Natural Language Processing and User Modeling: Synergies and Limitations

INGRID ZUKERMAN[1] and DIANE LITMAN[2]

[1]*School of Computer Science and Software Engineering, Monash University, Clayton, Victoria 3800, Australia, E-mail: ingrid@csse.monash.edu.au*
[2]*AT&T Labs – Research, Florham Park, New Jersey 07932, U.S.A., E-mail: diane@research.att.com*

Abstract. The fields of user modeling and natural language processing have been closely linked since the early days of user modeling. Natural language systems consult user models in order to improve their understanding of users' requirements and to generate appropriate and relevant responses. At the same time, the information natural language systems obtain from their users is expected to increase the accuracy of their user models. In this paper, we review natural language systems for generation, understanding and dialogue, focusing on the requirements and limitations these systems and user models place on each other. We then propose avenues for future research.

Key words: natural language generation, natural language understanding, plan recognition, surface features, dialogue systems

1. Introduction

One of the main goals of the field of natural language processing is to endow a computer with the ability to interact with people the way people interact with each other. It is both intuitively appealing and widely accepted by the research community that people use some model of their interlocutors when they interact with each other. This model assists them in all aspects of their interaction. For example, it helps them adjust the style and level of generated and accepted language to the style and capabilities of the interlocutor, understand the interlocutor's intentions even if they are not articulated precisely, and generate appropriate responses. In addition, people often update their models of their interlocutors during an interaction or as a result of an interaction.

This use of user models inspired several hopes regarding the advantages user models would bring to natural language systems. User models were expected to improve the ability of natural language systems to understand a user; help achieve adaptivity in natural language interactions; and increase the robustness of natural language systems, so that they could be used by anyone under various circumstances. However, these hopes have been achieved only partially. Research in plan recognition has produced Natural Language Understanding (NLU) systems that can infer a user's intentions even when they have not been articulated precisely; the incorpor-

ation of user models into Natural Language Generation (NLG) systems has yielded systems that adapt their output to users' beliefs and capabilities; and models of users' language usage have improved the robustness of natural language interfaces. However, most of these systems are research prototypes that were developed to test specific ideas, and are applicable only in restricted domains. In addition, few of these systems support fully interactive behaviour, and hence cannot demonstrate the contribution of user models to the entire interaction cycle: understanding a user's requirements, followed by a possible adjustment of the user model, the generation of a response, the understanding of new requirements, and so on. However, the development of such systems is now within our reach, both due to advances in natural language and user modeling, and due to the current emphasis on developing complete practical systems, albeit of limited scope.

The dream of the natural language community is a grand one. However, the reality is that advances are required in many sub-fields of natural language in order to achieve this dream. Parsing techniques must be robust enough to handle ill-formed and incomplete sentences and multimedia input; semantic components must be able to produce an internal representation from a user's input; discourse handling mechanisms must be able to put together the meaning of a piece of discourse; components that handle pragmatics must be able to make inferences that go beyond literal meaning; dialogue systems must be able to handle interruptions, false starts and changes in topic, and recover from mis-communication; and generation systems must be able to produce just the right discourse using appropriate media. These diverse requirements have prompted the separate investigation of different sub-fields of natural language. As a result, the research in user modeling for natural language has also been fragmented, since different aspects of a user model are required for each sub-field.

Interestingly, the fields of NLG and NLU have experienced similar trends regarding their use of user models, in the sense that user models have been considered mainly in relation to pragmatics. Most NLG systems that consult user models do so during content planning, i.e. when deciding what to say (Section 2), while most NLU systems are concerned with building user models that represent a user's plans and goals (Section 3). Only a few natural language systems consult user models in relation to surface features of language (Section 4). Finally, dialogue systems consult user models for a variety of high-level tasks (e.g. providing tailored and cooperative responses to users, or switching the control of the interaction), and also use the dialogue to dynamically update their user models (Section 5). The insights obtained from these natural language systems motivate the challenges and research avenues proposed in Section 6.

2. Natural Language Generation – Content Planning

One of the first NLG systems that consults a user model is described in (Wallis and Shortliffe, 1985). Wallis and Shortliffe recognized the need to model two aspects of a

user in order to generate appropriate explanations: his/her expertise in the subject matter and his/her preferences regarding the level of detail of an explanation. However, their model was rather coarse, consisting of a single number to represent each of these aspects. Their system tailored explanations to the user's level of expertise by omitting simple reasoning steps from explanations generated for an expert user and omitting detail from explanations generated for a novice.

From this start, it became clear that more sophisticated user models were required in order to enable NLG systems to generate appropriate and relevant discourse. Such discourse presents different types of information depending on the audience's perceived expertise or preferences, addresses a user's likely misconceptions and inferences, and takes into account contextual information. We now consider NLG systems that generate discourse that incorporates these features, and discuss the user models that support the requirements of these systems.

2.1. CONSIDERING A USER'S ATTRIBUTES – ONE-DIMENSIONAL USER MODELS

Many NLG systems consult user models which represent a single aspect of a user, e.g. expertise or preferences, where this aspect is influenced by the type of discourse being generated. For instance, systems that generated concept descriptions consulted a model of a user's expertise or interests (e.g. Paris, 1989; Tattersall, 1992; Stock et al., 1993), systems that produced evaluative discourse used a model of a user's preferences (e.g. Jameson, 1989; Carenini and Moore, 1999), and systems that generated health advice consulted a user model which consists of the user's medical record (e.g. Carenini et al., 1994; Binsted et al., 1995).

Paris (1989) observed that a user's level of expertise not only affects the amount of detail provided in an explanation, but also the kind of information given: process traces are normally presented to naive users, while descriptions based on the components of an object are given to expert users. This observation inspired the development of a system that tailors the type of information included in a description to a user's level of expertise. To support such explanations, Paris' user model distinguished between two types of information items in the domain, basic concepts and specific artifacts, recording the items that were known to the user. Although this model was rather coarse, it was sufficient to support the distinction between the two types of explanations under consideration. Tattersall's user model was slightly finer-grained, distinguishing between three levels of expertise a user could have with respect to a concept (Tattersall, 1992). As for Paris' system, this distinction was motivated by the three main explanation strategies considered by Tattersall: *initial*, which describes unknown concepts; *consolidating*, which describes previously introduced concepts; and *reminding*, which describes known concepts. These strategies also determined whether comparisons were optional or mandatory, and whether known or unknown target concepts should be used in comparisons.

The system described in (Stock et al., 1993) consulted a model of a user's interests in order to select information items to be included in multimedia presentations about

Fourteenth Century Italian frescoes. This model represented the relationship between different areas of interest by means of an activation/inhibition network. Each node in the network represented an area of interest, e.g. a painting school or a period of time, and was associated with a set of concepts. The activation of a node, e.g. by the user asking about a particular concept or area of interest, resulted in the activation of nodes in related areas (connected to this node by activation links) and the inhibition of nodes in incompatible areas (connected to this node by inhibitory links).

Jameson (1989) and Carenini and Moore (1999) developed mechanisms which generate discourse that supports an evaluative process. Carenini and Moore's system generated discourse that evaluates an object by consulting a model of the user's preferences represented by means of a multi-attribute value function. In contrast, the discourse generated by Jameson's system elicited the hearer's evaluative judgment implicitly, i.e. without expressing any value judgments in the discourse. This was done by consulting a model of the hearer's "evaluation standards" (i.e. preferences) in order to determine whether to include or exclude a piece of information or whether to be ambiguous or precise. For example, if an interviewer asked about the family situation of a job applicant, and the applicant thought her interviewer had a negative bias against people who are in a relationship, the applicant may choose to say "I live alone" instead of the more accurate "I have a boyfriend".*

The observation that conveying appropriate information to patients reduces the overall cost of health care has created a demand for health documents tailored to patients' needs and abilities. Several researchers developed systems that generate medical advice (Carenini et al., 1994; Binsted et al., 1995; Hirst et al., 1997; de Rosis et al., 1999).** Carenini et al. (1994) developed a system that advised patients about migraines and their treatment. A user model for each patient was obtained from an electronic questionnaire, which collected information about the patient's symptoms, habits, family history and medications taken. This supported the tailoring of the discourse to the patient's particular circumstances. In addition, the system modeled class concerns of migraine sufferers, e.g. the fear that migraines are a life threatening condition. However, the consideration of these concerns was hard-coded into the discourse planning system. As for Carenini et al.'s system, the explanations generated by Binsted et al.'s system mixed general medical information with specific information about a patient. Binsted et al. evaluated the quality of the generated text and the contribution of the personalization aspect by showing their system's output to health professionals. The results of this evaluation were encouraging. However, in order to improve system adaptivity, class concerns should be explicitly incorporated in user models.

---

* For a discussion on acquiring and addressing preferences in a dialogue, see Section 5.4.
** The research of Hirst et al. and de Rosis et al. is described in Section 2.3.

2.2. CONSIDERING INFERENTIAL PATTERNS – ENHANCING USER MODELS

Many NLG systems (e.g. Paris, 1989; Tattersall, 1992) laboured under the implicit assumption that a user's beliefs are a subset of the system's beliefs. In addition, as indicated above, users' beliefs were often modeled at a rather coarse level of detail (which was adequate for the generation phenomena considered by these systems). However, this level of detail does not support the generation of discourse that addresses a user's misconceptions or inferences.

McCoy (1989) and Milosavljevic (1997) considered the generation of explanations that take into account a user's perceptions regarding similarities between objects. Milosavljevic's system generated different types of comparisons, e.g. to distinguish between a target concept and a potential confusor, to convey a concept in terms of a similar concept, or to convey an attribute of a concept in terms of the same attribute of another concept. McCoy's system generated discourse that addresses two types of object-related misconceptions: misconceptions due to the misclassification of an object, and misconceptions due to the assignment of attribute values to an object which differ from those in the system's domain model. Both systems required a more detailed user model than that required by the systems mentioned above, as a user's beliefs regarding object attributes had to be explicitly represented. Further, in McCoy's system, the object taxonomy in the user model, the attributes associated with the objects in the user model and the values of these attributes were allowed to differ from those in the system's model of the world (this system is discussed further in Section 2.3).

Several researchers developed discourse planning systems which take into consideration inferences from the discourse, (e.g. Joshi et al., 1984; Lascarides and Oberlander, 1992; Mehl, 1994). However, the research described in (Zukerman and McConachy, 1993, 1995, 2001; Kashihara et al., 1995; Kashihara et al., 1996; Zukerman et al.,1996; Horacek, 1997) is of particular interest, since these researchers distinguish between the inferences drawn by different types of users.

Zukerman and McConachy (1993, 2001), Zukerman et al. (1996) and Horacek (1997) use inference rules to model a user's inferences in systems that generate explanations. Zukerman and McConachy's system generates concept descriptions that address anticipated erroneous inferences and omit easily inferred information, and both Zukerman et al.'s system and Horacek's generated arguments. However, Zukerman et al. focused on tailoring an argument to the user's capabilities, while Horacek considered the omission of easily inferred information. These researchers modeled different types of inferences and used different discourse planning mechanisms. Zukerman et al. (1996) modeled users' domain-specific inferences by attributing different degrees of belief in domain-specific inference rules to different types of users. These rules were used to build candidate arguments, and preference rules were then applied to select the argument which best suited a particular type of user. Horacek (1997) modeled domain-specific and problem-solving inference rules (attributed to different user stereotypes), and contextual inference

rules (attributed to all users). He used these inference rules to annotate the output of a goal-based discourse planner, and applied preference rules to determine which information could be omitted on the basis of these inferences. Finally, Zukerman and McConachy (1993, 2001) maintain one set of inference rules applicable to a concept hierarchy, whose effect is moderated by the type of the user (as done by Zukerman et al.). They use an optimization-based discourse planner which determines the most concise combination of propositions to be presented, while considering the inferential effect of these propositions. Zukerman and McConachy (1995) described an extension of this discourse planner which took into account the system's uncertainty regarding the stereotypical model to which a user belongs, and used a constraint-based mechanism to generate discourse that avoids boredom and cognitive overload while achieving as much of the communicative goal as possible.

Kashihara et al. (1995) considered a student's inferences for the generation of explanations which impose a cognitive load that is optimal for knowledge acquisition.* According to Kashihara et al., such a cognitive load is achieved by omitting information from explanations. This may be information a student must recall to understand an explanation, or information the student requires to build a mental structure from an explanation. During discourse planning, their system activated inferences to determine the level and type of cognitive load imposed by a candidate explanation, and adjusted the explanation when this level was deemed excessive. These inferences are similar to those described by Zukerman and McConachy (1993), in the sense that they are applicable to a concept hierarchy and their effect is moderated by factors that reflect a student's capabilities. However, unlike Zukerman and McConachy's system, Kashihara et al.'s system learned the values of these factors from interactions with each student. Kashihara et al.'s system was later extended so that it can select follow-up conversational actions, e.g. clarification questions or directives that restrict the topic of discussion, in order to ease a student's perceived cognitive load (Kashihara et al., 1996).

2.3. IMPROVING RELEVANCE AND APPROPRIATENESS – MULTI-DIMENSIONAL USER MODELS

The discourse planning systems described in Section 2.1 consulted user models which represented one aspect of a user, e.g. beliefs, preferences or medical record. The relevance and appropriateness of the generated discourse can be improved by integrating in a user model several factors, such as beliefs, contextual information and preferences.

In McCoy's system (McCoy, 1989), contextual information took the form of an object perspective, which highlighted the attributes in the model of a user's beliefs that were most relevant to the user's topic of discussion. This perspective assisted in the identification of the likely source of a user's misconception. Zukerman et

---

* See Kay's article in this issue (Kay, 2001) for a discussion of student modeling.

al. (1998) modeled a similar aspect of discourse context – focus of attention – in a system that generated arguments. This was done by spreading activation from recently mentioned propositions through a semantic network that was part of the user model (such a network was also part of the "system model"). The propositions activated in this manner constituted focal points around which the search for an argument was conducted (this system is discussed further in Section 2.5).

The systems described in (van Beek, 1987; Sarner and Carberry, 1992) consulted a multi-dimensional user model to generate extended responses to a user's plan-based queries. The responses generated by van Beek's system pointed out incompatibilities between a user's plan and his/her goals or preferences, and provided alternatives to the user's sub-optimal plans. These responses were generated by consulting a user model that represented contextual information (in the form of a user's immediate plans, goals and preferences); the user's background (e.g. the degree in which a student is enrolled), and default goals and plans (e.g. avoiding failure). In addition to contextual information in the form of a user's plans and goals, Sarner and Carberry's multifaceted user model represented the user's beliefs and the user's stylistic preferences for different types of rhetorical predicates. The combination of the user's beliefs with the contextual information enabled Sarner and Carberry's system to evaluate the potential usefulness of different candidate propositions to the user; the most useful propositions were then included in the system's reply.

The system described in Hovy (1988) used a multi-dimensional user model that represented a hearer's level of expertise, emotional state, interest in the topic of discussion, and opinion regarding this topic to generate descriptions of events. This model affected decisions regarding different content planning aspects, such as the level of detail of a description, which depended on the hearer's expertise and interest; and the partiality of the description, which was influenced by the hearer's opinion about the topic.

Two systems that generate health advice also consulted multi-dimensional user models (Hirst et al., 1997; de Rosis et al., 1999). Hirst et al. (1997) developed an authoring tool for use by technical writers, which produced a fluent rendering of the information selected by these writers. In addition to basic information from a patient's medical record, their user model contained information about a patient's attitude to health care, e.g. locus of control and desire to read technical detail, which normally is not part of a medical record. This model afforded a degree of explanation customization that could not be achieved with user models that consisted of a patient's medical record only (Carenini et al., 1994; Binsted et al., 1995).

de Rosis et al.'s system, which generated explanations about drug prescriptions, maintained two user models, direct and indirect (de Rosis et al., 1999). The direct user model represented doctors, who interacted with the system and had the 'final word' regarding which information was presented. The indirect model represented nurses or patients, who were the recipients of the final explanations and did not interact with the system. The patient models contained stereotypical information as well as information extracted from the medical record of individual patients, while

the models of doctors and nurses contained only stereotypical information. The stereotypical model of direct users represented their propensity to discuss certain topics and their stylistic preference for conciseness or verboseness, while the stereotypical model of indirect users represented their interest in certain topics plus attributes which influence their discourse understanding ability. This information affected both content planning (i.e. deciding what to say) and surface generation (i.e. deciding how to present the information). During content planning, an operation mode given as input to the system determined the extent to which the discourse should reflect the propensities of direct users and the interests of indirect users. During surface generation, the text was tailored to the stylistic preferences of direct users and the comprehension capabilities of indirect users (Section 4).

Finally, multi-dimensional user models play an important role in multimedia interfaces. The input and output modalities of these interfaces may include linguistic modalities, such as text and speech, and non-linguistic modalities, such as graphics, animations and pointing. Thus, in addition to making decisions about the content of a presentation, multimedia interfaces must determine its modality, which is affected by users' preferences and interests. The systems described in (Bonarini, 1993; Chin et al., 1994) consulted a multi-dimensional user model when planning multimedia presentations. Bonarini's artificial co-pilot consulted a model of a driver's beliefs and goals to decide which advice to give to the driver, and a model of the driver's psychological states (attention, agitation, irritation and tiredness), attitudes and preferences to determine the modality of this advice (speech, text, map or icon). An interesting feature of Bonarini's system is that the model of the psychological states was inferred mainly from sensor input, e.g. pressing pedals or beeping the horn. Chin et al.'s MC (Maintenance Consultant) system used a multi-dimensional user model to perform different tasks in a multimedia interface. Their model represented static information, such as a user's job type, skill level and security access level; dynamic information regarding the user's programming expertise and display preferences (inferred from the interaction with the user); contextual information pertaining to the current task; and the conversational and visual context shared by the user and the system. Skill level and security access level helped the system interpret the user's requests, e.g. ''Can I do X?'' could be a direct speech act that asks whether the user has permission to do X, or an indirect speech act that requests the system to perform X; contextual task information enabled the system to guide inexperienced users; programming expertise was used to decide whether to volunteer explanations about the system's tools; the modality of the presentation was determined according to the user's display preferences; and the conversational context was used to parse the user's utterances.

2.4. REACTING TO THE USER'S FEEDBACK – TOWARDS INTERACTIVE SYSTEMS

All the systems described so far carried the burden of generating appropriate discourse, and most of these systems (except the system developed by Kashihara

et al., 1996) generated 'one-shot' explanations, and assumed that the user model is obtained from another component. In fact, Kashihara et al. foreshadowed questions that must be answered as systems move towards more extended interactions with users: (1) which user modeling information can we realistically expect to obtain from interactions with a user? and (2) how appropriate is the discourse that can be generated on the basis of this information?

In this section, we consider five NLG systems which adopt different approaches to answering these questions. These systems used users' responses to gather user modeling information required to plan their presentation.* Two of these systems focused on planning the content of their explanations (Moore and Paris, 1992; Peter and Rösner, 1994), while the remaining systems planned both the content and the type of their contribution (Cawsey, 1990, 1993; Shifroni and Shanon, 1992; Asami et al., 1996).

Moore and Paris (1992) adopted the opposite approach to that of the systems described above, rejecting the reliance of NLG systems on complete and accurate user models, and arguing that a system's ability to react to a user's feedback compensates for the lack of reliability of user models. This approach effectively placed on the user the majority of the burden of obtaining an acceptable explanation. However, this burden was eased by the ability of Moore and Paris' system to identify a user's requirements from vaguely articulated follow-up questions. Such an ability was afforded by a goal-based discourse planner, which yielded sufficient contextual information to support the interpretation of such queries. It is important to note that Moore and Paris' system still consulted a user model during discourse planning (this model was obtained from stereotypes, interactions with the user, and observable artifacts such as the user's program). However, little emphasis was placed on the accuracy of this model, as the system made only limited attempts to update it as the dialogue progressed.

In contrast, both Cawsey (1990, 1993) and Peter and Rösner (1994) stressed the need to maintain user models that are as complete and accurate as possible, while taking into account that these models may be flawed. Both systems used *double stereotypes* (Chin, 1989) to determine the content of the discourse (double stereotypes represent the relation between a user's level of expertise and the difficulty of the concepts in the system's knowledge base). In Cawsey's system, the user model influenced the content of the explanation and the selection of rhetorical devices, as well as the style of the interaction (by asking the user questions when the user model did not provide sufficient information for content planning). In contrast, in Peter and Rösner's system, the user model influenced mainly the content of the explanation.

Both systems used explicit and implicit user model acquisition. However, they differed in their approach to perform these tasks. Cawsey's system initially consulted a stereotypical user model, and updated this model implicitly from a user's

---

* Dialogue systems that focused on analyzing users' responses to build up a user model are described in Section 5.3.

clarification questions and acknowledgments, and explicitly from the user's answers to the system's questions. Implicit user model acquisition was also performed through indirect inferences drawn from information already present in the user model. However, these inferences were dynamically activated only when a piece of information about the user was needed for discourse planning, and their result was not recorded in the user model. This obviated the need for procedures that handled inconsistencies in the user model. Peter and Rösner's system performed explicit user model acquisition briefly at the beginning of an interaction; during the remainder of the interaction, it performed only implicit acquisition from the user's clarification questions. In addition, Peter and Rösner's rules of inference modified the user model, which in turn mandated a mechanism for resolving inconsistencies. Shifroni and Shanon (1992) refined the distinction between implicit and explicit user-model acquisition in that they postulated a spectrum of model-acquisition actions which may be part of a generated explanation. Specifically, their system determined which instruction/acquisition strategy (*explanation, question, explicit assumption* or *implicit assumption*) should be chosen in light of the information in the user model. The system then made inferences from the user's response to the presented text. These inferences updated the (stereotypical) model to which the user was assigned, which in turn affected subsequent parts of the explanation.

Asami et al. (1996) adopted a similar view to that of Shifroni and Shanon, in the sense that they considered both questions and explanations as strategies which help convey information to a student. Specifically, Asami et al.'s tutoring system consulted a student model representing causal understanding of physical systems in order to determine the following aspects of the system's dialogue contribution: the topic to be discussed, the focus of the discussion (identifying the cause of an observation or predicting the effect of an action), the level of granularity of the contribution (detailed or outline), and the style of the system's utterance (question, explanation or alteration of the problem). The student model was in turn updated on the basis of the student's answers.

Finally, the two multimedia systems described in Section 2.3 also acquired some user modeling information from their interaction with a user. Bonarini's system (Bonarini, 1993) inferred a user's psychological states from the user's actions, and updated its model of a user's beliefs and goals. Chin et al.'s system (Chin et al., 1994) inferred the user's expertise, display preferences and task (an initial conjecture regarding the user's expertise was made using double stereotypes, Chin, 1989).

The above discussion enables us to propose some answers to the questions posed at the beginning of this section. Regarding Question (1), we see that most of the systems discussed in this section (except the systems described in Bonarini, 1993; Chin et al., 1994) made inferences regarding one user modeling dimension only: his/her beliefs. The reply to Question (2) follows from this answer. The discourse generated by these systems is similar to that generated by early content planning systems (e.g. Paris, 1989; Tattersall, 1992), in the sense that this discourse did not take into account

contextual information, a user's interests and preferences, or his/her inferences, since these aspects of a user model were not automatically obtained from the interaction with the user.* As indicated above, Bonarini and Chin et al. point the way towards systems that integrate the acquisition and use of multi-dimensional user models. Such systems would acquire a model of a user's beliefs using the acquisition techniques described above and in Section 5.3, and infer contextual information and preferences using plan recognition techniques (Section 3 and Section 5.4). However, the acquisition of information regarding a user's inferential patterns is a topic for future research.

2.5. USING BAYESIAN NETWORKS TO MAKE ENHANCED USER MODELS FEASIBLE IN INTERACTIVE SYSTEMS

Bayesian networks (BNs) (Pearl, 1988) have been used for a variety of user modeling tasks (Jameson, 1996). They offer a potential solution to the problem of maintaining enhanced user models, as they represent both beliefs and inferences. BNs are directed acyclic graphs, where each node represents a belief in the value of a variable. Each node is associated with a *conditional probability table* which represents the effect of other nodes on the probability of each possible value of this node. BNs can be used both to make predictions and to analyze outcomes. In the context of user models for discourse planning systems, this means that a BN can be used to anticipate a user's beliefs from planned discourse, and that its assumptions about the user can be updated according to the user's response. This avoids problems of inconsistency in user models.

The systems described in (Zukerman et al., 1998; Jitnah et al., 2000) used BNs to generate arguments and rebuttals. Zukerman et al. (1998) used one BN to represent normative beliefs and inferences (i.e. the system's beliefs) and another BN to represent the user's beliefs and inferences in a system that generated arguments. Bayesian propagation was used during content planning in order to determine the effect of a planned argument on the system's and the user's beliefs. This distinction between normative beliefs and the user's beliefs supported the generation of arguments that balance normative correctness and persuasiveness. In follow-on work, the system described in (Jitnah et al., 2000) generated rebuttals to rejoinders posed by a user to the system's arguments. This was done by consulting the user model BN and the normative model BN, as well as contextual information in the form of a line of reasoning inferred from the user's rejoinder (Zukerman et al., 2000).

These systems (and others described in Jameson, 1996) point the way towards interactive systems that have two important features: (1) they take advantage of the ability of BNs to represent beliefs and inferences in order to maintain enhanced

---

* Cawsey's system identified and corrected erroneous beliefs, but it lacked the user modeling information to determine the source of these beliefs, as done in McCoy's system (McCoy, 1989).

user models; and (2) they use the predictive and analytical capabilities of BNs to automatically maintain consistent and accurate user models. Although this solves some of the problems pointed out earlier, there are questions that must be answered in order to fully support enhanced user models for interactive systems: (1) how do we determine the structure of a BN and the conditional probability tables? (2) can we adapt the structure of a network and the conditional probability tables to accommodate individual users and changes in users? At present, the structure of BNs is mainly hand-crafted, but the conditional probability tables can be automatically obtained from the observed behaviour of users. However, owing to the large amounts of data required to learn these tables, they are currently obtained from populations of users, rather than from individual users (Zukerman and Albrecht, 2001). The second question and the related question of learning the structure of BNs are currently being considered by the machine learning and uncertainty in Artificial Intelligence communities (Friedman and Goldszmidt, 1999).

3. Natural Language Understanding – Plan Recognition

As shown in the previous section, if a natural language system is to tailor its responses to a particular user, the system needs to have some sort of model of the user. One type of information that has often been included in such user models is knowledge about a user's plans and goals. Plan recognition in natural language is concerned with constructing a model of a user's plans and goals during the course of understanding the user's utterances. The system can then use this model to both better understand subsequent utterances, and to generate more cooperative and helpful responses.

More generally, plan recognition is an active research area in Artificial Intelligence, as well as a promising approach for handling many problems that arise in the area of natural language pragmatics. In fact, the complex problems faced by natural language processing systems have led to many significant advances in the field of plan recognition. Underlying the plan-based approach to natural language is a view of human communication as rational action. Speakers are assumed to have goals, which they are attempting to achieve via plans containing communicative actions (e.g. utterances) and possibly other types of actions. Plan recognition, in turn, is the process of inferring such goals and plans from a speaker's utterances. Here, we only briefly note some of the natural language understanding tasks that have been handled using plan recognition. Details regarding both plan recognition techniques and motivating applications (both within and outside the area of natural language processing) can be found in Carberry (2001) in this issue.

The ability to reason about plans has been used in diverse areas of natural language processing, such as understanding stories, understanding the intentions behind a user's utterances, providing new methods for explaining surface linguistic phenomena, and handling issues arising in extended dialogue. In the area of story understanding, Wilensky (1983) used plan-based inferences to explain the actions

in a story, while Charniak and Goldman (1993) assembled BNs from a plan-based representation in order to handle the probabilistic aspects of story understanding. In question-answering and dialogue systems, recognition of the intentions behind a user's utterances has enabled systems to generate a wide range of cooperative responses. For example, by inferring an underlying plan motivating a user's utterances, a system can anticipate obstacles that might prevent the user from successfully executing his/her plan, and provide augmented responses to remove these obstacles (Allen and Perrault, 1980). In addition to using plan recognition to supply more information than explicitly requested (Allen and Perrault, 1980), plan recognition has formed the basis of systems that understand indirect speech acts (Perrault and Allen, 1980), respond to ill-formed queries (Carberry, 1988), detect and correct misconceptions (Quilici, 1989), handle queries based on invalid plans (Pollack, 1990), and recognize complex discourse acts such as expressions of doubt (Carberry and Lambert, 1999). Plan-based reasoning has also provided new methodologies for handling traditional linguistic phenomena, such as resolving referring expressions (Grosz, 1977) and inter-sentential ellipsis (Carberry, 1985; Litman, 1986). Furthermore, extending plan recognition to track context and to recognize plans at multiple levels has proved to be effective for reasoning about issues arising in extended dialogue: understanding sub-dialogues entered into to clarify or correct plans (Litman and Allen, 1987), obtain information required to execute a plan (Lochbaum, 1995), or negotiate conflicting beliefs (Carberry and Lambert, 1999); inferring an agent's plan incrementally as a dialogue progresses (Carberry, 1990); and inferring communicative goals that are conveyed incrementally rather than in a single utterance (Lambert and Carberry, 1991). Finally, plan recognition systems that explicitly handle uncertainty are capable of tracking and evaluating multiple alternatives in order to recognize intentions that are revised over multiple utterances (Carberry, 1990; Raskutti and Zukerman, 1991; Charniak and Goldman, 1993).

4. Considering Surface Features of Language

NLU systems that consult user models in relation to surface features have focused on two main tasks: (1) assistance with second language acquisition (Schuster, 1985; McCoy et al., 1996), and (2) adaptation of natural language interfaces (Fain Lehman and Carbonell, 1989; Allen and Bryant, 1996). Both types of systems consulted user models which encoded deviations of the user's language from the system's language. Schuster's system, which covered only verbs and prepositions, represented these deviations implicitly in a grammar of Spanish – the native language of the users being modeled (Schuster, 1985). In contrast, Fain Lehman and Carbonell (1989), Allen and Bryant (1996) and McCoy et al. (1996), whose systems covered a variety of linguistic phenomena, represented these deviations by means of rules that performed dynamic modifications to correct grammar productions. McCoy et al. used a large set of mal-rules (Sleeman, 1984) which represented common errors

performed by deaf students of written English whose native language was American Sign Language. Fain Lehman and Carbonell used a set of four devices (insertion, deletion, substitution and transposition of terms in correct grammar productions) to model users' possible departures from standard English when interacting with a natural language interface. Following in Fain Lehman and Carbonell's footsteps, Allen and Bryant investigated the use of a unification-based formalism to represent the linguistic knowledge that supports an adaptive parser, instead of Fain Lehman and Carbonell's case-based formalism. This led to the identification of additional opportunities for language adaptation.

Both McCoy et al. (1996) and Fain Lehman and Carbonell (1989) enhanced their user models with additional information that moderated the applicability of different deviations from a normative grammar. These enhancements were influenced by the task performed by their system and by the system's operating conditions. McCoy et al.'s system corrected a student's grammar mistakes, which were expected to diminish over time. Their system maintained a feature-based model of English, where the ordering of the features was tailored to a user's characteristics (e.g. the user's native language). This model together with the mal-rules (each of which was applicable to only a few features) supported the identification of language features where the user was likely to make mistakes. In contrast, Fain Lehman and Carbonell's system dynamically adapted the grammar accepted by a natural language interface in order to facilitate a user's interaction with the interface. This was done by learning 'ungrammatical' language patterns that reflected the user's grammatical style, and using these patterns when parsing the user's subsequent input.

The NLG systems described in Bateman and Paris (1989) and de Rosis et al. (1999) consulted a user model during surface generation. Bateman and Paris defined *registers* (Halliday, 1978) which specified language features employed by different types of users (e.g. system developers and end users). By using these registers, their system generated different phrasings for the same propositional content according to the type of the target audience. In de Rosis et al.'s system, the verbalization of individual sentences was tailored to the comprehension capabilities of a user, which were determined from the user's age and level of instruction. A future avenue of research which parallels the developments in NLU systems involves the automatic adaptation of the vocabulary presented by a reading tutor as a student's reading ability improves (a non-adaptive reading tutor is described in Mostow and Aist, 1997). In addition, as indicated in Marcu (1996), a promising area of investigation, in particular in the area of health education, pertains to tailoring the verbalization of discourse according to its persuasiveness for different types of users.

5. Dialogue Systems

While the majority of the earliest user modeling research was in the area of dialogue systems (e.g. Kass and Finin, 1988; Kobsa and Wahlster, 1989), research in user

modeling and dialogue has both dramatically dropped off over the years, and changed greatly in character. The initial dialogue systems were typically text-based, where a user was expected to ask a question or enter some input. The system would then use sophisticated user modeling capabilities to tailor its responses to individual users, to provide more cooperative responses, to correct or prevent user misconceptions, etc. The dimensions modeled by early user models included goals/plans,★ capabilities, attitudes, and beliefs regarding the domain, the world and other agents (although many systems focused on one dimension). Often, the dialogue itself was used to incrementally acquire or update the user model, both via system inferences on user utterances and via system queries to the user. Unfortunately, most of these early systems typically worked on only a few carefully hand-crafted examples. Furthermore, while the adaptations performed by these systems were intuitive, the utility of these adaptations was never seriously evaluated. This type of system has largely disappeared in recent years. Recent research has instead emphasized the development of often shallower but more robust systems, where the user models are more empirically motivated and sometimes even acquired automatically, the logic-based knowledge representations are less expressive but more tractable, and where new representations that are robust to uncertainty are used. In addition, speech is increasingly supported, as are experimental and quantitative evaluations of system performance.

5.1. KNOWLEDGE REPRESENTATION AND REASONING – PROVIDING INFERENTIAL SUPPORT FOR DIALOGUE

There has been a long tradition of modeling the participants in a dialogue as rational agents who perform plan recognition (see Carberry, 2001 in this issue) and other types of rational inferences. As a result, much of the work in the field of user modeling for dialogue has been concerned with developing representational and reasoning mechanisms to support such inferential approaches.

Initially, the field was dominated by knowledge-based formalisms such as plans (Section 3) and various types of logics. Hustadt (1994) argued that modal logic, which is a very expressive representation, is needed to deal with the important notions of belief and desire; he then presented a modal logic for supporting stereotype-based user modeling in mixed-initiative dialogue. His work was motivated by a previous proposal to use modal logic, again with the goal of representing and reasoning about agents using notions such as belief, intention and argument (Allgayer et al., 1992). Dols and van der Sloot (1992) presented a belief-based formalization of communicative conventions to support the acquisition of new beliefs as a dialogue progressed. Models of an agent's beliefs (both about him/herself and the conversational partner) have also often been represented in languages derived from epistemic logic (Moore, 1980; Konolige, 1986) (see

---

★ See Section 3 for a discussion of plan recognition applications to natural language.

Hintikka, 1962 for a discussion on epistemic logic). However, Taylor, Carletta and Mellish (1996) argued that such complex belief models supporting unlimited nesting of beliefs are in fact unnecessary if dialogue participants are cooperative. They proposed a simpler belief representation for such cases, and argued that besides being more computationally tractable, their belief model yields more human-like referring and repair strategies.

Increasingly, probabilistic representation systems have been used to support inference in dialogue systems. Jameson et al. (1995) showed how multi-attribute utility theory and BNs can provide a unifying framework for dialogue (and other) systems that perform adaptive evaluation-oriented information provision. The users of such systems had the goal of making evaluative judgments, while the system provided the users with the information needed to make these judgments. Such an approach integrates dialogue tasks such as move planning within a general framework of quantitative user modeling. More recently, Horvitz and Paek (1999) illustrated the use of Bayesian user models for a variety of conversational tasks. In particular, they described a Bayesian representation, inference strategies and control procedures that can be used to infer speaker goals from linguistic and other inputs, to control question asking, and to control dialogue flow. As discussed below, Berthold and Jameson (1999) used a BN to assess the cognitive load of a user, and Chu-Carroll and Brown (1998) used Dempster–Shafer theory for modeling mixed initiative.

5.2. Mixed-initiative dialogue – allowing for flexible control of an interaction

In collaborative situations, expertise is often distributed among multiple agents. As a result, the agent in control (i.e. the agent with the initiative) often switches back and forth to other agents. Similarly, a system that can participate in mixed-initiative dialogue must be able to recognize when to take control of the interaction, and when to relinquish control to the user. Thus, from a user modeling perspective, one way a system can tailor a dialogue is to adapt the level of initiative that the system takes in response to the user's needs and preferences. However, as will become evident from the discussion below, there have been many different uses of the term "mixed initiative" in the literature. Cohen et al. (1998) presented a useful survey and synthesis of existing approaches.

Chu-Carroll and Brown (1998) argued that in a collaborative problem-solving environment, it is necessary to distinguish between two types of initiative – task and dialogue. An agent has the task initiative when s/he controls how a task plan should be accomplished (e.g. by proposing a domain action). In contrast, an agent has the dialogue initiative when s/he controls the conversation (e.g. by establishing mutual beliefs between the agents). Chu-Carroll and Brown used an evidential approach based on Dempster–Shafer theory to learn a model for predicting when both types of initiative shifted from one conversational participant to another, given the current initiative holders and a set of user cues. In particular, who held each type

of initiative and various cues for shifting initiative (e.g. silence at the end of an utterance) were hand-labeled in a training set of utterances. An evaluation showed that user cues improved the accuracy of predictions about who held both types of initiative.

In follow-on work, Chu-Carroll used this framework to build an adaptive mixed-initiative spoken dialogue system for providing information about films (Chu-Carroll, 2000). The system predicted whether it or the user had each type of initiative, and generated different responses depending on the initiative holder. An empirical evaluation with human subjects demonstrated that the use of mixed initiative increased user satisfaction and dialogue efficiency (Chu-Carroll and Nickerson, 2000). Smith and Hipp (1994) also collected human–computer dialogues (using their circuit trouble-shooting spoken dialogue system), in order to empirically evaluate whether the use of a *directive* or *declarative* initiative mode would yield more efficient dialogues. In the directive mode their system always tried to achieve its own goals, while in the declarative mode the system tried to find and adopt common goals with the user. Their experimental results demonstrated that the declarative mode allowed the system to achieve a goal using fewer utterances.★

Other research has used computer–computer dialogue simulations to explore the utility of allowing mixed-initiative dialogue behaviors. Guinn (1998) presented a prescriptive model for automating mixed initiative, and used experimental computer–computer dialogue simulations to assess the efficiency of various schemes. Similarly, Ishizaki et al. (1999) used computer–computer simulations to determine when a mixed-initiative dialogue strategy (using an initiative model based on Whittaker and Stenton, 1988) improves dialogue efficiency in a route finding domain. Their experimental results showed that for easy problems (where difficulty was estimated as the ratio of the length of the shortest route and the length of the found route), mixed-initiative dialogues are sometimes a little more efficient than non-mixed-initiative dialogues. However, for difficult problems, non-mixed-initiative dialogues are in fact more efficient.

Although not in the area of natural language systems *per se*, a recent research trend has been to apply ideas from mixed-initiative dialogue systems to other types of interactive systems.★★ Stein et al. (1999) focused on providing mixed-initiative capabilities in the context of information retrieval interactions, where users often opportunistically change their goals and strategies. Cesta and D'Aloisi (1999) considered issues of mixed-initiative interaction in the context of delegation-based agent systems (e.g. a meeting scheduling assistant), while Lester et al. (1999) focused on the area of life-like agents for learning environments. Rich and Sidner (1998) developed a collaboration manager for software interface agents, based on collaborative theories from the area of natural language discourse.

★ Other work on initiative and spoken dialogue (Allen et al., 1996; Hagen, 1999; Litman and Pan, 1999) is described below in Section 5.5.
★★ Several of these systems hope to incorporate full-blown natural language understanding components in the future.

Finally, the research described in Chu-Carroll and Carberry (1995) and Green and Carberry (1999) is in the intersection of the areas of mixed initiative and NLG (Section 2). Chu-Carroll and Carberry developed a computational strategy for deciding when to initiate an information-sharing sub-dialogue and determining the focus of this sub-dialogue in the context of a collaborative activity. Green and Carberry focused on the use of mixed initiative for reply generation. They proposed that a system should maintain and evaluate a set of stimulus conditions as supplements to discourse plan operators, in order to know when and how to incorporate extra information into replies to yes–no questions. Both systems relied on a model of the user's beliefs and the user's plan. Such a model can be explicitly or implicitly conveyed (sometimes incrementally) from previous dialogue, or inferred using stereotypes, by applying standard techniques for inferring user models (e.g. Kass, 1991).

5.3. USING DIALOGUE TO INCREMENTALLY BUILD USER MODELS

User modeling systems have often relied on hand-crafted user models provided in advance by a system designer. In the context of a dialogue system (and as discussed in Section 2.4), there is also the opportunity for a system to dynamically and incrementally update its user model as a dialogue progresses. For example, each time the user makes an utterance, the system can use the contents of the utterance (explicitly or via inference) to update the user model. A system can also take a more active role, by initiating a sub-dialogue whenever it feels that it would be useful to obtain some missing user modeling information (e.g. Cawsey, 1990, 1993, discussed in Section 2; Wu, 1991 and van Beek, 1991, discussed in Carberry, 2001 in this issue). The intuition behind the active approach is that the length added by the extra utterances in a knowledge-acquisition sub-dialogue might in fact reduce the length of subsequent dialogue, thus reducing dialogue length overall (Shifroni and Shanon, 1992).

Chin (1989) and Kass (1991) made a first step towards developing a domain independent module for acquiring a user model, where information about a user was inferred during the user's dialogue with an advisory system. This was done by using information obtained during the interaction to activate hand-crafted domain-independent heuristics for acquiring a user model (e.g. infer that the user knows the generalization of a concept if the user talks about three specific instantiations of the concept, infer that the user understands some terminology if the user does not ask for a clarification). Interestingly, the domain-independent rules separately developed by Kass and by Chin had very little overlap.

More recently, machine learning techniques have been used to automatically derive user model acquisition rules. However, while the automatically learned rules are similar in form to those developed by hand by Chin and Kass, their content and focus differ. For example, rules have been automatically learned to predict speech misrecognitions (Litman and Pan, 2000), as discussed below in Section 5.6.

Quilici (1994) showed how to use specific types of negative user-feedback during plan-oriented dialogues with an advisory system in order to infer a set of plan-oriented user beliefs that are likely contributors to misconceptions. This work differs from work discussed earlier in this article (e.g. 'one-shot' approaches to handling misconceptions described in McCoy, 1989; Pollack 1990) in that the user model was inferred gradually from user feedback as the dialogue progressed. This allowed the system to (1) provide an initial "standard" answer which is tuned as the dialogue progresses, (2) avoid inference by waiting for the user to provide missing details, and (3) control inference by trying to relate feedback only to previous dialogue responses. In addition, Quilici's work focused primarily on understanding a user's feedback, while previous work on combining explanation planning and user feedback focused on generating a response (e.g. Moore and Paris, 1992 and Cawsey, 1990, 1993 in Section 2.4). As Quilici himself noted, work in explanation and feedback could benefit by integrating these approaches.

5.4. MODELING PREFERENCES IN CONSULTATION DIALOGUES

Morik (1989) and Elzer et al. (1994) argued that to better support consultation dialogues, users' preferences should be recognized during a dialogue. Both Morik and Elzer et al. considered the acquisition of a model of a user's preferences. However, Morik used this model to generate replies to the user's questions, while Elzer et al. used it to evaluate and improve the user's plan.

Morik's system (Morik, 1989), used both explicit and implicit user model acquisition to determine the user's preferences (which she called "evaluation criteria" or "evaluation standards"). Explicit user model acquisition was performed to obtain basic facts about a user from which a set of stereotypes was inferred. These stereotypes, as well as the user's follow-up questions, were used to infer the user's evaluation criteria. The match between these criteria and the domain properties known to the system enabled the system to moderate the strength of its recommendations and to determine whether and how to provide additional information when answering the user's questions. For instance, when asking whether a proposition P is true, the user may want P or $\neg P$. Depending on the user's evaluation criteria, the system may offer additional domain information related to the features of P or to those of $\neg P$.*

Elzer et al. (1994) presented a recognition strategy that utilizes characteristics of both a user's utterances and the dialogue to model attribute-value preferences. For example, a user of a course-advising system might have a preference for afternoon courses. This preference could be explicitly conveyed by the user (e.g. "I like afternoon courses"), or deduced by the system based on an analysis of rejected suggestions. In addition to recognizing a user's preferences, the system used the way in which the preferences were conveyed to generate a strength for these prefer-

---

* This idea is similar to Jameson's (1989), whereby the inclusion or exclusion of information in a reply depends on the perceived requirements of the interlocutor (Section 2.1).

ences (i.e. how important are these preferences to the user). Finally, the system main-
tained endorsements to represent its confidence in its beliefs (i.e. how strongly the
system believes that the preferences in its model reflect the user's). By exploiting
this model of user preferences, Elzer et al. showed how a dialogue system can detect
that a user has proposed a sub-optimal solution, and then suggest a better
alternative.

5.5. TOWARDS SPOKEN DIALOGUE SYSTEMS

Due to recent technological advances, it is now possible to build real-time, interactive
spoken dialogue systems, where the user's input is passed through an automatic
speech recognizer, and the system's output is sent to a text-to-speech synthesizer.
The use of speech in dialogue systems presents new opportunities for user modeling,
including the development of new types of user models, and the acquisition of
traditional types of user models using new methodologies and triggers.

For example, recent research in the area of mixed-initiative dialogues has focused
on new issues that are particularly relevant for spoken dialogue systems.
Hagen (1999) developed an approach to mixed-initiative dialogue that was tailored
to the types of restrictions often found in telephone-based, spoken language inter-
faces to databases. Such systems must operate in real time, and often have access
to only simple linguistic representations (in their case, speech recognition was based
on phrase spotting techniques). Thus, the user modeling techniques used in these
systems need to respect these constraints. Hagen showed how a computationally
tractable dialogue grammar could be used to parse dialogues into a dialogue history
of speech acts. By modeling the dialogue history and the data needs of the underlying
information retrieval system, the dialogue manager of the system was able to adapt
to a user's attempts to change initiative. Similarly, the TRAINS mixed-initiative
spoken dialogue system (Allen et al., 1996) focused on adapting ideas underlying
the plan-based approach in order to handle the demands of robustness to recognition
errors and real-time performance inherent in speech-based systems. Finally, the
VERBMOBIL speech-to-speech translation system (Alexandersson et al., 1997)
focused on achieving robustness by combining knowledge-based and statistical
approaches to dialogue processing (this system is discussed further in Section 5.6).

Litman et al. (1999) proposed a new type of user model for managing initiative in
dialogue systems with speech input. They argued that a spoken dialogue system
should dynamically estimate the performance of the speech recognition component
for every user in every dialogue. They applied a machine learning approach to learn
classification rules that predict poor speech recognition performance from features
available in system logs, using previously labeled human–computer dialogue corpora
as training data. In follow-on work, Litman and Pan (2000) implemented a spoken
dialogue system which used these predictions to automatically change dialogue
initiative, e.g. the system started asking focused questions when a sustained high
level of misrecognition was inferred.

Because the use of speech recognition often leads to errorful input, the issue of system verification of user input takes on extra importance in spoken dialogue systems. Smith (1998) designed and evaluated strategies for dynamically deciding whether to confirm each user utterance during a task-oriented dialogue. Simulation results suggested that context-dependent adaptation strategies (e.g. using dialogue expectations based on task knowledge, in addition to using local information from the parser) can improve performance, especially when the system has greater initiative.

Berthold and Jameson (1999) also proposed a new component for incorporation into a user model for a dialogue system, namely a representation of the cognitive load of a user. Since high cognitive load results from situational distractions, the authors speculated that it would be useful for a system to adapt to inferred cognitive load by using simpler input or output strategies. As a first step toward this goal, the authors used an empirically motivated BN to assess a user's cognitive load from characteristics of his/her speech (e.g. sentence fragments and articulation rate).

5.6. Towards empirical methods – strengthening the empirical basis of user models, and experimentally evaluating their utility

Most of the early work in user modeling for dialogue was motivated and justified by intuition, or at best by fairly informal analyses of human–human example dialogues. Furthermore, traditional prototype dialogue systems typically relied on hand-crafted plan libraries or detailed models of users' beliefs. As in many other areas of user modeling (see the papers in this issue by Webb et al., 2001; Zukerman and Albrecht, 2001; and Chin, 2001), the area of user modeling for dialogue has recently seen a welcome and increasing use of empirical methods. Machine learning and reasoning under uncertainty techniques have been used to automatically acquire user models from dialogue corpora, where the accuracy of the learned models can be quantitatively evaluated. In addition, experiments with human participants as well as with computer–computer simulations have been used to evaluate the impact that a user modeling component has on the rest of a dialogue system.

In the context of the work on cognitive load discussed above, Berthold and Jameson (1999) presented a method for synthesizing previous experimental results, which they used to derive qualitative and quantitative constraints for a BN, thus strengthening its empirical basis. They also showed how to use such analyses to generate artificial user data that can be used for evaluation purposes. Recall that other uses of artificial data were discussed above, namely the use of computer–computer dialogue simulations in the context of evaluating mixed-initiative dialogue systems (Guinn, 1998; Ishizaki et al., 1999; Smith, 1998).

Advances in machine learning and reasoning under uncertainty approaches have also strengthened the empirical basis of user models in dialogue systems. As discussed above, the systems described in (Chu-Carroll and Brown, 1998; Litman

et al., 1999) automatically learned user models which were used to determine when initiative should change during the course of a dialogue. To this effect, both systems used observations regarding the progress of the dialogue as training data. Then, the accuracy of the acquired user models was quantitatively evaluated on different test data. In the dialogue module of the VERBMOBIL speech-to-speech translation system (Alexandersson et al., 1997), models that made plan-based inferences and models that predicted dialogue acts were acquired automatically (at least partially). Declarative plan operators for plan recognition were both hand-coded and automatically derived from a corpus using a Bayesian learning technique transferred from the field of grammar extraction. The corpus was also used to generate predictive models which calculate probabilities of follow-up dialogue acts given a previous dialogue act sequence.

Finally, as recent systems have become robust enough to try out with people, controlled experiments involving human–computer interactions have become possible. Litman and Pan (1999) empirically demonstrated the utility of initiative adaptation in a train-timetable spoken dialogue system where the user (as opposed to the system) triggered the adaptation process. The experimental evaluation of the follow-on system described in Litman and Pan (2000), where the adaptation process was automated, showed that the adaptive system outperformed a non-adaptive version. Other experimental evaluations of spoken dialogue systems with human users were discussed above (Smith and Hipp, 1994; Chu-Carroll and Nickerson, 2000).

6. Thoughts for the Future

As stated in Section 1, user models were expected to improve the ability of natural language systems to understand a user, help achieve adaptivity in natural language interactions, and increase the robustness of natural language systems. These hopes have been achieved only partially, and represent a continuing challenge. In this section, we consider additional challenges for user models in natural language systems in the context of insights obtained from the preceding discussion.

As seen in the previous sections, a substantial proportion of natural language systems which consult user models focused on achieving specific natural language capabilities. Researchers working on these systems stated the demands these capabilities imposed on user models, but often gave little consideration to how these demands were to be satisfied. At first glance, such research may appear deficient in its coverage of the relevant factors. However, it is crucial for such research to continue, as it highlights the desired capabilities of a user model, without being restricted by what is currently possible. This gives researchers who investigate the acquisition of user models clear targets to aim for.

At the same time, we should build systems that integrate several natural language components. The development of such systems offers several advantages from the user modeling perspective: (1) they will require the development of comprehensive

user models that can contribute to several aspects of natural language processing, e.g. interpreting users' utterances, addressing possible misconceptions, and adapting the content of the discourse and the style of the interaction to the users' requirements; (2) these systems will afford the opportunity to evaluate the contribution of different parts of user models to natural language interactions as a whole; and (3) they will provide additional opportunities for the deployment of user modeling systems (as well as natural language systems).

Natural language systems that consult user models should be able to obtain sufficient information from their interactions with users so that they can maintain consistent and accurate user models. In addition, they should be able to recover from flaws in these models and adapt to the changing requirements of users. Multi-dimensional user models and enhanced user models make these problems more difficult. Advances in machine learning and predictive statistical models offer promising techniques for acquiring and updating user models, thereby increasing the applicability of the systems which consult these models (Webb et al., 2001; Zukerman and Albrecht, 2001). However, the application of these techniques to learn features of user models that affect natural language systems (or natural language features that affect user models) is a relatively recent occurrence. To date, machine learning techniques have been used to learn mainly coarse features of natural language systems, e.g. when to change initiative in the course of a dialogue (Chu-Carroll and Brown, 1998; Litman et al., 1999), and how to derive plan operators and predict dialogue acts (Alexandersson et al., 1997). Bayesian networks have been applied to a variety of user modeling tasks related to natural language, such as plan recognition (Charniak and Goldman, 1993), argument generation (Zukerman et al., 1998) and dialogue modeling (Jameson et al., 1995; Horvitz and Paek, 1999), but these networks were hand-crafted. In Section 2 we have projected future advances that will enable Bayesian networks to better support the requirements of adaptive natural language systems. However, more general questions must be answered in the context of using techniques from machine learning and reasoning under uncertainty to support the user modeling requirements of natural language systems: (1) which natural language and user modeling capabilities can be supported by the currently available techniques? and (2) which additional advances are required to support capabilities that are not catered for at present?

The evaluation of the contribution of user models to natural language systems is an important and difficult issue. Such an evaluation must assess the accuracy of the user model consulted by a natural language system, and demonstrate a measurable improvement in the performance of this system as a result of consulting this model (which assesses indirectly the accuracy of the user model). Although the evaluation of the performance of natural language systems is the topic of much debate, and some researchers have conducted both types of evaluations with respect to the specific capabilities of their systems (e.g. Litman and Pan, 1999; Zukerman and McConachy, 2001), generally accepted guidelines for the evaluation of user models and natural language systems are still forthcoming.

Finally, the deployment of natural language systems that consult user models will increase the immediacy of the demand for robust natural language systems. Such systems are expected to operate in realistic application domains and to be useful to a wide variety of people. The construction of such systems will also require the extension of currently available user modeling techniques. Here too we look to advances in machine learning and predictive statistical models to provide insights regarding such extensions.

Acknowledgments

References

Alexandersson, J., Reithinger, N. and Maier, E.: 1997, Insights into the Dialogue Processing of VERBMOBIL. In: *Proceedings of the Fifth Conference on Applied Natural Language Processing.* Washington, D.C., pp. 33–40.

Allen, C. S. and Bryant, B. R.: 1996, Learning a User's Linguistic Style: Using an Adaptive Parser to Automatically Customize a Unification-Based Natural Language Grammar. In: *UM96 – Proceedings of the Fifth International Conference on User Modeling.* Kona, Hawaii, pp. 35–42.

Allen, J. and Perrault, C.: 1980, Analyzing Intention in Utterances. *Artificial Intelligence* **15**, 143–178.

Allen, J. F., Miller, B. W., Ringger, E. K. and Sikorski, T.: 1996, A Robust System For Natural Spoken Dialogue. In: *Proceedings of the Thirty-Fourth Annual Meeting of the Association of Computational Linguistics (ACL).* Santa Cruz, California, pp. 62–70.

Allgayer, J., Ohlback, H. J. and Reddig, C.: 1992, Modelling Agents with Logic. In: *UM92 – Proceedings of the Third International Workshop on User Modeling.* Dagstuhl, Germany, pp. 22–34.

Asami, K., Takeuchi, A. and Otsuki, S.: 1996, Methods for Modeling and Assisting Causal Understanding in Physical Systems. In: *UM96 – Proceedings of the Fifth International Conference on User Modeling.* Kona, Hawaii, pp. 145–152.

Bateman, J. and Paris, C.: 1989, Phrasing a Text in Terms a User can Understand. In: *IJCAI89 – Proceedings of the Eleventh International Joint Conference on Artificial Intelligence.* Detroit, Michigan, pp. 1511–1517.

Berthold, A. and Jameson, A.: 1999, Interpreting Symptoms of Cognitive Load in Speech Input. In: *UM99 – Proceedings of the Seventh International Conference on User Modeling.* Banff, Canada, pp. 235–144.

Binsted, K., Cawsey, A. and Jones, R.: 1995, Generating Personalized Patient Information Using the Medical Record. In: *Proceedings of the Fifth Conference on Artificial Intelligence in Medicine – Europe, Lecture Notes in Computer Science,* **934**, 29–41.

Bonarini, A.: 1993, Modeling Issues in Multimedia Car-Driver Interaction. In: M. T. Maybury (ed.): *Intelligent Multimedia Interfaces.* Menlo Park, California: AAAI Press/The MIT Press, pp. 353–371.

Carberry, S.: 1985, A Pragmatics-based Approach to Understanding Intersentential Ellipsis. In: *Proceedings of the Twenty-Third Annual Meeting of the Association for Computational Linguistics.* Chicago, Illinois, pp. 188–197.

Carberry, S.: 1988, Modeling the User's Plans and Goals. *Computational Linguistics* **14**(3), 23–37.

Carberry, S.: 1990, Incorporating Default Inferences into Plan Recognition. In: *AAAI90 – Proceedings of the Eight National Conference on Artificial Intelligence.* Boston, Massachusetts, pp. 471–478.

Carberry, S.: 2001, Techniques for Plan Recognition. *User Modeling and User-Adapted Interaction* **11**(1–2), 31–48 (this issue).

Carberry, S. and Lambert, L.: 1999, A Process Model for Recognizing Communicative Acts and Modeling Negotiation Subdialogues. *Computational Linguistics* **25**(1), 1–53.

Carenini, G., Mittal, V. and Moore, J. D.: 1994, Generating Patient Specific Interactive Natural Language Explanations. In: *Proceedings of the Eighteenth Symposium on Computer Applications in Medical Care.* Banff, Canada.

Carenini, G. and Moore, J. D.: 1999, Tailoring Evaluative Arguments to a User's Preferences. In: *UM99 – Proceedings of the Seventh International Conference on User Modeling.* Banff, Canada, pp. 299–301.

Cawsey, A.: 1990, Generating Explanatory Discourse. In: R. Dale, C. Mellish, and M. Zock (eds.): *Current Research in Natural Language Generation.* Academic Press, pp. 75–102.

Cawsey, A.: 1993, User Modeling in Interactive Explanations. *User Modeling and User-Adapted Interaction* **3**(3), 221–248.

Cesta, A. and D'Aloisi, D.: 1999, Mixed-Initiative Issues in an Agent-Based Meeting Scheduler. *User Modeling and User-Adapted Interaction* **9**(1-2), 45–78.

Charniak, E. and Goldman, R. P.: 1993, A Bayesian Model of Plan Recognition. *Artificial Intelligence* **64**(1), 50–56.

Chin, D.: 1989, KNOME: Modeling What the User Knows. In: A. Kobsa and W. Wahlster (eds.): *User Models in Dialog Systems.* Springer-Verlag, pp. 74–107.

Chin, D.: 2001, Empirical Evaluation of User Models. *User Modeling and User Adapted Interaction* **11**(1–2), 181–194 (this issue).

Chin, D. N., Inaba, M., Pareek, H., Nemoto, K., Wasson, M. and Miyamoto, I.: 1994, Multi-Dimensional User Models for Multi-media I/O in the Maintenance Consultant. In: *UM94 – Proceedings of the Fourth International Conference on User Modeling.* Hyannis, Massachusetts, pp. 139–144.

Chu-Carroll, J.: 2000, MIMIC: An Adaptive Mixed Initiative Spoken Dialogue System for Information Queries. In: *Proceedings of the Sixth ACL Conference on Applied Natural Language Processing (ANLP).* Seattle, Washington, pp. 97–104.

Chu-Carroll, J. and Brown, M. K.: 1998, An Evidential Model for Tracking Initiative in Collaborative Dialogue Interactions. *User Modeling and User-Adapted Interaction* **8**(3-4), 215–253.

Chu-Carroll, J. and Carberry, S.: 1995, Generating Information-Sharing Subdialogues in Expert-User Consultation. In: *IJCAI95 – Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence.* Montreal, Canada, pp. 1243–1250.

Chu-Carroll, J. and Nickerson, J. S.: 2000, Evaluating Automatic Dialogue Strategy Adaptation for a Spoken Dialogue System. In: *Proceedings of the First Conference of the North American Chapter of the Association for Computational Linguistics (NAACL).* Seattle, Washington, pp. 202–209.

Cohen, R., Allaby, C., Cumbaa, C., Fitzgerald, M., Ho, K., Hui, B., Latulipe, C., Lu, F., Moussa, N., Pooley, D., Quian, A. and Siddiqi, S.: 1998, What is Initiative? *User Modeling and User-Adapted Interaction* **8**(3-4), 171–214.

de Rosis, F., Grasso, F. and Berry, D.: 1999, Refining Instructional Text Generation after Evaluation. *Artificial Intelligence in Medicine* **17**, 1–36.

Dols, F. J. and van der Sloot, K.: 1992, Modelling Mutual Effects in Belief-based Interactive Systems. In: *UM92 – Proceedings of the Third International Workshop on User Modeling.* Dagstuhl, Germany, pp. 3–19.

Elzer, S., Chu-Carroll, J. and Carberry, S.: 1994, Recognizing and Utilizing User Preferences in Collaborative Consultation Dialogues. In: *UM94 – Proceedings of the Fourth International Conference on User Modeling.* Hyannis, Massachusetts, pp. 19–24.

Fain Lehman, J. and Carbonell, J.: 1989, Learning the User's Language: A Step Towards Automated Creation of User Models. In: A. Kobsa and W. Wahlster (eds.): *User Models in Dialog Systems.* Springer-Verlag, pp. 163–194.

Friedman, N. and Goldszmidt, M.: 1999, Learning Bayesian Networks from Data. *Tutorial D3 – The Sixteenth International Joint Conference on Artificial Intelligence.* Stockholm, Sweden.

Green, N. and Carberry, S.: 1999, A Computational Mechanism for Initiative in Answer Generation. *User Modeling and User-Adapted Interaction* **9**(1–2), 93–132.

Grosz, B. J.: 1977, The Representation and Use of Focus in Dialogue Understanding. Technical Report 151, SRI International, Menlo Park, California.

Guinn, C. I.: 1998, An Analysis of Initiative Selection in Collaborative Task-Oriented Discourse. *User Modeling and User-Adapted Interaction* **8**(3-4), 255–314.

Hagen, E.: 1999, An Approach to Mixed Initiative Spoken Information Retrieval Dialogue. *User Modeling and User-Adapted Interaction* **9**(1-2), 45–78.

Halliday, M. A.: 1978, *Language as Social Semiotic.* London: Edward Arnold.

Hintikka, J.: 1962, *Knowledge and Belief.* New York: Cornell University Press.

Hirst, G., DiMarco, C., Hovy, E. and Parsons, K.: 1997, Authoring and Generating Health-Education Documents that are Tailored to the Needs of the Individual Patient. In: *UM97 – Proceedings of the Sixth International Conference on User Modeling.* Sardinia, Italy, pp. 107–118.

Horacek, H.: 1997, A Model for Adapting Explanations to the User's Likely Inferences. *User Modeling and User-Adapted Interaction* **7**(1), 1–55.

Horvitz, E. and Paek, T.: 1999, A Computational Architecture for Conversation. In: *UM99 – Proceedings of the Seventh International Conference on User Modeling.* Banff, Canada, pp. 201–210.

Hovy, E. H.: 1988, *Generating Natural Language under Pragmatic Constraints.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Hustadt, U.: 1994, A Multi-Modal Logic for Stereotyping. In: *UM94 – Proceedings of the Fourth International Conference on User Modeling.* Hyannis, Massachusetts, pp. 87–92.

Ishizaki, M., Crocker, M. and Mellish, C.: 1999, Exploring Mixed-Initiative Dialogue Using Computer Dialogue Simulation. *User Modeling and User-Adapted Interaction* **9**(1–2), 45–78.

Jameson, A.: 1989, But What Will the Listener Think? Belief Ascription and Image Maintenance in Dialog. In: A. Kobsa and W. Wahlster (eds.): *User Models in Dialog Systems.* Springer-Verlag, pp. 255–312.

Jameson, A.: 1996, Numerical Uncertainty Management in User and Student Modeling: An Overview of Systems and Issues. *User Modeling and User-Adapted Interaction* **5**, 193–251.

Jameson, A., Schafer, R., Simons, J. and Weis, T.: 1995, Adaptive Provision of Evaluation-Oriented Information: Tasks and Techniques. In: *IJCAI95 – Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence.* Montreal, Canada, pp. 1886–1893.

Jitnah, N., Zukerman, I., McConachy, R. and George, S.: 2000, Towards the Generation of Rebuttals in a Bayesian Argumentation System. In: *Proceedings of the International Natural Language Generation Conference.* Mitzpe Ramon, Israel, pp. 39–46.

Joshi, A., Webber, B. L. and Weischedel, R. M.: 1984, Living Up to Expectations: Computing Expert Responses. In: *AAAI84 – Proceedings of the Fourth National Conference on Artificial Intelligence.* Austin, Texas, pp. 169–175.

Kashihara, A., Hirashima, T. and Toyoda, J.: 1995, A Cognitive Load Application in Tutoring. *User Modeling and User-Adapted Interaction* **4**(4), 279–303.

Kashihara, A., Nomura, K. Hirashima, T. and Toyoda, J.: 1996, Collaboration and Student Modeling in Instructional Explanation. In: *UM96 – Proceedings of the Fifth International Conference on User Modeling.* Kona, Hawaii, pp. 161–168.

Kass, R.: 1991, Building a User Model Implicitly from a Cooperative Advisory Dialog. *User Modeling and User-Adapted Interaction* **1**(3), 203–258.

Kass, R. and Finin, T.: 1988, Modeling the User in Natural Language Systems. *Computational Linguistics* **14**(3), 5–22.

Kay, J.: 2001, Learner Control. *User Modeling and User Adapted Interaction* **11**(1–2), 111–127 (this issue).

Kobsa, A. and Wahlster, W. (eds.): 1989, *User Models in Dialog Systems.* Springer-Verlag.

Konolige, K.: 1986, *A Deduction Model of Belief.* Morgan Kaufman.

Lambert, L. and Carberry, S.: 1991, A Tripartite Plan-based Model of Dialogue. In: *Proceedings of the Twenty-Ninth Annual Meeting of the Association for Computational Linguistics.* Berkeley, California, pp. 47–54.

Lascarides, A. and Oberlander, J.: 1992, Abducing Temporal Discourse. In: R. Dale, E. H. Hovy, D. Rösner, and O. Stock (eds.): *Aspects of Automated Language Generation.* Springer-Verlag, Berlin, pp. 167–182.

Lester, J. C., Stone, B. A. and Stelling, G. D.: 1999, Lifelike Pedagogical Agents for Mixed-Initiative Problem Solving in Constructivist Learning Environments. *User Modeling and User-Adapted Interaction* **9**(1-2), 1–44.

Litman, D. J.: 1986, Understanding Plan Ellipsis. In: *AAAI86 – Proceedings of the Fifth National Conference on Artificial Intelligence.* Philadelphia, Pennsylvania, pp. 619–624.

Litman, D. J. and Allen, J. F.: 1987, A Plan Recognition Model for Subdialogues in Conversation. *Cognitive Science* **11**, 163–200.

Litman, D. J. and Pan, S.: 1999, Empirically Evaluating an Adaptable Spoken Dialogue System. In: *UM99 – Proceedings of the Seventh International Conference on User Modeling.* Banff, Canada, pp. 55–64.

Litman, D. J. and Pan, S.: 2000, Predicting and Adapting to Poor Speech Recognition in a Spoken Dialogue System. In: *AAAI00 – Proceedings of the Seventeenth National Conference on Artificial Intelligence.* San Antonio, Texas, pp. 722–728.

Litman, D. J., Walker, M. A. and Kearns, M. J.: 1999, Automatic Detection of Poor Speech Recognition at the Dialogue Level. In: *Proceedings of the Thirty-Seventh Annual Meeting of the Association for Computational Linguistics (ACL).* College Park, Maryland, pp. 309–316.

Lochbaum, K.: 1995, The Use of Knowledge Preconditions in Language Processing. In: *IJCAI95 – Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence.* Montreal, Canada, pp. 1260–1266.

Marcu, D.: 1996, The Conceptual and Linguistic Facets of Persuasive Arguments. In: *Proceedings of ECAI-96 Workshop – Gaps and Bridges: New Directions in Planning and NLG.* Budapest, Hungary, pp. 43–46.

McCoy, K. F.: 1989, Highlighting a User Model to Respond to Misconceptions. In: A. Kobsa and W. Wahlster (eds.): *User Models in Dialog Systems.* Springer-Verlag, pp. 233–254.

McCoy, K. F., Pennington, C. and Suri, L. Z.: 1996, English Error Correction: A Syntactic User Model Based on Principled "Mal-Rule" Scoring. In: *UM96 – Proceedings of the Fifth International Conference on User Modeling.* Kona, Hawaii, pp. 59–66.

Mehl, S.: 1994, Forward Inferences in Text Generation. In: *ECAI94 – Proceedings of the Eleventh European Conference on Artificial Intelligence.* Amsterdam, The Netherlands, pp. 525–529.

Milosavljevic, M.: 1997, Augmenting the User's Knowledge via Comparison. In: *UM97 – Proceedings of the Sixth International Conference on User Modeling.* Sardinia, Italy, pp. 119–130.

Moore, J. D. and Paris, C. L.: 1992, Exploiting User Feedback to Compensate for the Unreliability of User Models. *User Modeling and User-Adapted Interaction* **2**(4), 287–330.

Moore, R. C.: 1980, Reasoning about Knowledge and Action. Ph. D. thesis, Department of Electrical Engineering and Computer Science.

Morik, K.: 1989, User Models and Conversational Settings: Modeling the User's Wants. In: A. Kobsa and W. Wahlster (eds.): *User Models in Dialog Systems.* Springer-Verlag, pp. 364–385.

Mostow, J. and Aist, G.: 1997, The Sounds of Silence: Towards Automated Evaluation of Student Learning in a Reading Tutor that Listens. In: *AAAI97 – Proceedings of the Fourteenth National Conference on Artificial Intelligence.* Providence, Rhode Island, pp. 355–361.

Paris, C. L.: 1989, The Use of Explicit User Models in a Generation System for Tailoring Answers to the User's Level of Expertise. In: A. Kobsa and W. Wahlster (eds.): *User Models in Dialog Systems.* Springer-Verlag, pp. 200–232.

Pearl, J.: 1988, *Probabilistic Reasoning in Intelligent Systems.* San Mateo, California: Morgan Kaufmann Publishers.

Perrault, C. and Allen, J.: 1980, A Plan-Based Analysis of Indirect Speech Acts. *American Journal of Computational Linguistics* **6**, 167–182.

Peter, G. and Rösner, D.: 1994, User-Model-Driven Generation of Instructions. *User Modeling and User-Adapted Interaction* **3**(4), 289–320.

Pollack, M.: 1990, Plans as Complex Mental Attitudes. In: P. Cohen, J. Morgan, and M. Pollack (eds.): *Intentions in Communication.* MIT Press, pp. 77–103.

Quilici, A.: 1989, Detecting and Responding to Plan-Oriented Misconceptions. In: A. Kobsa and W. Wahlster (eds.): *User Models in Dialog Systems.* Springer-Verlag, pp. 108–132.

Quilici, A.: 1994, Forming User Models by Understanding User Feedback. *User Modeling and User-Adapted Interaction* **3**(4), 321–358.

Raskutti, B. and Zukerman, I.: 1991, Generation and Selection of Likely Interpretations during Plan Recognition. *User Modeling and User-Adapted Interaction* **1**(4), 323–353.

Rich, C. and Sidner, C. L.: 1998, COLLAGEN: A Collaboration Manager for Software Interface Agents. *User Modeling and User-Adapted Interaction* **8**(3–4), 315–350.

Sarner, M. H. and Carberry, S.: 1992, Generating Tailored Definitions Using a Multifaceted User Model. *User Modeling and User-Adapted Interaction* **2**(3), 181–210.

Schuster, E.: 1985, Grammars as User Models. In: *IJCAI85 – Proceedings of the Ninth International Joint Conference on Artificial Intelligence.* Los Angeles, California, pp. 20–22.

Shifroni, E. and Shanon, B.: 1992, Interactive User Modeling: An Integrative Explicit-Implicit Approach. *User Modeling and User-Adapted Interaction* **2**(4), 331–366.

Sleeman, D.: 1984, Mis-Generalization: An Explanation of Observed Mal-rules. In: *Proceedings of the Sixth Annual Conference of the Cognitive Science Society.* Boulder, Colorado, pp. 51–56.

Smith, R. W.: 1998, An Evaluation of Strategies for Selectively Verifying Utterance Meanings in Spoken Natural Language Dialog. *International Journal of Human-Computer Studies* **48**, 627–647.

Smith, R. W. and Hipp, D. R.: 1994, *Spoken Natural Language Dialog Systems: A Practical Approach.* Oxford University Press.

Stein, A., Gulla, J. A. and Thiel, U.: 1999, User-Tailored Planning of Mixed Initiative Information-Seeking Dialogues. *User Modeling and User-Adapted Interaction* **9**(1–2), 133–166.

Stock, O. and the ALFRESCO Project Team: 1993, ALFRESCO: Enjoying the Combination of Natural Language Processing and Hypermedia for Information Exploration. In: M. T. Maybury (ed.): *Intelligent Multimedia Interfaces.* Menlo Park, California: AAAI Press/The MIT Press, pp. 197–2244.

Tattersall, C.: 1992, Generating Help for Users of Application Software. *User Modeling and User-Adapted Interaction* **2**(3), 211–248.

Taylor, J. A., Carletta, J. and Mellish, C.: 1996, Requirements for Belief Models in Cooperative Dialogue. *User Modeling and User-Adapted Interaction* **6**(1), 23–68.

van Beek, P.: 1987, A Model for Generating Better Explanations. In: *Proceedings of the Twenty-Fifth Annual Meeting of the Association for Computational Linguistics.* Stanford, California, pp. 215–220.

van Beek, P. and Cohen, R.: 1991, Resolving Plan Ambiguity for Cooperative Response Generation. In: *IJCAI91 – Proceedings of the Twelfth International Joint Conference on Artificial Intelligence.* Sydney, Australia, pp. 938–944.

Wallis, J. W. and Shortliffe, E. H.: 1985, Customized Explanations Using Causal Knowledge. In: B. C. Buchanan and E. H. Shortliffe (eds.): *Rule-based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project.* Addison-Wesley Publishing Company, pp. 371–388.

Webb, G. I., Pazzani, M. J. and Billsus D.: 2001, Machine Learning for User Modeling. *User Modeling and User Adapted Interaction* **11**(1–2), 19–29 (this issue).

Whittaker, S. and Stenton, P.: 1988, Cues and Control in Expert-Client Dialogues. In: *Proceedings of the Twenty-Sixth Annual Meeting of the Association of Computational Linguistics (ACL).* Buffalo, New York, pp. 123–130.

Wilensky, R.: 1983, *Planning and Understanding.* Reading, Massachusetts: Addison-Wesley.

Wu, D.: 1991, Active Acquisition of User Models: Implications for Decision-Theoretic Dialog Planning and Plan Recognition. *User Modeling and User-Adapted Interaction* **1**(2), 149–172.

Zukerman, I. and Albrecht, D.: 2001, Predictive Statistical Models for User Modeling. *User Modeling and User Adapted Interaction* **11**(1–2), 5–18 (this issue).

Zukerman, I., Jitnah, N. McConachy, R. and George, S.: 2000, Recognizing Intentions from Rejoinders in a Bayesian Interactive Argumentation System. In: *PRICAI2000 – Proceedings of the Sixth Pacific Rim International Conference on Artificial Intelligence.* Melbourne, Australia, pp. 252–263.

Zukerman, I. and McConachy, R.: 1993, Consulting a User Model to Address a User's Inferences during Content Planning. *User Modeling and User Adapted Interaction* **3**(2), 155–185.

Zukerman, I. and McConachy, R.: 1995, Generating Discourse across Several User Models: Maximizing Belief while Avoiding Boredom and Overload. In: *IJCAI95 – Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence.* Montreal, Canada, pp. 1251–1257.

Zukerman, I. and McConachy, R.: 2001, WISHFUL: A Discourse Planning System that Considers a User's Inferences. *Computational Intelligence* **17**(1).

Zukerman, I., McConachy, R. and Korb, K. B.: 1996, Consulting a User Model while Generating Arguments. In: *UM96 – Proceedings of the Fifth International Conference on User Modeling.* Kona, Hawaii, pp. 153–160.

Zukerman, I., McConachy, R. and Korb, K. B.: 1998, Bayesian Reasoning in an Abductive
    Mechanism for Argument Generation and Analysis. In: *AAAI98 – Proceedings of the
    Fifteenth National Conference on Artificial Intelligence.* Madison, Wisconsin, pp. 833–838.

Authors' Vitae

Ingrid Zukerman is an Associate Professor in Computer Science at Monash
University. She received her B.Sc. degree in Industrial Engineering and Management
and her M.Sc. degree in Operations Research from the Technion – Israel Institute of
Technology. She received her Ph.D. degree in Computer Science from UCLA in
1986. Since then, she has been working in the School of Computer Science and Soft-
ware Engineering at Monash University. Her areas of interest are discourse
planning, plan recognition and agent modeling.

Diane Litman is a Principal Technical Staff Member at AT&T Labs – Research. She
received her A.B. degree in Mathematics and Computer Science from the College of
William and Mary in Virginia in 1980, and her M.S. and Ph.D. degrees in Computer
Science from the University of Rochester in 1982 and 1986, respectively. Since then,
she has been working in the Artificial Intelligence Principles Research Department,
AT&T Labs – Research (formerly Bell Laboratories); from 1990–1992, she was also
an Assistant Professor of Computer Science at Columbia University. Her research
interests include computational linguistics, knowledge representation and reasoning,
natural language learning, plan recognition, and spoken language processing.