Natural Language Engineering **6** (3–4): 323–340. Printed in the United Kingdom

© 2000 Cambridge University Press

Information state and dialogue management in the TRINDI dialogue move engine toolkit

STAFFAN LARSSON
Department of Linguistics, Göteborg University, Box 200, SE-405 30 Göteborg, Sweden
e-mail: sl@ling.gu.se

DAVID R. TRAUM
USC Institute for Creative Technologies, 13264 Fiji Way, Marina del Rey, CA 902, USA
e-mail: traum@ict.usc.edu

(*Received 23 September 1999, revised 26 March 2000*)

Abstract

We introduce an architecture and toolkit for building dialogue managers currently being developed in the TRINDI project, based on the notions of *information state* and *dialogue move engine*. The aim is to provide a framework for experimenting with implementations of different theories of information state, information state update and dialogue control. A number of dialogue managers are currently being built using the toolkit, and we present overviews of two of them. We believe that this framework will make implementation of dialogue processing theories easier, also facilitating comparison of different types of dialogue systems, thus helping to achieve a prerequisite for arriving at a best practice for the development of the dialogue management component of a spoken dialogue system.

1 Introduction

We discuss a candidate model for best practice in the development of the dialogue management component of a spoken dialogue system. At this early stage in development of such systems, there is a wide range of current practice, depending in part on the different tasks that such systems are put to, but also on external factors, such as preference and background of the developers. To determine what the *best practice* is, if indeed there can be a single model that is useful for the whole range of such systems, it is necessary to be able to compare and evaluate these components. A big difficulty is that not only are the dialogue management components themselves radically different, but the systems of which they are a part are also quite heterogeneous, making principled comparison of the components difficult, if not impossible.

What we propose is a view of dialogue management functions in terms of *information state*. Key to this approach is identifying the relevant aspects of information in dialogue, how they are updated, and how updating processes are controlled. This simple view can be used to compare a range of approaches and specific theories

https://doi.org/10.1017/S1351324900002539 Published online by Cambridge University Press

of dialogue management within the same framework (as well as facilitating hybrid approaches). Use of the framework will thus allow comparison to determine empirically which is the best practice.

The term *information state* of a dialogue represents the information necessary to distinguish it from other dialogues, representing the cumulative additions from previous actions in the dialogue, and motivating future action. For example, statements generally add propositional information; questions generally provide motivation for others to provide specific statements. Information state is also referred to by similar names, such as 'conversational score' or 'discourse context' and 'mental state'. Generally, although not necessarily, we will also talk about the information state of participants of the dialogue, representing the information that those participants have at a particular point in the dialogue – what they brought with them to the dialogue, what they pick up, and how they are motivated to act in the (near) future.

We present a particular view toward the formalization of the notion of information state in such a way that allows specific theories of dialogue to be formalized, implemented, tested, compared and iteratively reformulated. Key to this approach will be a notion of *update* of information state, with most updates related to the observation and performance of *dialogue moves*. We will present the framework for modeling information state in the next section. As well as the framework itself, we have been developing a toolkit to allow system designers to build dialogue management components according to their particular theories of information states, this will be described in Section 3. We have also begun to build several different sorts of systems using the toolkit, some of which will be briefly described in Section 4. We conclude, in Section 5, with a discussion comparing this approach to some other methods for building dialogue management components.

2 The information state approach to dialogue modeling

We view an information state theory of dialogue modeling as consisting of:

- a description of the *informational components* of the theory of dialogue modeling, including aspects of common context as well as internal motivating factors (e.g. participants, common ground, linguistic and intentional structure, obligations and commitments, beliefs, intentions, user models, etc.)
- *formal representations* of the above components (e.g. as lists, sets, typed feature structures, records, Discourse Representation Structures (DRSs), propositions or modal operators within a logic, etc.)
- a set of *dialogue moves* that will trigger the update of the information state. These will generally also be correlated with externally performed actions, such as particular natural language utterances. A complete theory of dialogue behavior will also require rules for recognizing and realizing the performance of these moves, e.g. with traditional speech and natural language understanding and generation systems.
- a set of *update rules*, that govern the updating of the information state, given various conditions of the current information state and performed dialogue moves, including (in the case of participating in a dialogue rather than just

https://doi.org/10.1017/S1351324900002539 Published online by Cambridge University Press

monitoring one) a set of selection rules, that license choosing a particular dialogue move to perform given conditions of the current information state

- an *update strategy* for deciding which rule(s) to select at a given point, from the set of applicable ones. This strategy can range from something as simple as "pick the first rule that applies" to more sophisticated arbitration mechanisms, based on game theory, utility theory, or statistical methods.

It is important to distinguish information state approaches to dialogue modeling from other, structural, dialogue state approaches. These latter approaches conceive a 'legal' dialogue as behaving according to some grammar, with the states representing the results of performing a dialogue move in some previous state, and each state licensing a set of allowable next dialogue moves. The 'information' is thus implicit in the state itself and the relationship it plays to other states. It may be difficult to transform an information state view to a dialogue state view, since there's no necessary finiteness restriction on information states (depending on the type of information modeled), and the motivations for update and picking a next dialogue move (using update rules, and update strategy) may rely on only a part of the information available, rather than the whole state. On the other hand, it is very easy to model dialogue state as information state: the information is the dialogue state, itself. This is easily modeled as a register indicating the state number (for finite state models, or a stack for recursive transition networks). The dialogue moves will be the same moves that are used in the dialogue state theory, the update rules will be the transitions in the dialogue state theory, formulated as an update to a new state, given the previous state and performance of the action, and the update strategy will be much the same as in the transition network (i.e. deterministic or non-deterministic, etc.)

Structural dialogue state approaches have often been contrasted with plan-based approaches to dialogue modeling (e.g. by Cohen (1996) and Sadek and De Mori (1998)). Structure-based approaches are usually viewed as viable for simple, scripted dialogues, while plan-based approaches, though more complex and difficult to embed in practical dialogue systems, are seen as more amenable to flexible dialogue behavior. Plan-based approaches are also criticized as being more opaque, especially given the large amount of procedural processing and lack of a well-founded semantics for plan-related operations. An information-state approach allows one to fruitfully combine the two approaches, using the advantages of each. The information state may include aspects of dialogue state as well as more mentalistic notions such as beliefs, intentions, plans, etc. Moreover, casting the updates in terms of update rules and strategies that apply the rules under appropriate conditions provides for a more transparent, declarative, representation of system behavior than most procedural programs, rendering the resulting dialogue manager easily amenable to experimentation with different dialogue strategies.

In the rest of this section, we will present the aspects of information state in a little more detail. To keep a degree of concreteness, we will make reference to an example theory of information state developed by Cooper and Larsson, discussed in Cooper, Larsson, Matheson, Poesio and Traum (1999).

https://doi.org/10.1017/S1351324900002539 Published online by Cambridge University Press

2.1 Informational components

Information state is usually not conceived of as a monolithic node in a transition network (as with dialogue state), but rather as consisting of several interacting components. There is a wide range of possibilities as to what kinds of components should be used to model dialogue. The first choice point comes as to whether to model the participants' internal state, or more external aspects of the dialogue. There are also many ways of modeling the internal state of a participant. One can choose to model the mental state of the agent (attitudes such as belief, desire, intention, along with social correlates such as mutual belief, joint intention and obligation) (e.g. Breiter and Sadek 1996; Traum and Allen 1994), or one can take a more structural view of the dialogue, concentrating on the performance of actions and various sorts of accessibility relationships. (e.g. Ahrenberg, Dahlbäck and Jönsson 1990). It may also be useful to distinguish components of information state into *static* and *dynamic* aspects. The former are those aspects of information state that are not expected to change during the course of a dialogue, but are still very useful to modeling the progression of the dialogue. Examples of static information state components could include domain knowledge, knowledge of dialogue conventions, or similar sorts of information. It depends upon the type of dialogue being modeled, and the scope of the conversation as to which aspects will be assumed to be static vs. dynamic (e.g. contrasting a knowledge acquisition system vs. a question answering system – the former may want to treat domain knowledge as dynamic, while the latter would see it as static). Marking some information as static may have some advantages in efficient implementation, since various compilation shortcuts and memory allocation techniques could be performed. It is still good practice to have declarative knowledge sources rather than implicit knowledge in program (or finite state automaton) control structure, in order to be able to reuse the same knowledge for different dialogue situations.

Our example information state is a simplified version of the dialogue game board proposed by Ginzburg (1996a, 1996b, 1998). There is some information assumed to be private (including *beliefs*, and an *agenda* of actions to perform in the dialogue) and some that is assumed to be shared (propositions assumed to be shared *beliefs*, questions under discussion (*QUD*), and the latest dialogue move performed (*lm*)). This small set of informational elements was used to track the behavior of participants in information seeking dialogues, including asking and answering (potentially elliptical) questions and accumulating information (Cooper *et al.* 1999; Poesio, Cooper, Larsson, Matheson and Traum 1999).

2.2 Formal representations

Given a choice of what aspects of the dialogue structure and the participants' internal state to model, the question then arises as to *how* to model them. There are a wide number of choices, from simple abstract data types, to more complex informational systems, such as logics (with associated inference systems) and statistical systems of various flavors. These choices will be related to the particular theory of accessibility of these elements, and will also affect other processing issues, such as

https://doi.org/10.1017/S1351324900002539 Published online by Cambridge University Press

comprehensiveness and efficiency. As an example, consider an aspect of information state such as actions to be performed in the dialogue (e.g. an agenda, plan, or other such bundle of intentions). There is a choice as to whether to represent tokens separately (e.g. with some sort of list) or just types, not distinguishing between multiple tokens of the same type (e.g. with a set). Given a choice of representing an agenda as a list, there is still the question of accessibility – should it be a FIFO queue, a LIFO stack, or some more open structure, allowing access to the whole list? Likewise, if an agent's beliefs are represented, should this be a set, some sort of ordered list, or a complete logical inference system, in which implicit beliefs are also said to hold given some configuration of explicit beliefs?

(1)
$$\begin{bmatrix} \textsc{private} : \begin{bmatrix} \textsc{bel} : \textsc{Set(Prop)} \\ \textsc{agenda} : \textsc{Stack(Action)} \end{bmatrix} \\ \textsc{shared} : \begin{bmatrix} \textsc{bel} : \textsc{Set(Prop)} \\ \textsc{qud} : \textsc{Stack(Question)} \\ \textsc{lm} : \textsc{Move} \end{bmatrix} \end{bmatrix}$$

Our example information state is represented as a record (Cooper and Larsson 1999), as shown in (1). Here, private and shared information are represented as sub-records, each with several fields. Each field is either a value, a set or a stack, with the type of information (proposition, action, question or move) indicated.

2.3 Dialogue moves

Dialogue moves are meant to serve as an abstraction between the large number of different possible communications (especially in a natural language) and the types of update to be made on the basis of these communications. Dialogue moves can also provide an abstract level for content generation. There are also a number of dialogue move taxonomies to choose from; some principles regarding this issue are outlined in Traum (1999). There must be at least sufficient types of dialogue moves to provide the different kinds of updates desired. The set of dialogue moves to choose is also influenced by the task of language interpretation – how easy will it be to (reliably) determine that one move vs. another has been performed? Another complicating issue is how to capture the inherent multi-functionality of utterances – with one complex move per utterance, where each move has multiple functions, or with a (potentially complex) structure of simple moves for each utterance. Dialogue moves are often conceived of as something like *speech-acts* in the sense of (Searle 1969), though this is not a requirement; dialogue moves can be any mediating input, e.g. logical forms or even word-lattices augmented with likelihoods.

Our example information state theory uses only two types of moves, **ask** and **answer**.

2.4 Update rules

Update rules formalize change in information state as the dialogue progresses. Each rule consists of a set of applicability conditions and a set of effects. The applicability conditions specify constraints on the information state that must hold for the rule to be appropriate. Effects are changes that will be made to the information state

https://doi.org/10.1017/S1351324900002539 Published online by Cambridge University Press

when the rule has been applied (assuming that all conditions hold). Update rules are meant to encapsulate coherent bundles of change to the information state, given a particular theory of dialogue. While conditions and effects are related to coherent operations on an abstract datatype, regardless of application to a theory of dialogue, update rules specialize these operations further to be specific (potentially complex) building blocks of a dialogue theory.

Continuing our example information state theory, the rule for adding a question to QUD if an **ask** move has been performed is shown in (2). This rule has two conditions: that the latest move was of type **ask**, and that the top of the agenda was the action of raising a question, the effects are to pop this item from the agenda, and push onto the QUD the question that is the content of both the *raise* agenda item and the **ask** dialogue move. Another rule is shown in (3), which pops a question from the QUD, and adds a proposition to the shared beliefs, when an answer has been performed. Other update rules in our sample information state theory include rules to add an action to answer the top question on the QUD (if asked by the other agent), and to select an answer move on the basis of the top of the agenda.

(2) U-RULE: **integrateSysAsk**

$$\text{PRE: } \begin{cases} \text{val(SHARED.LM, ask(sys,}Q\text{))} \\ \text{fst(PRIVATE.AGENDA, raise-question(}Q\text{))} \end{cases}$$

$$\text{EFF: } \begin{cases} \text{push(SHARED.QUD, } Q\text{)} \\ \text{pop(PRIVATE.AGENDA)} \end{cases}$$

(3) U-RULE: **integrateUserAnswer**

$$\text{PRE: } \begin{cases} \text{val(SHARED.LM, answer(usr,}A\text{))}, \\ \text{in(SHARED.QUD, } Q\text{)} \\ \text{answer-to(} A, Q \text{)} \\ \text{apply(} Q, A, P \text{)} \end{cases}$$

$$\text{EFF: } \begin{cases} \text{pop(SHARED.QUD, } Q\text{)}, \\ \text{add(SHARED.BEL, } P\text{)} \end{cases}$$

2.5 Update strategy

Along with the set of update rules, a strategy for how to apply the rules is needed. This is, in many cases, going to be crucial for the design of the rules themselves. Given a particular update strategy, one may need to make adaptations to the rules, and perhaps also aspects of the information state itself, in order to guarantee a particular sequence of rule applications. There's also a question of whether to have separate strategies for choosing different types of update rules (i.e. for observation and selection of dialogue moves), and whether these processes can be ordered or asynchronously applied. Some of the types of update strategies we have considered include:

1. Take the first rule that applies (iteratively until no rules apply).
2. Apply each rule (if applicable) in sequence.
3. Apply rules according to class.
4. Choose among applicable rules using probabilistic information.
5. Present choices to user to decide (for development modes).

For our sample information state, we use algorithm 1.

https://doi.org/10.1017/S1351324900002539 Published online by Cambridge University Press

2.6 Discussion

There is thus a synergy between choices for the components of information state: the conceptual notions, formal representations, dialogue moves, update rules and update strategy. A complete theory of dialogue update will need to include a smoothly interacting combination of these aspects. However, it is still very possible to hold some of them constant while trying out other possibilities for the others. For example, different ways of representing a concept, different rules for doing updates, or different update strategies for applying rules. This experimentation is a central part of our current work. In the next section, we present an architecture for implementing and testing this approach to information state update.

Given the sample informational components described above, we can track information states in simple information-seeking dialogue. A short question-answer exchange is illustrated in (4). Before the exchange, the system has an agenda item to raise the question about the user's destination. This meets the conditions for the update rule that selects an **ask** move. After the system utterance, the update algorithm will first apply the rule shown in (2). After the user utterance, the rule in (3) will check that the answer matches the question topmost on QUD, and will then pop the question off the QUD, and integrate the proposition resulting from applying the question to the answer into the shared beliefs.

(4) $\left[\text{PRIVATE} = \left[\text{AGENDA} = \langle\ \text{RAISE}(\text{X}\hat{}(\text{TO}{=}\text{X})\ \rangle\ \right]\right]$

`Sys: Where do you want to go?`

(triggers **integrateSysAsk**)

$$\left[\begin{array}{ll}\text{PRIVATE} = & \left[\text{AGENDA} = \langle\rangle\ \right] \\ \text{SHARED} = & \left[\begin{array}{ll}\text{QUD} = & \langle\ \text{X}\hat{}(\text{TO}{=}\text{X})\ \rangle \\ \text{LM} = & \text{ASK}(\ \text{X}\hat{}(\text{TO}{=}\text{X})\ )\end{array}\right]\end{array}\right]$$

`Usr: Malvern`

(triggers **integrateUsrAnswer**)

$$\left[\text{SHARED} = \left[\begin{array}{ll}\text{BEL} = & \{\ (\text{TO}{=}\text{MALVERN})\ \} \\ \text{QUD} = & \langle\rangle \\ \text{LM} = & \text{ANSWER}(\text{MALVERN})\end{array}\right]\right]$$

3 TrindiKit: A dialogue move engine toolkit

Developing a dialogue theory using the information-state approach described in the previous section yields a computational theory of dialogue that naturally lends itself to implementation. We call the implementation of such a theory of dialogue dynamics a Dialogue Move Engine (DME), since its main functions are updating information state based on the observance of moves and selecting moves to be performed. This DME, together with some connective material, forms the dialogue management and discourse tracking aspects of a dialogue system. A complete dialogue system would need modules (or sets of modules) to perform at least the following additional functions:

- user interface – to receive input from and present output to the user;

https://doi.org/10.1017/S1351324900002539 Published online by Cambridge University Press

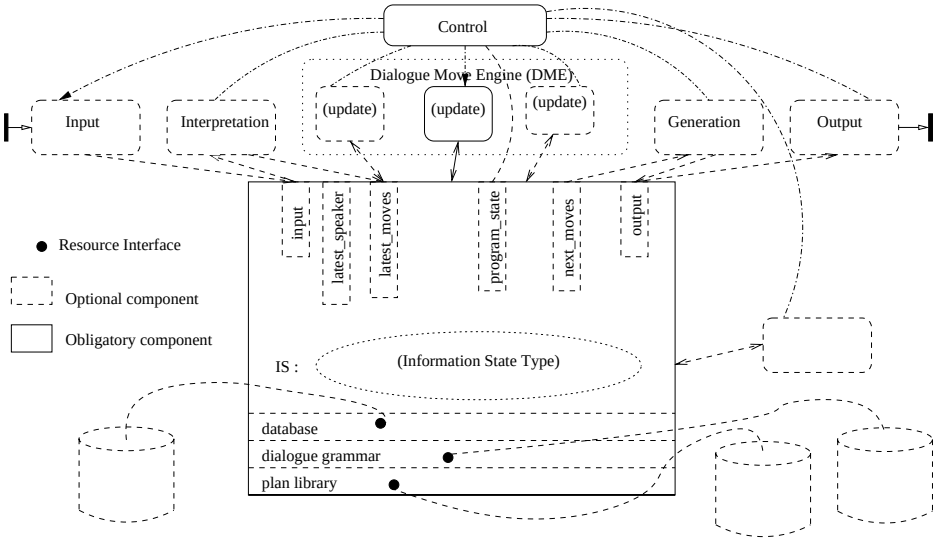


Fig. 1. The TRINDI DME architecture.

- interpretation – to calculate from the input which dialogue moves have been performed, adding these to a special *latest move* part of the information state;
- generation – to take the contents of the special *next move* part of the information state, and produce the output;
- control – to wire together the other modules, either serially or in parallel.

As part of the TRINDI project, and described more fully in Larsson, Berman, Bos, Gronqvist, Ljunglöf and Traum (2000), we are developing a DME toolkit called TrindiKit, which provides the basic architecture as well as facilities for implementing theories of information state. The general architecture of TrindiKit is shown schematically in Figure 1. Dialogue management is handled by the control module, the DME[1] and the information state.[2] The DME can consist of one or multiple UPDATE MODULES, each containing a different set of rules and potentially a different algorithm. Typically the DME contains an update module and a selection module, encapsulating the functions of integrating observed dialogue moves, and selecting new ones for the system to say. Under an architecture of this sort, it is up to the control module as to how to interleave these two functions.

[1] We use the term DME in two senses, a broader sense, meaning the implementation of the five components of the information-state based dialogue modeling approach described in the previous section, and more narrowly, within such an implementation, as the collection of modules implementing the update algorithm, applying update rules and 'consuming' and 'producing' dialogue moves, in the process.

[2] Likewise, Information State is used both to denote the components of the theoretical approach to dialogue modeling, and the specific blackboard-like system module that allows inspection and update (via update rules) of the current state.

https://doi.org/10.1017/S1351324900002539 Published online by Cambridge University Press

The components of the architecture are the total information state (TIS), consisting of the information state proper (IS), as well as interface variables for communicating with language processing modules and non-linguistic resources; the Dialogue Move Engine, consisting of one or more DME modules; other dialogue system modules (DME-external); and a control module, wiring together the other modules, either in sequence or through some asynchronous mechanism. The IS is specified using abstract data types, each permitting a specific set of queries to inspect the type and operations to change it. These are the building blocks of update rules, which can be used by other modules to inspect and change the information state in coherent ways.

Some of the components are obligatory, and others are optional or user-defined. To build a system, one must minimally supply an information state type, at least one DME module, consisting of TIS update rules and an algorithm, and a control module, operating according to a control algorithm. Any useful system is also likely to need additional modules, e.g. for getting input from the user, interpreting this input, generating system utterances, and providing output for the user. The other modules will generally communicate with the rest of the system through designated interface variables. Other resources, such as databases, plan libraries, etc., can also be integrated into the system, using an interface that allows the same kinds of queries and operations as for the IS proper, allowing update rules to be oblivious as to whether the components are part of the information state or external resources. A possible setup of DME-external modules, is indicated by dashed lines in Figure 1. Note that the illustrated module setup is just an example. The TrindiKit provides methods for defining any number of both DME-modules and DME-external modules, with associated interface variables.

Apart from the general architecture defined above, the TrindiKit provides the following dialogue system building blocks: definitions of datatypes (for use in TIS variable definitions), a language and format for specifying TIS update rules, methods for accessing the TIS, an algorithm definition language for DME and control modules, default modules for input, interpretation, generation and output, methods for converting items from one type to another, methods for visually inspecting the TIS, and debugging facilities.

4 Implementations using TrindiKit

A number of systems are currently being developed using the TrindiKit. We will look at two of them in some detail: GoDiS, developed at Gothenburg University (Bohlin, Cooper, Engdahl and Larsson 1999a), which uses an extension of the information state theory used as an example in Section 2, and the EDIS system (Matheson, Poesio and Traum 2000), developed at University of Edinburgh which uses a notion of information state based on Poesio and Traum (1998). More details of these and other TrindiKit systems can be found in Bos, Ericsson, Larsson, Lewin, Matheson and Milward 2000).

4.1 GoDiS

GoDiS is an experimental dialogue system built using the TrindiKit. It uses fairly simple algorithms for control, update and selection modules, keyword-based inter-

https://doi.org/10.1017/S1351324900002539 Published online by Cambridge University Press

pretation and template-based generation. The notion of information state used in GoDiS is an extension of that illustrated in Section 2, and is currently able to handle simple grounding phenomena as well as *question accommodation*, allowing users to answer unasked but salient questions. The GoDiS system currently distinguishes eight dialogue move types: **ask, answer, repeat, request_repeat, greet, goodbye, thank** and **quit**.

The main division in the information state is between information that is private to the system and that which is assumed to be shared between the dialogue participants. What we mean by shared information here is that which has been explicitly established during the conversation.[3] The SHARED field is divided into three 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 final SHARED field contains information about the latest move (speaker, move type and content).

The private information contains four subfields. The BEL field contains propositions that the system holds to be true. The AGENDA field contains the system's short term intentions for the next turn. The PLAN field is a list of actions that are longer-term dialogue goals. This plan can, however, be changed during the course of the conversation. We also have 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. In this way it is easy to delete information which the agent has optimistically assumed to have become shared if it should turn out that the other dialogue participant does not understand or accept it. A variant of the system (with a few different update rules) allows a system to pursue a *cautious* rather than *optimistic* strategy with respect to grounding. In this version, information will at first only be placed on TMP until it has been acknowledged by the other dialogue participant whereupon it can be moved from TMP to the appropriate shared field, in a manner similar to the way (described in the next section) CDU works in the EDIS system.

The update rules in GoDiS include an ability to perform several kinds of *accommodation* when the appropriate expected structures are not on the QUD or PLAN. In real human-human dialogues, dialogue participants can address questions that have not been explicitly raised in the dialogue. However, the question is still needed, to allow interpretation of elliptical utterances. An example from a travel agency dialogue[4] is shown in (4).

(5) J: what month do you want to go
P: well around 3rd 4th april / some time there
P: as cheap as possible

The strategy we adopt for interpreting elliptical utterances is to think of them as short answers to questions on QUD (in the sense of Ginzburg (1996a, 1996b,

[3] Akin to the 'conversational scoreboard' in Lewis (1979).
[4] This dialogue has been translated from Swedish. It was collected by the University of Lund.

https://doi.org/10.1017/S1351324900002539 Published online by Cambridge University Press

1998)). A suitable question here is *What kind of price does C want for the ticket?* This question is not under discussion at the point when the customer says "as cheap as possible", but it can be inferred from the context. In fact *J* will have as part of his plan, the action of raising this question. On our analysis it is this fact which enables *J* to interpret the ellipsis. He finds the matching question on his plan, accommodates by placing it on QUD and then continues with the integration of the information expressed by *as cheap as possible* as normal.

A similar situation arises when there is no current plan to use as context for interpretation. For example, if a travel agent discovers that his customer wishes to get information about a flight he will adopt a plan to ask her where she wants to go, when she wants to go, what price class she wants and so on. In cases where the customer does not state her errand explicitly, but rather answers some question(s) (e.g. about destination and means of transport), the agent must infer what the task is. We call this process *task accommodation*, and it is closely related to question accommodation.

A sample dialogue handled by GoDiS is shown in (6). The information state which results from utterance 3 is shown in (7).

(6)
```
1 Sys: welcome to the travel agency!
2 Usr: flights to paris
3 Sys: what city do you want to go from?
4 Usr: london
5 Sys: when do you want to travel?
6 Usr: in april, as cheap as possible
7 Sys: do you want a return ticket?
8 Usr: yes please
9 Sys: the price is $300
```

(7)
$$\left[\begin{array}{lcl} \text{PRIVATE} & = & \left[\begin{array}{lcl} \text{BEL} & = & \{\} \\ \text{AGENDA} & = & \langle\rangle \\ \text{PLAN} & = & \left\langle\begin{array}{l} \text{RAISE}(\text{R}\hat{}(\text{RETURN}{=}\text{R})), \\ \text{RAISE}(\text{M}\hat{}(\text{MONTH}{=}\text{M})), \\ \text{RAISE}(\text{C}\hat{}(\text{CLASS}{=}\text{C})), \\ \text{RESPOND}(\text{P}\hat{}(\text{PRICE}{=}\text{P})) \end{array}\right\rangle \\ \text{TMP} & = & (\text{same as SHARED}) \end{array}\right] \\ \text{SHARED} & = & \left[\begin{array}{lcl} \text{BEL} & = & \{(\text{TO}{=}\text{PARIS}),(\text{HOW}{=}\text{PLANE})\} \\ \text{QUD} & = & \langle \text{X}\hat{}(\text{FROM}{=}\text{X})\rangle \\ \text{LM} & = & \text{ASK}(\text{SYS},\text{Y}\hat{}(\text{FROM}{=}\text{Y})) \end{array}\right] \end{array}\right]$$

Utterance 2 is seen as answering two questions; however, no task or plan has yet been established and no questions have been raised. To be able to integrate the utterance, the system must find a task associated with a plan which includes the raising of questions which match the answers given. Once the task (in this case getting price information about a trip) has been accommodated and the plan entered into the PLAN field, the questions can be accommodated and the answers integrated. As a consequence of this process, the information state now contains a plan which guides the system behavior.

Given the kind of information state illustrated in (7), we can provide update rules which accommodate questions. A rule formalizing question accommodation is given

https://doi.org/10.1017/S1351324900002539 Published online by Cambridge University Press

in (8). When interpreting the latest utterance by the other participant, the system makes the assumption that it was a **reply** move with content A. This assumption requires accommodating some question Q such that A is a relevant answer to Q. The condition 'answer-to(A, Q)' is true if A is a relevant answer to Q given the current information state, according to some (possibly domain-dependent) definition of question-answer relevance.

(8) U-RULE: **accommodateQuestion**(Q, A)

PRE: { val(SHARED.LM, answer(usr,A)), in(PRIVATE.PLAN, raise(Q)), answer-to(A, Q)

EFF: { del(PRIVATE.PLAN, raise(Q)) push(SHARED.QUD, Q)

GoDiS uses an update algorithm where different types of rules are applied at different stages of the update process. There are currently six rule types and 28 rules:

- *refill*: puts new actions on the AGENDA (8 rules)
- *grounding*: handles optimistic or cautious grounding (2 rules)
- *integrate*: integrates the effects of the latest move (12 rules)
- *accommodate*: handles question (and task) accommodation (3 rules)
- *database*: performs database search (2 rules)
- *store*: stores current SHARED in PRIVATE.TMP (1 rule)

and the update algorithm is shown in (9).

(9)
```
if ( latest_moves = failed )
   then repeat( refill )
   else ( grounding;
          repeat( integrate;
                  accommodate;
                  database );
          if ( latest_speaker = usr )
          then repeat( refill )
          else store
          )
```

If the interpreter failed on the latest utterance, the agenda is refilled (usually with a request_repeat move); otherwise, a *grounding* rule is applied which enters the latest move into SHARED.LM. After grounding, the effects of the move are integrated into the information state, which may require question and/or task accommodation. If necessary, database searches are performed. Finally, if the latest move was performed by the user, the agenda is refilled; otherwise, the current SHARED field is stored in PRIVATE.TMP in case the next optimistic grounding assumption should prove to be wrong.

The current control algorithm in GoDiS simply calls each module in turn in a serial fashion. The selection algorithm simply picks the first applicable rule.

GoDiS has been adapted to the travel agency and autoroute domains, and there are lexicon resources for both English and Swedish. GoDiS is also being used in teaching and student projects, including adaptations to domains such as cinema ticket booking, managing a handheld computer agenda, and interfacing with mobile phones.

https://doi.org/10.1017/S1351324900002539 Published online by Cambridge University Press

4.2 EDIS

The EDIS system (Matheson *et al.* 2000), uses a notion of information state based on Poesio and Traum (1997, 1998), using the record representation used for coding information states in Cooper *et al.* (1999) and Poesio *et al.* (1999). The informational components consist of a common ground part, a semi-public part, and a private part, as with GoDiS. The common part includes four types of information: obligations of dialogue participants to perform actions (OBL), social commitments of participants that propositions hold (SCP), a dialogue history containing acts that have been performed (DH), and conditional statements that will establish obligations or commitments, given the performance of appropriately typed dialogue acts (COND). The semi-public part, analogous to TEMP in GoDiS, are a collection of *discourse units* (DUs) (Traum and Hinkelman 1992), which represent coherent bundles of information that are *grounded* (added to the common ground (Clark and Schaefer 1987)) together. Private information includes the intentions of the agent being modeled. The formal representations of EDIS are shown in (10), where PT-R is a record containing the type of information contained in common ground, shown to the right. Only two DUs are represented, CDU for current, and PDU, the previous one. UDUs is a list of the DUs (which may include the ones identified in PDU and/or CDU) which are not grounded.

(10)
$$\begin{bmatrix} \text{G} & : & \textit{PT-R} \\ \text{CDU} & : & \begin{bmatrix} \text{C} & : & \textit{PT-R} \\ \text{ID} & : & \text{DU-ID} \end{bmatrix} \\ \text{PDU} & : & \begin{bmatrix} \text{C} & : & \textit{PT-R} \\ \text{ID} & : & \text{DU-ID} \end{bmatrix} \\ \text{UDUs} & : & \text{List(DU-ID)} \\ \text{INT} & : & \text{List(Action)} \end{bmatrix}$$

$$\textit{PT-R} : \begin{bmatrix} \text{DH} & : & \text{List(Action)} \\ \text{OBL} & : & \text{List(Action)} \\ \text{SCP} & : & \text{List(Prop)} \\ \text{COND} & : & \text{List(Action)} \end{bmatrix}$$

EDIS uses a modified version of the dialog acts from the DRI coding scheme (Discourse Resource Initiative 1997), giving them precise effect conditions according to aspects of the information state in (9). EDIS uses the same general pipelining of modules as GoDiS, however the update algorithm is a bit different. Whenever a set of dialogue acts are placed in `latest_moves`, the following algorithm is performed, applying a set of update rules in each step.

(11) a. Create a new DU and push it on top of UDUs (and set CDU to this one, first moving the old CDU to PDU).
b. Perform updates on the basis of backwards grounding acts, such as merging PDU.C with G for an **acknowledgement**.
c. If any other type of act is observed, record it in the dialogue history in CDU and apply the update rules for this kind of act.
d. Apply update rules to all parts of the IS which contain newly added acts.

There is also a deliberation step, applied for each system turn, which leads to the system developing new intentions on the basis of obligations, potential obligations that would result from conditions (in the COND field of G or CDU) if an intended act were performed, as well as insufficiently understood dialogue acts and intentions to perform complex acts. Following deliberation, dialogue acts are selected to fulfil

https://doi.org/10.1017/S1351324900002539 Published online by Cambridge University Press

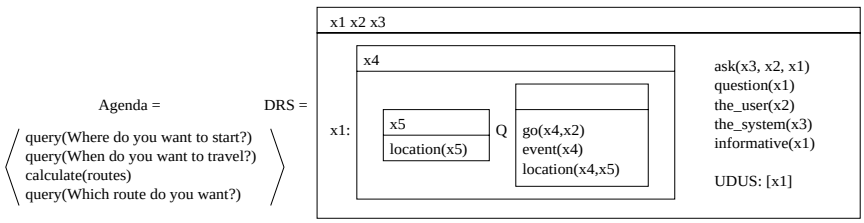
any intentions, and placed in the `next_moves` interface variable, for the generation module to act on.

4.3 Other TrindiKit systems

We briefly mention two other TrindiKit systems. More details on the theories of information state underlying these systems can be found in Traum, Bos, Cooper, Larsson, Lewin, Matheson and Poesio (1999), while details of the systems themselves can be found in Bos *et al.* (2000).

The MIDAS system uses the DRS structures of DRT (Kamp and Reyle 1993) as a major component of its information state. Using a simplified version of the theory proposed in (Poesio and Traum 1998), the root DRS represents the assumed common ground, which will include subordinate DRSes representing events mentioned in the dialogue, as well as DRSes for tracking grounding. Multiple theorem provers are used both for pragmatic aspects of dialogue act interpretation and to implement some of the conditional tests on the update rules. An example of a MIDAS information state can be seen in (12), in which the DRS represents a question the system has asked. This is one of a set of questions, with other future questions remaining on an agenda.

(12)



Another system focuses on conversational game theory, recasting a previous Autoroute system (Lewin 1998) in the current framework. Conversational games are formalized as recursive transition networks. The top-level information state type is a record, as shown in (12). The top level distinguishes a dialogue participant's role as rational agent from that as conversational game-player, with the intuition that the details of the latter should not be major factors in the former. Actions and agenda-items are themselves records with multiple fields.

(13)

$$\left[\begin{array}{lll} \textsc{rational_agent} & : & \left[\begin{array}{lll} \textsc{plan} & : & \textsc{Stack(Action)} \\ \textsc{scoreboard} & : & \textsc{Set(Prop)} \end{array}\right] \\ \textsc{game_player} & : & \left[\begin{array}{lll} \textsc{agendaitem} & : & \textsc{Stack(Agendaitem)} \\ \textsc{currtoken} & : & \textsc{Token} \\ \textsc{alltokens} & : & \textsc{Stack(Set(Prop))} \end{array}\right] \end{array}\right]$$

5 Discussion

We feel the work presented here makes several modest contributions toward an objective of 'best practice' in dialogue systems design. The first is a toolkit for building dialogue managers that can be closer to the level of a theory of dialogue

https://doi.org/10.1017/S1351324900002539 Published online by Cambridge University Press

processing than is available in most systems. There are a number of dialogue system toolkits available (e.g. those reviewed in Bohlin *et al.* 1999b), however these are mostly at the level of wiring together global dialogue states and actions, rather than at the level of considering and acting on specific information in the dialogue context. We hope that a tool such as TrindiKit, and especially the computational view of dialogue modeling as using precisely defined information states and updates can help bridge the currently fairly considerable gap between practical dialogue systems and more theoretical approaches to dialogue.

Second, we hope that the specification of components of information state can help clarify both what is meant by *dialogue management* and the role of different aspects of dialogue management. There are a number of dialogue system functions closely related to dialogue management, though there is a wide difference in particular systems as to how these functions are divided into (conceptual or implementational) modules, and whether, even if there is a module called *dialogue manager*, it will have these functions. For example, in many systems, the key dialogue management task is control of the dialogue, e.g. by tracking dialogue state and performing state-specific functions. In other systems (Allen, Miller, Ringger and Sikorski 1996), there is a central hub controller, and the dialogue manager is one of many 'spokes'. Control can also be distinguished into program control and data-flow aspects. If data flow is too implicit or inaccessible, this may block certain kinds of dialogue management functions, such as timely explicit feedback on the source of problems, when the system has trouble deciding what was said or what to do. Finally, there is a distinction between features of the current state of processing and operations to perform on the basis of that state. The framework presented in this paper, with clear separations between information state, update mechanisms, and control flow can help clarify some of these tricky issues, allowing for more flexible experimentation on a wide range of dimensions.

Indeed, perhaps the main advantage of implementing dialogue systems in a framework like this one is the ability to compare many aspects of such systems, e.g. the different sorts of examples presented in Section 4. While any well-designed system (e.g. Smith and Hipp 1994; Walker *et al.* 1998) should allow a range of experimentation of different parameters, it is much more difficult to compare different systems or approaches to the same phenomena. By specifying different notions of information state (and related updates and algorithms), we can compare, e.g. whether an approach to question-answering is best made with reference to dialogue game grammars, *Questions Under Discussion*, or elements of rational and social agency, such as intentions and obligations (see also Larsson (1998)).

5.1 Next steps

We are currently extending the work presented here in several directions. First, gaining more experience with using the TrindiKit to build dialogue managers, and improving the tools provided to the system designer. Along with this will come evaluation, both of the resulting systems and the utility of the toolkit for building such systems.

https://doi.org/10.1017/S1351324900002539 Published online by Cambridge University Press

Another avenue is visualization tools to better examine the contents of the information state as the dialogue progresses, and allow for easier writing of rules. We are currently exploring integration with the THISTLE visualization tool (Calder 1998). THISTLE is a parameterizable display engine and editor for diagrams, allowing the inclusion of interactive diagrams.

Our aim is that one could approach the task of dialogue system development from any of several different levels, depending on expertise and interest. Pure software engineering effort can be put into developing better versions of the TrindiKit, or similar software, to allow better prototyping of systems and more efficient run-time behavior. Theories of dialogue can be more directly implemented and tested using such a kit, without extensive software engineering or programming abilities required. Finally, given a domain independent theory of dialogue modelling, such as that underying GoDiS or EDIS, the addition of domain and task-specific resources (such as plan libraries and databases) can allow a dialogue developer to customize the system for a particular application, without extensive knowledge of dialogue dynamics.

Acknowledgements

This work was supported by the TRINDI (Task Oriented Instructional Dialogue) project, EU TELEMATICS APPLICATIONS Programme, Language Engineering Project LE4-8314. Other members of the consortium were instrumental in developing the ideas and systems described here. Peter Bohlin and Johan Bos helped develop the Trindikit distribution. GoDiS was developed by the first author and Bohlin, Robin Cooper, Stina Ericsson and Elisabet Engdahl. EDIS was developed by Colin Matheson, Massimo Poesio, and the second author. The MIDAS system was developed by Bos. Ian Lewin developed the SRI conversational game theory system. In addition, all of the above and other TRINDI consortium participants have contributed to the development of the framework presented here. As usual, all mistakes and misrepresentations are due to the authors of this paper.

References

Ahrenberg, L., Dahlbäck, N. and Jönsson, A. (1990) Discourse representation and discourse management for a natural language dialogue system. *Proc. 2nd Nordic Conference on Text Comprehension in Man and Machine*.

Allen, J. F., Miller, B. W., Ringger, E. K. and Sikorski, T. (1996) A robust system for natural spoken dialogue. *Proc. Annual Meeting of the Association for Computational Linguistics (ACL-96)*, pp. 62–70.

Bohlin, P., Cooper, R., Engdahl, E. and Larsson, S. (1999a) Information states and dialogue move engines. *Proc. IJCAI99 Workshop: Knowledge and reasoning in practical dialogue systems*, pp. 25–31.

Bohlin, P., Bos, J., Larsson, S., Lewin, I., Matheson, C. and Milward, D. (1999b) *Survey of existing interactive systems*. Deliverable D1.3, Trindi Project.

Bos, J., Ericsson, S., Larsson, S., Lewin, I., Ljunglf, P. and Matheson, C. (2000) *Dialogue dynamics in restricted dialogue systems*. Technical report D3.2, Trindi.

Bretier, P. and Sadek, M. D. (1996) A rational agent as the kernel of a cooperative spoken dialogue system: Implementing a logical theory of interaction. In: Müller, J. P., Wooldridge,

https://doi.org/10.1017/S1351324900002539 Published online by Cambridge University Press

M. J. and Jennings, N. R. (eds.), *Intelligent Agents iii — Proceedings of the 3rd International Workshop on Agent Theories, Architectures, and Languages (ATAL-96). Lecture Notes in Artificial Intelligence*. Springer-Verlag.

Calder, J. (1998) *Thistle: diagram display engines and editors*. Technical Report HCRC/TR-97, HCRC, University of Edinburgh.

Clark, H. H. and Schaefer, E. F. (1987) Collaborating on contributions to conversation. *Language and Cognitive Processes*, **2**: 1–23.

Cohen, P. (1996) Dialogue modeling. In: Cole, R., Mariani, J., Uszkoreit, H., Zaenen, A. and Zue, V. (eds.), *Survey of the State of the Art of Human Language Technology*. Cambridge University Press.

Cooper, R., Larsson, S., Matheson, C., Poesio, M. and Traum, D. (1999) *Coding instructional dialogue for information states*. Deliverable D1.1, Trindi Project.

Cooper, R. and Larsson, S. (1999) Dialogue moves and information states. In: Bunt, H. C. and Thijsse, E. C. G. (eds.), *Proc. 3rd International Workshop on Computational Semantics*.

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

Ginzburg, J. (1996a) Dynamics and the semantics of dialogue. In: Seligman, J. and Westerståhl, Dag (eds.), *Logic, Language and Computation, vol. 1*. CSLI Publications.

Ginzburg, J. (1996b) Interrogatives: Questions, facts and dialogue. In: Lappin, S. (ed.), *The Handbook of Contemporary Semantic Theory*. Blackwell.

Ginzburg, J. (1998) Clarifying utterances. In: Hulstijn, J. and Niholt, A. (eds.), *Proc. Twente Workshop on the Formal Semantics and Pragmatics of Dialogues*, pp. 11–30.

Kamp, H. and Reyle, U. (1993) *From discourse to logic*. D. Reidel.

Larsson, S. (1998) Questions under discussion and dialogue moves. *Proc. TELT13/Twendial '98: Formal semantics and pragmatics of dialogue*.

Larsson, S., Berman, A., Bos, J., Görnqvist, L., Ljunglöf, P. and Traum, D. (2000) (Trindikit 2.0 manual available from http://www.ling.gu.se/research/projects/trindi/trindikit/documentations.html.

Lewin, I. (1998) *The autoroute dialogue demonstrator*. Technical report CRC-073, SRI Cambridge Computer Science Research Centre.

Lewis, D. K. (1979) Scorekeeping in a language game. *J. Philosophical Logic*, **8**(3): 339–359.

Luz, S. (1999) *State-of-the-art survey of dialogue management tools*. Deliverable D2.7a, Disc Project.

Matheson, C., Poesio, M. and Traum, D. (2000) Modelling grounding and discourse obligations using update rules. *Proc. 1st Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 1–8.

Poesio, M., Cooper, R., Larsson, S., Matheson, C. and Traum, D. (1999) Annotating conversations for information state update. *Proc. Amstelogue'99 Workshop on the Semantics and Pragmatics of Dialogue*.

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

Poesio, M. and Traum, D. R. (1998) Towards an axiomatization of dialogue acts. *Proc. Twendial'98, 13th Twente Workshop on Language Technology: Formal semantics and pragmatics of dialogue*.

Sadek, D. and De Mori, R. (1998) Dialogue systems. In: De Mori, R. (ed.), *Spoken Dialogues with Computers*. Academic Press.

Searle, J. R. (1969) *Speech Acts*. Cambridge University Press.

Smith, R. W. and Hipp, D. R. (1994) *Spoken Natural Language Dialog Systems: A practical approach*. Oxford University Press.

Traum, D., Bos, J., Cooper, R., Larsson, S., Lewin, I., Matheson, C. and Poesio, M. (1999) *A model of dialogue moves and information state revision*. Technical report D2.1, Trindi Project.

https://doi.org/10.1017/S1351324900002539 Published online by Cambridge University Press

Traum, D. R. (1999) 20 questions on dialogue act taxonomies. *Proc. Amstelogue'99 Workshop on the Semantics and Pragmatics of Dialogue.*

Traum, D. R. and Allen, J. F. (1994) Discourse obligations in dialogue processing. *Proc. 32nd Annual Meeting of the Association for Computational Linguistics*, pp. 1–8.

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.

Walker, M. A., Fromer, J. C. and Narayanan, S. (1998) Learning optimal dialogue strategies: A case study of a spoken dialogue agent for email. *Proc. COLING ACL-98.*

https://doi.org/10.1017/S1351324900002539 Published online by Cambridge University Press