PLANNING DIALOGUE CONTRIBUTIONS WITH NEW INFORMATION

KRISTIINA JOKINEN and HIDEKI TANAKA and AKIO YOKOO
ATR Interpreting Telecommunications Research Laboratories
email: {kjokinen|tanakah|ayokoo}@itl.atr.co.jp

Abstract

The paper discusses a framework for planning contributions in a spoken dialogue system, and focuses especially on the three *I*'s: Incrementality, Immediacy, and Interactivity. The emphasis is on communicative principles and the notion of NewInfo, or the information focus of the utterance. NewInfo provides a natural way of to conceptualize the planning process and to generate utterances on the level of granularity required in spoken interaction.

1. INTRODUCTION

The question that we will investigate is the starting point of generation, and we argue that this is NewInfo, the piece of new information exchanged in interaction, with which the mutual context gets updated. This may sound like sliding on the "slippery slope" of (McDonald, 1993), who points out that the answer to the question 'how far back does generation go?' is tied to the proportional amounts of linguistic and contextual information in the specification which serves as the source of generation. However, we would like to stress the separation of communicative knowledge from the minimal information units used as the basis for generation, and that the selection of the information units and the way they are actually communicated are subject to conditions which may require changes in one of the tasks before the other one is properly completed. In (spoken) interaction, utterances consist of fine-grained information units to which the listener immediately reacts by giving feedback or a response, and this feedback then directs the speaker to modify her utterance accordingly. Hence, we might as well slide the slippery slope all the way down, and conclude that generation starts simultaneously with interpretation, as a reaction to the presented information. The initial 'message' is then gradually specified into a linguistic expression with respect to language-specific knowledge and communicative requirements of the situation.

Consequently, in this paper we focus on the three *I*'s in generation: Incrementality, Immediacy, and Interactivity. The research is still ongoing, so we pose questions more than provide answers. After introducing the three *I*'s, the specific questions we will discuss are: (1) What are suitable utterance units for exchanging information in spoken dialogues? (2) What is the relationship between NewInfo and organisation of the task/domain information? (3) What kind of requirements are imposed on the generator?[1]

[1]Throughout the paper we use *generator* to refer to the whole system that generates rather than analyses natural language. The component of the generation system that mostly deals with world-knowledge, tasks, plans and communicative goals, is called *planner*, while *realiser* is the component which concerns lexico-semantic and syntactic information. Sentence planning is also called *micro-planning*.

2. THE THREE *I'* S

Consider the following telephone dialogue, taken from the ATR EMMI corpus (Loken-Kim et al., 1993). A conference participant calls the conference office and asks for information on how to get to the conference center from Kyoto station.

(1) IE: /breath/ (and I wondered if you could) [ah] I'm at Kyoto Station just now and I wondered if you could give me information on how to get to the conference center
A: Certainly [ah] where exactly are you in Kyoto Station
IE: [um] (I) I've just come on the Shinkansen from [ah] Tokyo so I'm just outside the tracks
A: /breath/ OK [um] you're going to want to look right ahead of you and you'll see a {large} (staircase) stairwell
IE: {[umhuh]} [uhuh]
A: you're going to want to go up that set of stairs
IE: [uhuh]
A: walk across a platform
IE: [uhuh]
A: and then down the stairs and out what is referred to as the central exit
IE: nK
A: you'll probly be able to see signs from where you're standing now signs towards the central exit
IE: [ah] yes I see {(ha') hachi} something
A: {(O') OK}
IE: OK
A: OK you wanna follow those signs until you're basically out {of Kyo}to station
IE: {[mhm]} [mhm]

A provides detailed instructions of how IE can get out of Kyoto station, as the first step towards the main goal to get to the conference center. The route is divided into parts, and the pieces of information are given *incrementally* to IE. After each piece of information, A pauses for a while and IE acknowledges the receipt of the information. Later in the dialogue the following occurs:

(2) IE: /ls/ all right that sounds [ah] easy enough (could) I wonder if you could tell me are there any interesting [ah] sights around that area (I) I believe there's a break at the conference at some point and I was just wondering if there's anything interesting to see
A: [ah] yes actually that's a very [um] interesting part of (the city) Kyoto city [um] close to thi conference center is a shrine called Heian shrine and ({it da}tes back) I don't know if you're familiar with Kyoto history
IE: {[hmm]}
A: or not but Kyoto (y') was the former capital of Japan
IE: [uhuh]
A: and in that particular time (instead of being referred to) the city itself was not referred to as Kyoto it was referred to as Heian
IE: [unhu{n]}
A: ({an}d so) which in Japanese has the idea of peace or tranquility
IE: [unh{un]}
A: {so} that shrine is actually very historic in the city itself
IE: [uhuh]
A: and depending on the day that you go there could be various events held
IE: [uhuh]
A: and (it's) it's very easy to locate because you'll see a large orange gate
IE: [uhuh]
A: [um] and (it's) it's a well-known landmark within the city
IE: nK
A: and across the street from thi shrine itself there's a museum {Kyoto} Art Museum
IE: {[uhuh]} +[aha]+

A: +so+ many people are also familiar with that so it's actually quite easy to find your way
IE: nK
A: I would say from the conference center if you're going to walk it would probly be about [um] fifteen minutes
IE: [uhuh]
A: fifteen minutes on foot
IE: [oh] OK sounds very easy to reach

This example has a similar structure to the previous one: A gives information which IE acknowledges. However, there is no clear plan structure that guides A's incremental descriptions. Rather, each piece of information is connected to the previous one via *topical associations:* the name of the shrine associates with the old name of Kyoto and peace and tranquility, while the gate with which the shrine is easy to locate, also serves as a landmark of the whole city. These topical chains, shown in Fig. 2, are quite different from the task structure in Fig. 1 which underlies dialogue 1: A relies on her knowledge of the domain and the different relations between the concepts rather than on a hierarchial plan. This type of generation provides an interesting challenge to NLG systems in general, since it not only requires flexible focus shifting (McCoy and Cheng, 1991; Hovy and McCoy, 1989) but also that the communicative principles governing associative chains are spelled out.

Two unusual interactions, though characteristic of spoken dialogues, also take place in dialogue 2. In the beginning, A starts to provide information about the interesting places to visit around the conference center, but soon realises that a foreign conference participant does not automatically possess knowledge of such historical facts as what period a shrine dates back to. A thus repairs her utterance completely and produces a remark about her ignorance. IE's evasive feedback confirms A's tacit assumptions of IE's scarce knowledge of Kyoto's history, and is embedded inside A's turn. Although A's remark can also function as an indirect request for IE to indicate her knowledge state on the matter, A continues her original utterance as if IE had only provided backchannelling and not taken a full turn. Mutual knowledge has thus been established without an explicit question-answer exchange. This is an example of *immediacy* of reaction: the speakers monitor their own contributions, and closely follow what the partner says. The information exchange is managed locally by presenting new information to the partner who then analyses, evaluates, and reacts to the new information in the current dialogue context (Clark and Schaefer, 1989; Traum, 1994;

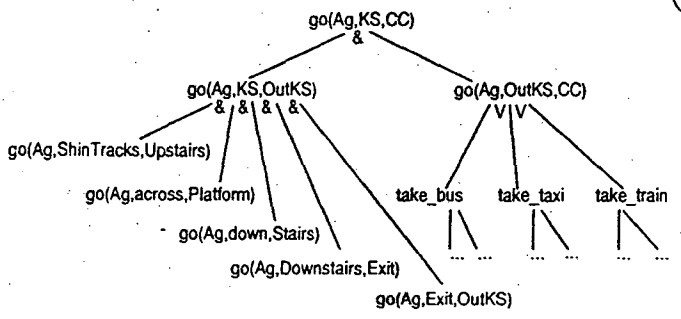

Figure 1: A task structure. & represents conjunctive goals, V alternatives.

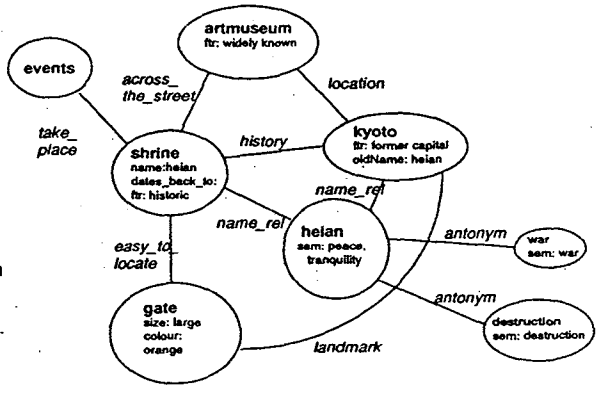

Figure 2: Topic association network.

Jokinen, 1996). However, the listener does not initiate the response only after the speaker has finished speaking, but rather, starts response generation immediately, simultaneously as the speaker speaks. The listener signals her understanding of what is being said by explicit or implicit feedback[2] (Allwood et al., 1992), and she may even co-produce utterance units (Fais, 1994).

The other peculiar interaction occurs at the end of the dialogue. A has volunteered information about how long it takes to walk from the conference center to the museum, and IE acknowledges this. A may have reasons to suspect that IE has not really understood the presented information, so she repeats it to make sure that it is correctly integrated into mutual knowledge. This is an example of the *interactivity* of communication. The whole dialogue of course is already an interactive event, but the subtle point here is the use of repetition as a means of interaction *management*. Since the factual information in a repetition is already known, the relevance of such an utterance arises from the very act of repetition and new information is looked for on a metalevel, i.e. on the level of interaction management. In this example, repetition functions as an effective turn release, signalling to the partner that she needs to confirm the information in a more convincing way.

In NL generation, conversational aspects have been addressed especially in interactive explanation and instruction generation (Cawsey, 1993; Carletta, 1992; Moore, 1995; Inui et al., 1996). In this paper we approach the problem from the view-point of spoken interaction, and outline a reactive response planner which takes into account the speaker's communicative needs and the new information intended to be conveyed, based on the three *I*'s discussed above.

3. INFORMATION STRUCTURE OF UTTERANCES

3.1. Prosodic information units

It is commonly agreed that sentences are not appropriate units for spoken interactions; rather, the object of study is the utterance, variably defined with the help of speech acts, turns, turn-constructional units, and intonational units. In generation, the general control flow goes from conceptual information to the string of words (ultimately: sounds), and the question of a suitable utterance unit gets rephrased as a question of the minimal information unit that constitutes the basis for generation and can deal with the three *I*'s as well as be prosodically identified.

According to Stenström (1994), the speakers' turns are organised into *information units*: each unit has its information focus marked with a nuclear tone, and usually the word with a nuclear tone occurs at the end of the unit. For instance, phrases like WALK *across a* PLATFORM, where capitalization marks the pitch-prominent words, are segmentd into two information units consisting of the two nuclear words. However, from the view-point of generation, we regard the phrase as a single unit, since both 'walking' and 'across a platform' are new information on the discourse level. We thus distinguish between information units which are minimal constinuents on the prosodic level (accented words), and information units which are minimal units on the discourse level (new information). Our NewInfo unit can be prosodically complex, and it thus corresponds to what Pierrehumbert and Hircshberg (1990) call an *intermediate phrase*: it contains one or more accented words and a phrase accent (high or low tone) at the end. One or more intermediate phrases plus a boundary tone then make up an *intonational phrase*, roughly corresponding to an utterance.

[2]Ward (1997) discusses the reflexive nature of backchannels and demonstrates how they can be generated in a highly interactive system relying only on acoustic features like the pitch and the length of pauses.

3.2. NewInfo and Central Concept

To model the information content of utterances, we use the notions *Central Concept (CC)* and *NewInfo (NI)* (Jokinen, 1994; Jokinen and Morimoto, 1997). These notions are related to the linguistic topic-comment structure (what is talked about *vs.* what is said about it) and the focus-ground structure (new *vs.* old information), but defined in terms of discourse referents, i.e. objects used by the planner.[3] They can be *realised* by linguistic phrases: analogously to Grosz et al. (1995), we say that *U realises d* if *U* is a phrase for which *d* is a discourse referent in the context model.

CC is the discourse referent which the utterance is about or which the participants focus their actions on. Given the plan in Fig. 1, at the beginning of the dialogue CC is fixed to the instantiated discourse referent `go(id1,ag1,ks,cc)`, corresponding to the top goal, and can then shift to subgoal instantiations, depending on the planner's action. In the topic network Fig. 2, CC is first the instantiated discourse referent of the node *shrine*, then shifts to *kyoto*, *heian*, etc. The shifting, however, is now constrained by the organisation of domain knowledge and topic associations: the current NI becomes the next CC, and the next NI is one of its salient properties or property values.

CC fixes the view-point from which NI is presented, and its realisation depends on the context: object-type CCs may be realised as pronouns (IT *'s a well-known landmark*), but if recoverable from the context, CC need not be explicitly present at all (*fifteen minutes on foot* has its CC "distance from conference center to shrine" omitted). CC is not necessarily *old information*: for instance, in dialogue 1, when A gives the first step in directing IE out of Kyoto Station (*go up that set of stairs*), CC is `go(id2,ag1,ks,outks)` ("get out of Kyoto Station") which is not mentioned in the context before. On the other hand, this is not NewInfo either, since it is not realised, but *inferrable* (Prince, 1979): expansion of the goal has resulted in the NewInfo `go(id3,ag1,shintracks,upstairs)`, which is realised in the utterance, and from which CC is to be inferred. Conversely, old information need not be CC. For instance, before instructing IE to go up the stairs, A has introduced the stairwell and then refers to it as part of mutually known *background* information: *go up* THAT *set of stairs*. However, A is not talking about the stairs but the way to get out of the station.

NewInfo is the information centre of the utterance, identified as the discourse referent(s) to be presented, but not yet established as part of mutual knowledge. NI is always explicitly realised, with the prominent pitch accent. It is selected on the basis of the fine-grained task structure, if such exists, or the topic associations in the domain, and further specified with, or rather, *wrapped in*, communicatively important information. The wrapping may only contain some morpho-syntactically required specifications, so NI becomes realised via a direct mapping to words (*walk across a platform*) and is prosodically marked as an intermediate phrase. A more complex NI (causal relation, comparision), or a more complex communicative situation (explanation, negotiation), may require more elaborate wrapping so NI becomes realised as a complex intonational phrase.

4. REQUIREMENTS FOR A GENERATOR

We now move on to the requirements that the three *I*s impose on a generator. Assuming that only relevant information is communicated to the partner and that the most relevant information in a given context is the new information, we conclude that the starting point for generation is NewInfo. Furthermore, considering the three *I*s, (1) NewInfo can be gradually specified as needed

[3] A similar distinction is made by Vallduví in terms of *link* and *focus*, but his concern is in cross-linguistic realisation of information packaging, not in dialogue management (Vallduví and Engdahl, 1996).

in incremental generation, (2) it is the unit which the immediate reaction is a reaction to, and (3) its obvious repetition directs the hearer to look for the relevant interpretation on the level of interaction rather than on the level of factual information exchange.

We do not discuss real-time planning, but the reactive nature of the generator is obvious: under time pressure, the planner may want to give the most important part of the message (NewInfo) to the realiser first, then provide further specification as necessary. Since content planning and realisation are theoretically parallel processes (de Smedt and Kempen, 1987), the realiser may thus start saying something immediately after NewInfo has been decided, and produce temporizers (*uhmm, errr*) while waiting for the next piece of information from the planner.

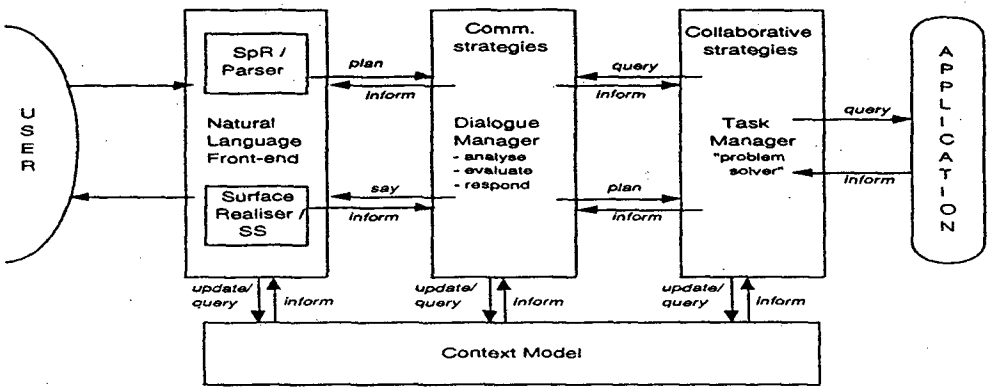

Figure 3: Architecture of a dialogue system. SpR and SS refer to speech recognition and speech synthesis, respectively.

To meet these requirements we consider a highly modular system architecture depicted in Fig. 3. The different components are independent "agents" which operate within their own expertise area but can communicate with each other via a simple "agent communication language" (Fig. 4). Each agent can also query and update the Context Model (CM) which records the ongoing dialogue state. Planning is divided into task and dialogue planning (Cawsey, 1993): the task manager (TM) produces plan recipes in regard to a particular application and the current plan, while the dialogue manager (DM) plans the system's communicative actions. DM can request TM to give a suggestion of what to do next, and TM can query DM of a parameter value. DM processes requests by the parser (PR) to plan a response to a user utterance in the cycle of analysis, evaluation, and response (Jokinen, 1996), and in particular, its planning also concerns content organisation into utterances. Language-specific knowledge is stored in a linguistic lexicon and used by an incremental surface realiser (SR), say of the type described in (Wilcock and Black, 1998). DM requests SR to realise a set of concepts, and the realiser must be capable of producing elliptical and fragmentary utterances. TM and DM operate on concepts defined in the World Model and the mapping from task-related entities to linguistic words[4] is described in the Conceptual Lexicon (not shown in Fig. 3).

The pieces of information communicated to the user depend on the task that gave rise to the communication in the first place. Task information is fine-grained and decomposed into subgoals,

[4]We assume that words can first be underspecified and can then be gradually specified into the final words by the surface realiser, cf. Zock (1993).

```
request(PR,DM,plan,Sem)      (PR requests DM to plan a response to Sem)
request(DM,TM,plan,Goal)     (DM requests TM to plan a recipe for Goal)
request(DM,SR,say,Sem)       (DM requests SR to realise Sem)
request(XX,CM,upd,Param)     (XX requests CM to update context with Param)
request(TM,DM,stat,Cond)     (TM queries DM whether a Cond is true or false)
request(TM,DM,val,Param)     (TM queries DM for a value of Param)
request(XX,CM,val,Param)     (XX queries CM for a value of Param)
inform(DM,PR,Sem,Stat)       (DM informs PR whether a response to Sem is planned or not)
inform(TM,DM,Goal,Plan)      (TM informs DM of Plan for Goal)
inform(SR,DM,Sem,Stat)       (SR informs DM whether Sem is realised or not)
inform(CM,XX,Param,Stat)     (CM informs XX whether update with Param succeeds or not)
inform(DM,TM,Cond,Stat)      (DM informs TM whether Cond is true or false)
inform(DM,TM,Param,Val)      (DM informs TM of Val for Param)
inform(CM,XX,Param,Val)      (CM informs XX of Val for Param)
```

Figure 4: Simple agent communication language.

each of which describes a basic act, or an aggregation of acts, in the full plan. The subgoals may have knowledge preconditions, such as a constraint on mutual knowledge concerning the agent's location, which must be fulfilled before the act can be executed.[5] The concepts describing the content of a plan are linked to a world model hierarchy which describes the ontology of the domain and also plays an important role in tracing topic associations. Each instantiated world model concept is a *discourse referent* in the Context Model; in particular, events and actions are discourse referents. A plan operator corresponding to the top level plan of Fig. 1 is represented as follows:

```
goal:         go(Id,Ag,ks,cc)
constraints:  location(Ag,inKS)
subgoal:      go(Id2,Ag,inKS,outKS)
subgoal:      go(Id3,Ag,outKS,cc)
effects:      location(Ag,cc)
```

The *immediate communicative goal, ICG,* is an intention to realise the current NewInfo, and the *communicative goal, CG,* is a generalisation of ICG, an intention to realise some concepts, not necessarily NI.[6] TM provides DM with NewInfo which can be a plan step, a knowledge precondition, or a concept in the world model. The content of the immediate communicative goal can thus vary from concepts and basic acts to general actions subsuming complex action sequences.

An important question, posed by Inui et al. (1996), is the granularity of the fine-grained units in the plan. There is a need to provide information in units which are suitable for incremental presentation and can function as minimal units for the partner's reaction, but there is also a need for aggregating the fine-grained units into bigger ones to maintain coherence of the dialogue. Similar considerations have been expressed by Hovy and Wanner (1996) on microplanning: one of the tasks of the sentence planner is sentence content delimitation, but so far little computational research has addressed the question of when and how to divide information into distinct sentences. We think that NewInfo is helpful in this respect, since it is a flexible unit: defined as a minimal information unit on a given planning level, it can be of different complexities, thus allowing efficient information exchange, cf. Inui et al. (1996). It can also be further specified ("wrapped"), if inappropriate for

[5] In fact, the first step of the plan in Fig. 1, `go(Ag,KS,OutKS)` can be considered as a plan to get the mutual knowledge precondition `mknow(sys,ag,location(ag,outKS)` of the second step fulfilled.

[6] The speakers normally have several other goals as well, concerning their intentions on the other levels of communication: on a task level the goal is to complete the current task; on a manipulative level the speaker may want to persuade, argue, agree, etc; and on a collaborative level she may have commitments to other (joint) goals.

communication. Dialogue coherence is thus a matter of communicative strategies imposed on the plans and domain knowledge rather than hierarchical organisation of knowledge sources as such.

5. GENERATION AS WRAPPING OF THE NEWINFO

DM decides on the appropriate communicative intention and the presentation of NewInfo, especially the level of explicitness in the utterance, and its content organisation. DM exploits a number of *communicative strategies*, and collects the concepts to be realised in *Agenda*. At each planning stage, DM evaluates Agenda with respect to the strategies, and augments it with relevant concepts as needed. At each stage, it can request SR to realise Agenda, i.e. the system can start "talking".

Agenda is initialised with the NewInfo concepts related to the current ICG (one of DM's own pending goals, or received from TM as a response to DM's "what-next"-request). This means that the simplest realisation for a communicative goal is the realisation of NI concepts. For instance, if IGC is `info_request(location(id1,ag1,X))` and there is no more time to plan further, DM can ask SR to realise this, the result being an elliptical, fragmentary utterance: *Location?/Where?* If the content of Agenda is not valid in the context (and there is more time to plan), DM continues its planning. It may notice that in the current dialogue situation, a complete intonational phrase is desirable, since this would force the partner to take an explicit turn. Moreover, if the NewInfo that user has just presented concerns location, a response with an elliptical question about location would get interpreted on the meta-level, and may, as in dialogue 1, convey false implicatures: *Where?* would most likely be interpreted as a sign of problems in telephone lines, while *Location?* be simply incomprehensible. DM may thus direct SR to produce a sentence instead of other syntactic phrases.

In evaluating the communicative adequacy of Agenda, DM may also notice that Agenda does not directly address the partner's intentions, and NI must be further specified. There are three different cases for NI wrapping. First, the present communicative goal may be ambiguous in its intentions. DM may notice that if it requests SR to realise the goal `info_request(location(id1,ag1,X))` as a full intonational phrase (*Where are you?*), the utterance is not accurate in the context: besides requesting specification of the partner's current location, it can also be understood as a question about her location in general. Since the partner has already said her location is Kyoto Station, the latter interpretation should be blocked, to avoid false implicatures being drawn (interpretation on the metal-level, or simply as being rude). Thus DM specifies the content of the communicative goal with the location information and the goal becomes `info_request(sys,ag1,location(id1,ag1,X), location(id2,X,ks))`. NewInfo is thus "wrapped" into a piece of information that makes the reference point clear (*in Kyoto Station*), probably with the empahsis added (*exactly*).

Second, NewInfo may contain reference to objects which are crucial in fulfilling the task, and so it is cooperative, and sometimes communicatively more efficient, to make sure that the objects are mutually known. For instance, the starting point in instructing IE out of Kyoto Station is the staircase which IE is to go up. If `staircase` appears in the Context Model as an uncertain discourse referent (i.e. it is not known whether the partner knows it or not), DM may introduce the concept via a separate inform-act, before giving the instruction to go up the stairs. In fact, this is what happens in the sample dialogue. NewInfo is `go(id4,ag,shintracks,upstairs)` and wrapped into the goal `inform(sys,ag1,location(staircase,infront))`, which is communicated first (after recursive planning for its suitable realisation).

Third, TM may give DM a conjunctive NewInfo, which consists of several plan steps. For instance,

instead of delivering each of the five steps of how to get out of Kyoto station separately to DM, TM may give them all at once (after reasoning that all the steps are leaf-nodes in the current plan and cannot be expanded). Since each conjunct is an independent NewInfo, DM has a choice of passing the conjuctive goal to SR as such (to be realised as a single, but long conjunction of utterances), or drop each NewInfo separately to SR, with a pause after each item requiring the partner's explicit acknowledgement. The decision is based on the intentions: `describe` would prefer the former, but `instruct` the latter realisation.

Context consideration also affects TM's planning. Collaborative task planning requires that the preconditions of a goal are fulfilled. If the Context Model does not provide necessary information to TM, it can query DM, which would then plan a request to the user, and forward the user's reply back to TM. Playing safe, TM can make sure that the preconditions for each plan step are fulfilled before providing DM with a plan. It can also choose a more risky strategy and provide DM with the plan (conjunctive NewInfo) at once. In this case, DM is responsible for realising the plan as well as monitoring its execution. TM may also keep the control of the plan execution in its own hands, but allow DM to handle (problems with) knowledge preconditions. The choice between the different control strategies is related to system's overall behaviour: communication between TM and DM takes time, and the dialogues become cumbersome if the user's knowledge is constantly queried, but if too much is assumed, backtracking and repairs may be necessary (Carletta, 1992).

Generation of associative topic shifts (dialogue 2) proceeds analogously. However, instead of relying on the decomposition of the task, TM uses a domain model (concept network). Topical associations are based on chaining the current NewInfo as the next CC, and selecting the next NI according to the topic shifting rules described in (McCoy and Cheng, 1991). For instance, the shift to "peace and tranquility" is justified as a shift to the attribute *Sem* of the object *Heian-name*. However, TM does not know whether its associations make communicatively appropriate topic shifts, so suggestions must be filtered by DM. A possible topic shift to "war and destruction" after lingering on "peace and tranquility" of *Heian-name* can thus be rejected by DM, if the shift violates the communicative strategy that says that the distance between the current CC and the main topic (sight-seeing information around the conference center) should not be larger than a given limit. On the other hand, DM does not know whether a topic is exhausted or there might be something more to say about it, so it has to request a new one.

6. CONCLUSION

The paper proposes a framework for planning dialogue contributions and emphasises the three *I*s for generation: Incrementality, Immediacy, and Interactivity. The starting point for generation is NewInfo, the new information intended to be realised, which in the course of planning, gets wrapped with regard to the communicative context and the communicative needs of the speakers. Modularity of the system architecture allows the planners to communicate with each other, and thus realisation can start once NewInfo has been decided. Planning can continue separately, and include pragmatic considerations like those described in (Hovy, 1988). NewInfo is realised via intermediate prosodic phrases which correspond to one or more words with a pitch accent. We envisage that the model also serves as a basis for integrating NLG research into speech synthesis (Black and Campbell, 1995). We continue research on the different constraints and their interaction.

7. ACKNOWLEDGEMENTS

We are grateful to Yasuharu Den, Graham Wilcock and Naoya Arakawa for their comments and discussion.

REFERENCES

J. Allwood, J. Nivre, and E. Ahlsén. 1992. On the semantics and pragmatics of linguistic feedback. *Journal of Semantics*, 9:1–29.

A. W. Black and N. Campbell. 1995. Predicting the intonation of discourse segments from examples in dialogue speech. In *Proc. of the ESCA Workshop on Spoken Dialogue Systems: Theories and Applications*, pp. 197–200.

J. Carletta. 1992. Planning to fail, not failing to plan: Risk-taking and recovery in task-oriented dialogue. In *Proc. of COLING-92*, pp. 896–900.

A. Cawsey. 1993. *Explanation and Interaction.* MIT Press.

H.H. Clark and E.F. Schaefer. 1989. Contributing to discourse. *Cognitive Science*, 13:259–294.

K. deSmedt and G. Kempen. 1987. Incremental sentence production, self-correction and co-ordination. In Kempen, ed, *Natural Language Generation: New Results in Artificial Intelligence, Psychology and Linguistics.* Martinus Nijhoff.

L. Fais. 1994. Conversation as collaboration: some syntactic evidence. *Speech Communication*, 15:231–242.

B. J. Grosz, A. K. Joshi, and S. Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):203–225.

E. H. Hovy and K. F. McCoy. 1989. Focusing your RST: A step toward generating coherent multisentential text. In *Proc. of the 11th Cognitive Science Conference*, pp. 667–674. Ann Arbor.

E. Hovy and L. Wanner. 1996. Managing sentence planning requirements. In *Gaps and Bridges: New directions in Planning and NLG*, pp. 53–58. Proc. of the ECAI'96 Workshop, Budapest.

E. H. Hovy. 1988. *Generating Natural Language Under Pragmatic Constraints.* Lawrence Erlbaum Associates.

K. Inui, A. Sugiyama, T. Takenobu, and H. Tanaka. 1996. Fine-grained incremental and interactive elaboration in explanatory dialogue. In *Gaps and Bridges: New directions in Planning and NLG*, pp. 77–82. Proc. of the ECAI'96 Workshop, Budapest.

K. Jokinen and T. Morimoto. 1997. Topic information and spoken dialogue systems. In *Proc. of the Natural Language Processing Pacific Rim Symposium 1997*, pp. 429–434. Phuket, Thailand.

K. Jokinen. 1994. Coherence and cooperation in dialogue management. In K. Jokinen, ed., *Pragmatics in Dialogue Management*, pp. 97–111. Gothenburg Papers in Theoretical Linguistics 71.

K. Jokinen. 1996. Goal formulation based on communicative principles. In *Proc. of COLING-96*, pp. 598–603.

K. Loken-Kim, F. Yato, K. Kurihara, L. Fais, and R. Furukawa. 1993. EMMI – ATR environment for multi-modal interactions. Technical Report TR-IT-0018, ATR Interpreting Telecommunications Research Laboratories.

K. McCoy and J. Cheng. 1991. Focus of attention: Constraining what can be said next. In C. L. Paris, W. R. Swartout, and W. C. Moore, eds., *Natural Language Generation in Artificial Intelligence and Computational Linguistics*, pp. 103–124. Kluwer Academic Publishers,

D. McDonald. 1993. Does natural language generation start from a specification? In H. Horacek and M. Zock, eds., *New Concepts in Natural Language Generation*, pp. 275–278. Pinter Publishers.

J. D. Moore. 1995. *Participating in Explanatory Dialogues.* MIT Press.

J. Pierrehumbert and J. Hircshberg. 1990. The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, and M. E. Pollack, eds., *Intentions in Communication*, pp. 271–311. MIT Press.

E. Prince. 1979. On the given/new distinction. *CLS*, 15.

A. Stenström. 1994. *An Introduction to Spoken Interaction.* Longman.

D. Traum. 1994. A computational theory of grounding in natural language conversation. Technical Report 545, Department of Computer Science, University of Rochester.

E. Vallduví and E. Engdahl. 1996. The linguistic realization of information packaging. *Linguistics*, 34:459–519.

N. Ward. 1997. A simple rule for the cooperative timing of utterances in spoken dialog. In *Collaboration, Cooperation and Conflict in Dialogue Systems*, pp. 85–90. Proc. of the IJCAI'97 Workshop, Nagoya.

G. Wilcock and W. J. Black. 1998. Incremental generation with Categorial Grammar and HPSG. Ms. University of Manchester Institute of Science and Technology.

M. Zock. 1993. Is content generation a one-shot process or a cyclical activity of gradual refinement? The case of lexical choice. In H. Horacek and M. Zock, eds., *New Concepts in Natural Language Generation.* Pinter Publishers.