Mixed Depth Representations for Dialog Processing

Susan W. McRoy (mcroy@uwm.edu)
Department of EECS
University of Wisconsin–Milwaukee
Milwaukee, WI 53201

Syed S. Ali (syali@uwm.edu)
Department of Mathematical Sciences
University of Wisconsin–Milwaukee
Milwaukee, WI 53201

Susan M. Haller (haller@cs.uwp.edu)
CS & E Department
University of Wisconsin–Parkside
Kenosha, WI 53141

Abstract

We describe our work on developing a general purpose tutoring system that will allow students to practice their decision-making skills in a number of domains. The tutoring system, B2, supports mixed-initiative natural language interaction. The natural language processing and knowledge representation components are also general purpose—which leads to a tradeoff between the limitations of superficial processing and syntactic representations and the difficulty of deeper methods and conceptual representations. Our solution is to use a *mixed-depth* representation, one that encodes syntactic and conceptual information in the same structure. As a result, we can use the same representation framework to produce a detailed representation of requests (which tend to be well-specified) and to produce a partial representation of questions (which tend to require more inference about the context). Moreover, the representations use the same knowledge representation framework that is used to reason about discourse processing and domain information—so that the system can reason with (and about) the utterances, if necessary.

Introduction

Building decision support systems involves the collection and representation of a large amount of knowledge. It also involves providing mechanisms for reasoning over this knowledge efficiently. To make the best use of these efforts, our project group, which involves researchers at the University of Wisconsin-Milwaukee, the University of Wisconsin-Parkside, and the Medical College of Wisconsin, is working on a system to redeploy our decision support tools to build new systems for educating students. Our aim is to give students an opportunity to practice their decision making skills by considering a number of scenarios that differ from each other in a controlled manner. We also wish to give students the opportunity to ask the system to explain what factors most influenced the system.

Another important goal of this work is to develop tools that will cover a number of different domains, including medical diagnosis, life skills such as good nutrition and budgeting, and student success in their degree program. Thus, the graphical parts of the system interface are fairly generic. Most interaction between the student and the system uses natural language. This approach allows the system to explain fairly sophisticated topics (such as why a positive computed tomography, CT, scan supports the diagnosis of gallstones) and to tailor the interaction to the user's concerns and apparent level of understanding. This approach also gives the student a great deal of flexibility in how she phrases her questions or requests. Students' utterances may be short and ambiguous, requiring extensive reasoning about the domain or the discourse model to fully resolve. However, full disambiguation is rarely necessary. Our solution is to use a *mixed-depth* representation.

A mixed-depth representation is one that may be shallow or deep in different places, depending on what was known or needed at the time the representation was created (Hirst and Ryan, 1992). Moreover, "shallow" and "deep" are a matter of degree. Shallow representations can be a surface syntactic structure, or it can be the text itself (as a string of characters). Deep representations might be a conventional first-order (or higher-order) AI knowledge representation, taking into account such aspects of language understanding as lexical disambiguation, marking case relations, attachment of modifiers of uncertain placement, reference resolution, quantifier scoping, and distinguishing extensional, intensional, generic, and descriptive noun phrases. Unlike quasi-logical form, which is used primarily for storage of information, mixed-depth representations are well-formed propositions, subject to logical inference. Disambiguation, when it occurs, is done by reasoning.

Other systems that represent sentence meaning at multiple levels (e.g., syntactic, semantic, and pragmatic) build separate complete structures for each level. So, for example, when processing an utterance, a parser must either defer all semantic processing or resolve semantic ambiguities without the full knowledge necessary to do so correctly. The mixed-depth approach opportunistically builds semantic constituent structures as soon as enough information is available. (Minimally, this will be information about the syntax of the utterance but it may include some semantic information about properties of discourse objects introduced by the utterance). This allows us to do as much semantic interpretation as possible, as soon as possible.

In our work, we use a parser with a linguistically based grammar to process the student's typed inputs and produce a structure that captures syntactic marking and bracketing, along with some conceptual information. Encoding decisions that require reasoning about the domain or about the discourse context are left to the knowledge representation and discourse processing components, respectively. The primary benefits of this approach are:

Generality The representation language and conceptual structures that are built during the initial parsing phase are not specific to any one domain.

Expressiveness The representation language is very expressive; our grammar covers a wide variety of syntactic constructions including fragments and sentences with embedded sentences (*e.g.* relative clauses and clause complements). For example, students can request a diagnostic exercise using a variety of forms including *Give me a story problem*, *I want you to describe a case for me.*, *Tell me a story.*, *A story please.* or *Another case.*

Uniformity The structures that are built by the parser are all subject to inference; they use the same conceptual framework that is used by the other components, including discourse processing, planning, and plan recognition. For example, students can request a diagnostic exercise by mentioning any step of the tutoring plan including *Tell me a story.* or *Quiz me.* or by mentioning the overall tutoring plan with *Tutor me.*

This paper gives an overview of our tutoring system, B2. It describes the natural language and knowledge representation components of B2, and our approach to the representation of questions and requests. The domain that we have developed most thoroughly helps medical students learn a statistical model for medical diagnosis. Many of the examples will be taken from this domain.

The Need for A Discourse Model

The first prototype of our current system is Banter (Haddawy et al., 1996). Banter is a tutoring shell that presents predefined story problems and short-answer questions on the basis of stored information about a particular medical situation, such as a patient who sees her doctor complaining of abdominal pains. This information comprises statistical relations among known aspects of a patient's medical history, findings from physical examinations of the patient, results of previous diagnostic tests, and the different candidate diseases. The information is represented as a Bayesian belief network. The system also includes a facility for explaining the system's reasoning to the student, The Banter shell has been designed to be general enough to be used with any network having nodes of hypotheses, observations, and diagnostic procedures.

A preliminary (and informal) user study of the Banter system with students at the Medical College of Wisconsin revealed two important facts: First, students like the idea of being able to set up hypothetical cases and witness how different actions might (or might not!) affect the statistical likelihood of a candidate diagnosis. Second, students do not like, and will not use, a system that overwhelms them with irrelevant information or that risks misleading them because it answers questions more narrowly than a teacher would. Students want to ask brief, context-dependent questions, such as *"Why CT?"* or *"What about ultrasound?"* and they prefer to give brief, context-dependent responses. Moreover, students like explanations that are tailored to their needs—sometimes only a single word answer, sometimes the answer along with its justification. A discourse model is necessary to deal with this.

The B2 Architecture

The B2 system performs three distinct, but interrelated, tasks that rely on a variety of information sources. The tasks are:

- Managing the interaction between the user and B2, including the interpretation of context-dependent utterances.
- Reasoning about the domain, for example, the relation between components of a medical case history and diseases that might occur.
- Meta-reasoning about the Bayesian reasoner and its conclusions, including an ability to explain the conclusions by identifying the factors that were most significant.

The B2 system consists of seven components (see Figure 1). In the diagram, solid, directed arrows indicate the direction of information flow between components. The system gets the user's input using a graphical user interface that supports both natural language interaction and mouse inputs. The *Parser* component of the *Parser/Generator* performs the first level of processing on the user input using its grammar and the domain information from the *Knowledge Representation Component*. The Parser interprets the user's inputs to form propositional representations of surface-level utterances for the Discourse Analyzer. The *Generator* produces natural language outputs from the text messages (propositional descriptions of text) that it receives from the *Discourse Planner*.

The system as a whole is controlled by a module called the *Discourse Analyzer*. The Discourse Analyzer determines an appropriate response to the user's actions on the basis of a model of the discourse and a model of the

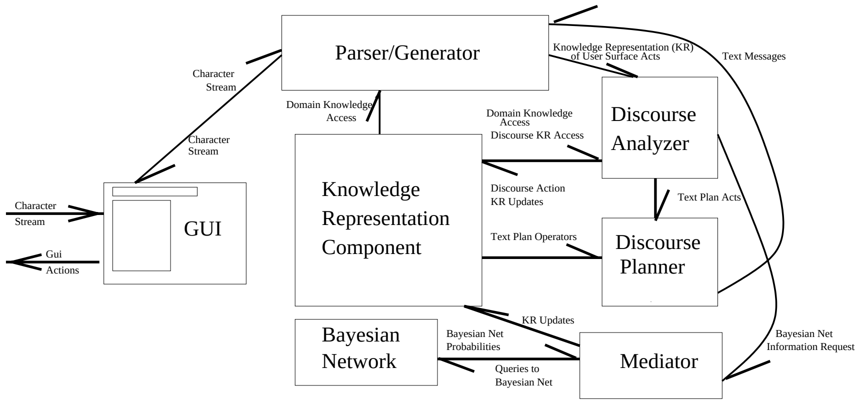

Figure 1: The B2 architecture

domain, stored in the knowledge representation component. The Analyzer invokes the *Discourse Planner* to select the content of the response and to structure it. The Analyzer relies on a component called the *Mediator* to interact with the Bayesian network processor. This Mediator processes domain level information, such as ranking the effectiveness of alternative diagnostic tests. All phases of this process are recorded in the knowledge representation component, resulting in a complete history of the discourse. Thus, the knowledge representation component serves as a central "blackboard" for all other components.

Together these seven components handle the three tasks mentioned above. They interact by addressing and handling queries to each other. However, the knowledge underlying these queries and the knowledge needed to generate a response can come from a variety of knowledge sources. Translating between knowledge sources is not an effective solution.

The information sources that B2 uses include:

- Linguistic knowledge — knowledge about the meanings of utterances and plans for expressing meanings as text.
- Discourse knowledge — knowledge about the intentional, social, and rhetorical relationships that link utterances.
- Domain knowledge — factual knowledge of the medical domain and the medical case that is under consideration.
- Pedagogy — knowledge about the tutoring task.
- Decision-support — knowledge about the statistical model and how to interpret the information that is derivable from the model.

In B2, the interaction between the tasks is possible because the information for all knowledge sources is represented in a uniform framework. The knowledge representation component serves as a central "blackboard" for all other components.

The Knowledge Representation Blackboard

B2 represents both domain knowledge and discourse knowledge in a uniform framework as a propositional semantic network. This allows the system to reason with (and about) utterances if necessary.

A propositional semantic network is a framework for representing the concepts of a cognitive agent who is capable of using language (hence the term *semantic*). The information is represented as a graph composed of nodes and labeled directed arcs. In a *propositional* semantic network, the propositions are represented by the nodes, rather than the arcs; arcs represent only non-conceptual binary relations between nodes. The particular systems that are being used for B2 are SNePS and ANALOG (Ali, 1994a; Ali, 1994b; Shapiro and Group, 1992) which provide facilities for building and finding nodes as well as for reasoning and truth-maintenance. These systems satisfy the following additional constraints:

- Each node represents a unique concept;
- Each concept represented in the network is represented by a unique node;
- The knowledge represented about each concept is represented by the structure of the *entire network* connected to the node that represents that concept.

These constraints allow efficient inference when processing natural language. For example, such networks can represent complex descriptions (common in the medical domain), and can support the resolution of ellipsis and anaphora, as well as general reasoning tasks such as subsumption (Ali, 1994a; Ali, 1994b; Maida and Shapiro, 1982; Shapiro and Rapaport, 1987; Shapiro and Rapaport, 1992).

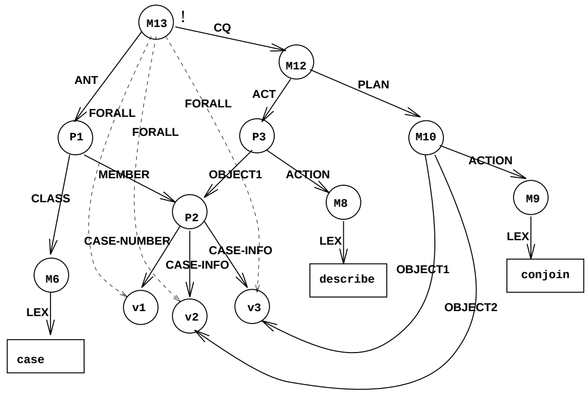

Figure 2: A rule stating that if `V1` is the case number of a case, and `V2` and `V3` are two pieces of case information, then a plan for generating a description of the case will present the two pieces of information in a coordinating conjunction.

We characterize a knowledge representation as *uniform* when it allows the representation of different kinds of knowledge in the same knowledge base using the same inference processes. The knowledge representation component of B2 is uniform because it provides a representation of the discourse knowledge, domain knowledge, and probabilistic knowledge (from the Bayesian net). This supports intertask communication and cooperation for interactive processing of tutorial dialogs.

The rule in Figure 2 is a good example of how the uniform representation of information in the semantic network allows us to relate domain information (a medical case) to discourse planning information (a plan to describe it). This network represents a text plan for describing a medical case to the user. Text plans are represented as rules in the knowledge representation. Rules are general statements about objects in the domain; they are represented by using *case frames*[1] that have `FORALL` or `EXISTS` arcs to nodes that represent variables that are bound by these quantifier arcs. In Figure 2, node `M13` is a rule with three universally quantified variables (at the end of the `FORALL` arcs), an antecedent (at the end of the `ANT` arc), and a consequent (at the end of the `CQ` arc). This means that if an instance of the antecedent is believed, then a suitably instantiated instance of the consequent is believed. Node `P1` represents the concept that something is a member of the class `case` and `P2` represents the concept that the case concept has a case number and case information. For more details about the knowledge representation, see (McRoy et al., 1997).

[1] Case frames are conventionally agreed upon sets of arcs emanating from a node that are used to express a proposition. For example, to express that A isa B we use the MEMBER-CLASS case frame which is a node with a MEMBER arc and a CLASS. arc (Shapiro et al., 1994) provides a dictionary of standard case frames. Additional case frames can be defined as needed.

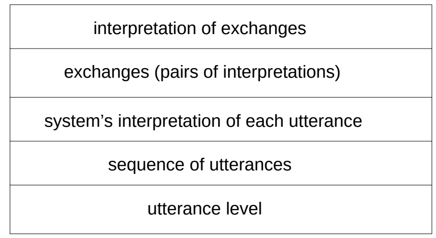

Figure 3: Five Levels of Representation

The Representation of the Discourse

The discourse model has five levels of representation, shown in Figure 3. These levels capture what the student and the system have each said, as well as how their utterances extend the ongoing discourse. Unlike many systems, B2's model of discourse will include a representation of questions and requests, as well as statements of fact. (Systems that do not represent questions and requests typically give these utterances a procedural semantics, interpreting them as operations to be performed.) Having an explicit representation of questions and requests simplifies the interpretation of context-dependent utterances such as *Why?* or *What about HIDA?*[2] (Haller, 1996). It also allows the system to recover from misunderstandings, should they occur (McRoy, 1995; McRoy and Hirst, 1995).

We will consider each of these levels in turn, starting with the utterance level, shown at the bottom of Figure 3.

The Utterance Level

For all inputs, the parser produces a representation of of its surface content, which the analyzer will assert as part of an occurrence of an event of type SAY. The content of the user's utterance is always represented by what she said literally. In the case of requests, the student may request a story problem directly, as an imperative sentence *Tell me a story* or indirectly, as a declarative sentence that expresses a desire *I want you to tell me a story.* The complete representation of the imperative sentence *Tell me a story* is shown in Figure 4.

For the system's utterances, the utterance level representation corresponds to a text generation event (this contains much more fine-grained information about the system's utterance, such as mode and tense.) The content of the system's utterance is the text message that is sent to the language generator.

Sequence of Utterances

The second level corresponds to the sequence of utterances. (This level is comparable to the linguistic structure in the tripartite model of (Grosz and Sidner, 1986)).

[2] HIDA stands for radio-nuclide hepatobilary imaging, a diagonistic test.

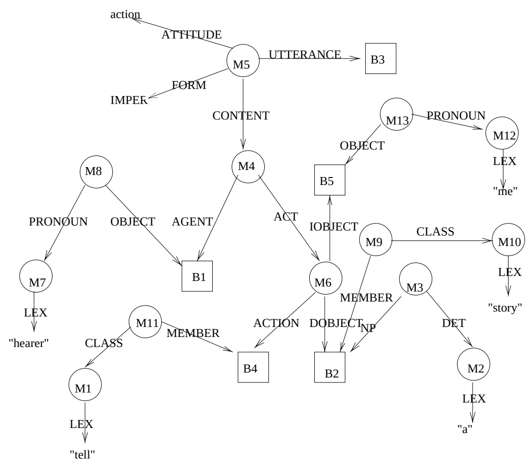

Figure 4: Node `B3` represents an utterance whose form is imperative, and whose content (`M4`) is the proposition that the hearer (`B1`) will tell a story (`B2`) to the speaker (`B5`).

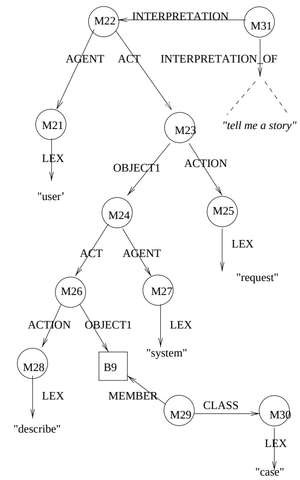

Figure 5: Node `M31` is a proposition that the interpretation of *Tell me a story* (which is glossed in this figure) is `M22`. Node `M22` is the proposition that the user requested that the system describe a case to the user.

In the semantic network, we represent the sequencing of utterances explicitly, with asserted propositions that use the BEFORE-AFTER case frame. The order in which utterances occurred (system and user) can be determined by traversing these structures. This representation is discussed in detail in (McRoy et al., 1997).

The Interpretation Level

In the third level, we represent the system's interpretation of each utterance. Each utterance event (from level 1) will have an associated system interpretation, which is represented using the INTERPRETATION_OF—INTERPRETATION case frame. For example, consider the interpretation of the utterance *Tell me a story* (as well as *I want you to tell me a story.*), shown in Figure 5. (Every utterance has one or more interpretations; at any time, only one is believed and a justification-based truth maintenance system is used to track changes in belief.)

The Exchange and Exchange Interpretation Levels

The fourth and fifth levels of representation in our discourse model are exchanges and interpretations of exchanges, respectively. A *conversational exchange* is a pair of interpreted events that fit one of the conventional structures for dialog (*e.g.* QUESTION–ANSWER). Figure 6 gives the network representation of a conversational exchange and its interpretation. Node `M113` represents the exchange in which the system has asked a question and the user has answered it. Using the MEMBER-CLASS case frame, propositional node `M115` asserts that the node `M113` is an exchange. Propositional node `M112` represents the system's interpretation of this exchange: that the user has accepted the system's question (*i.e.* that the user has understood the question and requires no further clarification). Finally, propositional node `M116` represents the system's belief that node `M112` is the interpretation of the exchange represented by node `M113`.

Interaction among the Levels

A major advantage of the network representation is the knowledge sharing between these five levels. We term this knowledge sharing *associativity.* This occurs because the representation is uniform and every concept is represented by a unique node. As a result, we can retrieve and make use of information that is represented in the network implicitly, by the arcs that connect propositional nodes. For example, if the system needed to explain why the user had said HIDA, it could follow the arcs from the node representing the utterance that *User said HIDA* to the system's interpretation of that utterance, node `M108`, to determine that

- The user's utterance was understood as the answer within an exchange (node `M113`), and
- The user's answer indicated her acceptance and understanding of the discourse, up to that point `M112`.

This same representation could be used to explain why the system believed that the user had understood the system's question. This associativity in the network is vital if the interaction starts to fail.

The Current Status of B2

B2 is being developed using the Common LISP programming language. We are using the SNePS 2.3.1 and ANALOG 1.1 tools to create the lexicon, parser, generator,

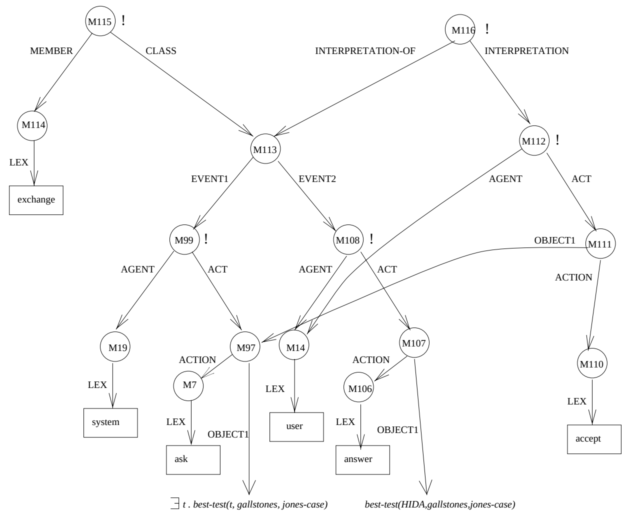

Figure 6: Node `M115` represents the proposition that node `M113` is an exchange comprised of the events `M99` and `M108`. `M108` is the proposition that *The user answered "HIDA is the best test to rule in Gallstones"*. Additionally, node `M116` represents the proposition that the interpretation of `M113` is event `M112`. `M112` is the proposition that the user has accepted `M96`. (`M96` is the question that the system asked in event `M99`.)

and underlying knowledge representations of domain and discourse information(Shapiro and Group, 1992; Shapiro and Rapaport, 1992; Ali, 1994a; Ali, 1994b).

An Internet-accessible, graphical front-end to B2 has been developed using the JAVA 1.1 programming language. It can be run using a network browser, such as Netscape. The interface that the user sees communicates with a server-side program that initiates a LISP process.

Summary

The goal of the B2 project is to give students an opportunity to practice their decision making skills where the primary modality of interaction is English. We give students the opportunity to ask the system to explain what factors were most influential to its decision and why.

The natural language processing and knowledge representation components of B2 are general purpose. It builds a five-level model of the discourse, that represents what was literally said, what was meant, and how each utterance and its interpretation relates to previous ones. This is necessary because students' utterances may be short and ambiguous, requiring extensive reasoning about the domain or the discourse model to fully resolve. We have shown how our mixed-depth representations encode syntactic and conceptual information in the same structure. This allows us to defer any extensive reasoning until needed, rather than when parsing. We use the same representation framework to produce a detailed representation of requests and to produce a representation of questions. The representations use the same knowledge representation framework that is used to reason about discourse processing and domain information—so that the system can reason with (and about) the utterances.

Acknowledgements

This work was supported by the National Science Foundation, under grants IRI-9701617 and IRI-9523666 and by a gift from the University of Wisconsin Medical School, Department of Medicine.

References

Ali, S. S. (1994a). A Logical Language for Natural Language Processing. In *Proceedings of the 10th Biennial Canadian Artificial Intelligence Conference*, pages 187–196, Banff, Alberta, Canada.

Ali, S. S. (1994b). *A "Natural Logic" for Natural Language Processing and Knowledge Representation*. PhD thesis, State University of New York at Buffalo, Computer Science.

Grosz, B. J. and Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational Linquistics*, 12.

Haddawy, P., Jacobson, J., and Jr., C. E. K. (1996). An educational tool for high-level interaction with bayesian networks. *Artificial Intelligence and Medicine*. (to appear).

Haller, S. (1996). Planning text about plans interactively. *International Journal of Expert Systems*, pages 85—112.

Hirst, G. and Ryan, M. (1992). Mixed-depth representations for natural language text. In Jacobs, P., editor, *Text-Based Intelligent Systems*. Lawrence Erlbaum Associates.

Maida, A. S. and Shapiro, S. C. (1982). Intensional concepts in propositional semantic networks. *Cognitive Science*, 6(4):291–330. Reprinted in R. J. Brachman and H. J. Levesque, eds. *Readings in Knowledge Representation*, Morgan Kaufmann, Los Altos, CA, 1985, 170–189.

McRoy, S., Haller, S., and Ali, S. (1997). Uniform knowledge representation for nlp in the b2 system. *Journal of Natural Language Engineering*, 3(2).

McRoy, S. W. (1995). Misunderstanding and the negotiation of meaning. *Knowledge-based Systems*, 8(2–3):126–134.

McRoy, S. W. and Hirst, G. (1995). The repair of speech act misunderstandings by abductive inference. *Computational Linguistics*, 21(4):435–478.

Shapiro, S., Rapaport, W., Cho, S.-H., Choi, J., Feit, E., Haller, S., Kankiewicz, J., and Kumar, D. (1994). A dictionary of SNePS case frames.

Shapiro, S. C. and Group, T. S. I. (1992). *SNePS-2.1 User's Manual*. Department of Computer Science, SUNY at Buffalo.

Shapiro, S. C. and Rapaport, W. J. (1987). SNePS considered as a fully intensional propositional semantic network. In Cercone, N. and McCalla, G., editors, *The Knowledge Frontier*, pages 263–315. Springer–Verlag, New York.

Shapiro, S. C. and Rapaport, W. J. (1992). The SNePS family. *Computers & Mathematics with Applications*, 23(2–5).