14: The printer will prepare the cartridge for printing.
U14: ok, thanks
S15: thank you for using the IMDiS homecentre assistant

Although fairly repetitive, this dialogue exhibits all the features mentioned above: instruction-confirmation pairs, answering minimal "how"-questions (U3-S7), question accommodation (U1), and grounding behavior (U13, U11-S12). In this case, however, the hierarchical structure is different from than exhibited by the original monologue and we see that the differences come not from linguistic information but from assumptions about the task structure. We rejoin here the observations made by the intentional school but with the

proviso of Grosz and Sidner [Grosz and Sidner(1986)]: the dialogue plan is not the task structure itself but it is derived from the task structure and knowledge about allowable moves.

4.2.4 Experiment and Theory

In building the experimental IMDiS, we have made several simplifications. All NL generation problems have been side-stepped by using canned text for output. Around 90% of the lexicon is used in both dialogue and monologue mode, while the rest is specific to one mode. It is a research issue to what extent canned text can be used, and how much "real" generation is necessary. In the perspective on transformation from classical manuals to more interactive systems, it seems that some clues can be used to do minimal textual transformations to adapt the existing text to the new use. But this needs further investigation.

IMDiS is also not capable of referent disambiguation dialogue. This type of dialogue would be needed for the system to be able to explain e.g. which component is being referred to and where it is to be found. The limited interaction capabilities we are assuming here do not allow for this.

4.3 Conclusion

In the previous section we illustrated an analysis of dialogue moves and obligations based on simple assumptions about more or less interaction between a manual and a user. This analysis shows that various degrees of interactivity change the way information gets organized and that it is not enough to recast the text as a set of questions and answers to approach real interactivity. The TRINDI tools allow us to describe the different cases of interactivity at a certain level: that of moves, information states and QUDs. From there we can reason back to plans.

The experiment suggests that some of the textual elements can be reused in the transformation from monologue to limited dialogue presentation but an important issue we have not addressed is is how to construct the task plans and connect them to the corresponding text output. In the last example given above, the structure of the task plans is not the same as that of the discourse as described in section 1 of chapter 2. In the text given as an example above, the various instructions will normally be seen as enumerations. So where could the additional structure come from? In [Zaenen and Larsson(1999)] we proposed to add an activity mark-up to the discourse structure mark-up. This type of structure allows

us to construct the task plan from an existing text but it does not tell us immediately how to present information about subtasks. E.g. in example (9) above, we could have given as an answer to the question "how?": "lower the print head into the position and gently push the cartridge lock lever up until it snaps into place." When we have the goal to describe a certain procedure, we surely need to give information about the core of the procedure but it is not immediately clear how much needs to be said about the preparatory and consecutive actions.

In any case for such a mark-up to be practically useful, we need to be able to extract this task structure semi-automatically. Linguistic (as partially specified in [Zaenen and Larsson(1999)]) and lay-out information can be used for this. It is, however, not sure that this will be an easy task. Further investigation is necessary to determine whether this kind of transformation is a worth while enterprise.

Regardless of the answer to this question, the material discussed here points to an approach to document creation which would start with a definition of the task structure and the construction of various interaction models, from traditional manual to full dialogue. These models would specify for each case which DM and IS information is available. On the basis of this, a schema for each case could be developed and the task descriptions could be augmented with annotations that allow the various realizations. This approach to document creation would be less abstract than the one mentioned in the introduction where it is envisioned to start from a abstract description of the device but it would be more abstract than the main theme pursued in this deliverable which was to transform a classical manual into a limited dialogue system.

Bibliography

[Allen and Perrault(1980)] Allen, J. F. and Perrault, C. R. (1980). Analyzing intention in utterances. *Artificial Intelligence*, **15**(3), 143–178.

[Allwood(1997)] Allwood, J. (1997). Notes on dialogue and cooperation. In D. K.Jokinen, D.Sadek, editor, *Collaboration, Cooperation and Conflict in Dialogue Systems*, IJCAI Workshop.

[Asher and Lascarides(1993)] Asher, N. and Lascarides, A. (1993). Lexical disambiguation in a discourse context. *Journal of Semantics*, **12**, 67–108.

[Carlson(1983)] Carlson, L. (1983). *Dialogue Games: an Approach to Discourse Analysis.* Reidel, Dortrecht.

[Clark and Schaefer(1989)] Clark, H. and Schaefer, E. (1989). Contributing to discourse. *Cognitive Science*, **13**, 259–294.

[Cohen and Levesque(1990)] Cohen, P. R. and Levesque, H. J. (1990). Rational interaction as the basis for communication. In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication.* MIT Press.

[Cohen and Levesque(1991)] Cohen, P. R. and Levesque, H. J. (1991). Confirmations and joint action. In *Proceedings IJCAI-91*, pages 951–957.

[Cohen and Perrault(1979)] Cohen, P. R. and Perrault, C. R. (1979). Elements of a plan-based theory of speech acts. *Cognitive Science*, **3**(3), 177–212.

[Cooper and Larsson(1999)] Cooper, R. and Larsson, S. (1999). Dialogue moves and information states. In H. Bunt and E. C. G. Thijsse, editors, *Proceedings of the Third International Workshop on Computational Semantics.*

[Discourse Resource Initiative(1997)] Discourse Resource Initiative (1997). Standards for dialogue coding in natural language processing. Report no. 167, Dagstuhl-Seminar.

[Ginzburg(1996)] Ginzburg, J. (1996). Interrogatives: questions, facts and dialogue. In *The Handbook of Contemporary Semantic Theory.* Blackwell, Oxford.

[Grosz and Sidner(1986)] Grosz, B. and Sidner, C. (1986). Attention, intention, and the structure of discourse. *Computational Linguistics*, **12**(3), 175–204.

[Heim(1982)] Heim, I. (1982). *The Semantics of Definite and Indefinite Noun Phrases.* Ph.D. thesis, University of Massachusetts at Amherst.

[Hobbs(1993)] Hobbs, J. (1993). Intention, information, and structure in discourse. *Proceedings of the NATO Advanced Research Workshop on Burning Issues in Discourse*, pages 41–66.

[Hovy(1991)] Hovy, E. (1991). Approaches to the planning of coherent text. In W. M. Cecile Paris, William Swartout, editor, *Natural Language Generation in Artificial Intelligence and Computational Linguistics.* Kluwer Academic.

[Kehler(1995)] Kehler, A. (1995). *Interpreting Cohesive forms in the context of Discourse Inference.* Ph.D. thesis, Harvard University.

[Knott and Dale(1993)] Knott, A. and Dale, R. (1993). Choosing a set of rhetorical relations for text generation. In *Proceedings of the Fourth European Workshop on Natural Language Generation.*

[Kuppevelt(1995a)] Kuppevelt, J. V. (1995a). Discourse structure, topicality and questioning. *Journal of Linguistics*, **31**, 109–147.

[Kuppevelt(1995b)] Kuppevelt, J. V. (1995b). Main structure and side structure in discourse. *Linguistics*, **33**, 809–833.

[Lascarides and Asher(1991)] Lascarides, A. and Asher, N. (1991). Discourse relations and defeasible knowledge. *Proceedings of the ACL.*

[Mann and Thompson(1988)] Mann, W. and Thompson, S. (1988). Rhetorical structure theory: Toward a functional theory of text organisation. *Text*, **22**, 243–281.

[Marcu(1997)] Marcu, D. (1997). The rhetorical parsing of natural language text. *Proceedings of the 35th Annual Meeting*, **35**.

[Matheson *et al.*(2000)] Matheson, C., Poesio, M., and Traum, D. (2000). Modelling grounding and discourse obligations using update rules. In *Proceedings of the First Conference of the North American Chapter of the Association for Computational Linguistics.*

[Moore and Paris(1993)] Moore, J. and Paris, C. (1993). Planning text for advisory dialogues. *Computational Linguistics*, **19**, 651–694.

[Moore and Pollack(1992)] Moore, J. and Pollack, M. (1992). A problem for rst: The need for multi-level discourse analysis. *Computational Linguistics*, **18**(4), 537–544.

[Poesio and Traum(1997)] Poesio, M. and Traum, D. R. (1997). Conversational actions and discourse situations. *Computational Intelligence*, **13**(3).

[Poesio and Traum(1998)] Poesio, M. and Traum, D. R. (1998). Towards an axiomatization of dialogue acts. In *Proceedings of Twendial'98, 13th Twente Workshop on Language Technology: Formal Semantics and Pragmatics of Dialogue.*

[Pollack(1990)] Pollack, M. E. (1990). Plans as complex mental attitudes. In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication.* MIT Press.

[Prust *et al.*(1994)] Prust, H., Scha, R., and van den Berg, M. (1994). Discourse grammar and verb phrase anaphora. *Linguistics and Philosophy*, **17**, 261–327.

[Sadek(1991)] Sadek, M. D. (1991). Dialogue acts are rational plans. In *Proceedings of the ESCA/ETR workshop on multi-modal dialogue.*

[Scha and Polanyi(1988)] Scha, R. and Polanyi, L. (1988). An augmented contextfree grammar for discourse. In *Proceedings of the 12th International Conference on Computational Linguistics*, COLING.

[Stalnaker(1979)] Stalnaker, R. (1979). Assertion. In P. Cole, editor, *Syntax and Semantics 9.* Academic Press.

[Traum(1994)] Traum, D. (1994). *A Computational Theory of Grounding in Natural Language Converstation.* Ph.D. thesis, University of Rochester.

[Traum *et al.*(1999a)] Traum, D., Bohlin, P., Bos, J., Ericson, S., Larsson, S., Lewin, I., Matheson, C., and Milward, D. (1999a). Dialogue dynamics and levels of interaction. Technical Report Deliverable D3.1, Trindi.

[Traum *et al.*(1999b)] Traum, D., Bos, J., Cooper, R., Larsson, S., Lewin, I., Matheson, C., and Poesio, M. (1999b). A model of dialogue moves and information state revision. Technical Report Deliverable D2.1, Trindi.

[Traum and Hinkelman(1992)] Traum, D. R. and Hinkelman, E. A. (1992). Conversation acts in task-oriented spoken dialogue. *Computational Intelligence*, **8**(3), 575–599. Special Issue on Non-literal language.

[van den Berg(1996)] van den Berg, M. (1996). Discourse grammar and dynamic logic. In *Proceedings of the Tenth Amsterdam Colloquium*, ILLC, Amsterdam.

[Webber and Joshi(1998)] Webber, B. L. and Joshi, A. (1998). Anchoring a lexicalized tree-adjoining grammar for discourse. *Proceedings of the ACL Workshop on Discourse Relations and Discourse Markers.*

[Zaenen and Larsson(1999)] Zaenen, A. and Larsson, S. (1999). Annotations for instructional texts. Deliverable D1.2, Trindi Project.