

UNDERSTANDING MEDICAL JARGON AS IF IT WERE A NATURAL LANGUAGE

Alain Bonnet

Heuristic Programming Project, Computer Science Department
Stanford University, Stanford CA 94305

This paper presents BAOBAB-2, a computer program built upon the MYCIN system [11], that is used for understanding medical summaries describing the status of patients. Due to the stereotypic way the physicians present medical problems in these summaries in addition to the constrained nature of medical jargon, these texts have a very strong structure. BAOBAB-2 takes advantage of this structure by having a model of this organization as a set of related *schemas* that facilitate the interpretation of these texts. Structures of the schemas and their relation to the surface structure of the text are described. Issues relating to selection and use of these schemas by the program during the interpretation are discussed.

1 Introduction

Early works on memory [1] and numerous psychological experiments [5] have suggested that people hearing a story make assumptions that they might revise or refine as more information comes in to confirm or contradict them. Making such assumptions entails building (or retrieving) models of the expected text contents. A corollary of this process is that, if the story adequately fits the model people have in mind, the story will be understood more easily.

Although it is difficult to give a formal definition of what constitutes a coherent text, we have an intuitive notion that sentences that compose it form episodes that are linked by different kinds of relationships, cause-effect, chronological orderings and the like. Medical summaries are highly structured; their structure can be described in terms of topic. An important problem is recognizing the different topics and deciding when a "shift of topic" occurs.

2 Related work and goals

AI research has recently explored strategies to recognize shifts of topic occurring during dialogs or written texts. The two main issues faced in doing so are the necessity to narrow down the space of possible referents of a linguistic object and the risk of combinatorial explosion. Thus, Grosz [6] studied the role of focus in the interpretation of utterances and its relation to domain structure. She used the task structure to resolve definite noun phrases in task-oriented dialogs. Sidner [12] extended this work to determine the use of focusing in the resolution of pronoun references and other kinds of anaphors occurring in dialogs. De Jong (in [10]) has used "scripts" [9] to study another kind of structured texts, such as accident reports.

All of this work and the present study have a common feature: they do not interpret sentences in isolation, but rather in the context of an ongoing discourse and hence they use *discourse structure*. BAOBAB-2 also explores issues of (1) what constitutes a model of highly-structured texts and (2) how and when topic shifts occur. The rationale to study schema shifts in the present work are twofold: first, it enables the program to narrow down the space of possible interpretations of inputs and second, it capitalizes on the coherence of texts, which is mainly a task of detecting anomalies, asking the user to clarify vague pieces of information or disappointed expectations, suggesting omissions.

* This research was supported by the Advanced Research Projects Agency under contract DAHC 15-73-C-0435 and a grant from l'Institut de Recherche pour l'Informatique et l'Automatique.

The domain of application is the medical summaries for which processing so far [8] has mainly consisted of filling in formatted grids and has not exhibited any interactive behavior. The program objective is to understand summaries about meningitis cases typed in "medical natural jargon" by a physician, interacting with her or him either to ask questions or to display what it has understood.

The program utilizes a model of what medical summaries typically look like to guide the comprehension. This model consists of a set of related schemas (structures containing domain-specific knowledge which resemble "frames" [7] or "scripts" [9]) described below. For example, it knows that there is a main character who is the patient. This patient presents symptoms. He is admitted to the hospital. A physician observes signs. Some exams are performed, cultures are taken and eventually results are obtained. The physician describes the status of a patient. The program uses both its medical knowledge and its model of the usual description of a medical case in order to produce an internal structure usable by MYCIN, which then attempts to make a diagnosis.

The program behaves like a clerk or a medical assistant who knows what the physician has to describe and how a malady is ordinarily presented. It reacts to violations of the model, such as a description ignoring symptoms or the mention of a culture that has been drawn but for which no result is ever given. It does not attempt to use its knowledge to infer any diagnosis but in certain cases can draw inferences that will facilitate MYCIN's task. It uses these to establish relationships between the concepts stated in order to interpret what is said; for example, it knows that "semi-coma" refers to the state of consciousness of the patient and "hyperthyroidism" to a diagnosis. A potential use of the program is to allow the physician to volunteer information before or during the consultation thereby decreasing her/his frustration at having to wait until asked by the system to mention a crucial symptom.

BAOBAB-2 is comprised of: (a) a parser mapping the surface input into an internal representation; (b) a set of schemas, representing a model of the kind of information it is ready to accept and the range of inferences it will be able to draw; (c) episode-recognition strategies, making it possible to focus on particular pieces of the texts and (d) a generator of English used to display in a non-ambiguous fashion what has been understood. This paper will emphasize on the episode-recognition strategies. Detail on the other parts can be found in Bonnet [2] and [3]. These techniques have been successfully implemented using INTERLISP in a program connected with MYCIN's data base, running on the DEC KA-10 of the medical center at Stanford University.

3 Text and model.

As noted earlier, medical summaries have a stereotypic structure. They can be viewed as a sequence of *episodes* that correspond to phrases, sentences or groups of sentences dealing with a single topic. These topics constitute the model and are represented by schemas. Understanding the content of an episode leads to building one or several internal clauses referring to the same schema. In other words, processing and understanding a text consists of mapping episodes in the text onto schemas that constitute the model. Matching a schema can be "discontinuous", that is, two episodes referring to the same schema are not necessarily adjacent in the text (they might be separated by an episode referring to another schema). We will refer to this phenomenon as a temporary schema shift.

A typical scenario is as follows: The medical case is introduced by giving general information such as the date and the reason for admission to the hospital. Then the patient is presented (name, age,...). Symptoms (noted by the patient) and signs (observed by the physician) are described. A physical exam is usually performed and cultures are taken for which results are pending or available. In the latter case, they are given in detail. The structure of such a text can be captured in a sequence of schemas as shown on Figure 2.

```

[ $SYMPTOMS
  (SYMP legalvals (CHILL HEADACHE PHOTOPHOBIA....)
    tobefilled T )
  (TEMP expect (BETWEEN 94 108)
    whenfilled INFERFEVER)
  :
  (STATE-OF-CONSCIOUSNESS expect (ALERT OBTUNDED..)
    default ALERT)]
  
```

Figure 1: Part of a schema.

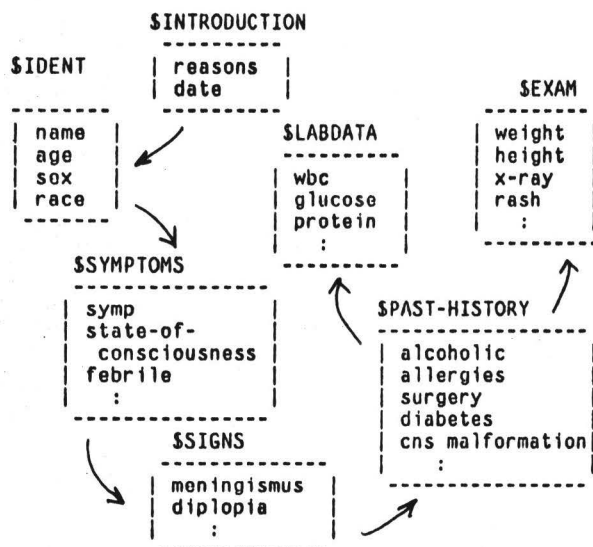


Figure 2: Set of schemas
(Arrows indicate sequencing preferences.)

4 Strategies to detect new episodes.

Top-down vs. bottom-up: Sometimes the schema is explicitly announced, as in "Results of the culture". This is a name-driven invocation of the schema. More often, the instantiation of the schema is content-driven. The clues used are : the attributes associated with the schema, their expected values (if any), and other concepts that might suggest the frame; for example "skin" is related to "rash" which belongs to the *physical exam* schema. These are indeed very simple indices. Research on more sophisticated methods for recognizing the relevant schema, such as discrimination nets, have been suggested by Charniak [4].

Suggest vs. confirm: Sidner [12] makes a distinction between potential and actual shifts of focus, pointing out that the cues suggesting a new frame must be confirmed by a subsequent statement in order to avoid making unnecessary shifts.

We handle this phenomenon in a different fashion. Instead of waiting for the suggestion to be confirmed, a qualitative distinction is made between the slots of a frame. The ones marked as suggesting but not confirming are regarded as weak clues and will not lead to a shift of focus, whereas the ones marked as confirming (hence suggesting) are sufficiently strong clues to command the shift. This distinction can be illustrated by the following two examples.

(1) The patient was found comatose. She was admitted to the hospital. A lumbar puncture was performed. She denied syncope or diplopia...

(2) The patient was found comatose. He was admitted to the hospital. The protein from csf was 58 mg%... (csf = cerebro spinal fluid)

In example 1, the lumbar puncture weakly suggests that "csf results" are available. In example 2, a detail of csf result (strong clue) is given directly; in other words, the physician jumps into detail and the topic "csf results" is directly confirmed.

Termination conditions and actions: A schema is usually terminated when all of its slots have been filled. This however is an ideal situation that does not occur very often. More often, the intervention of a new schema implies the current schema becomes out of focus. A "terminated-by" slot has been created to define which schemas can explicitly terminate others; for example, the \$SYMPTOM schema usually closes the \$IDENTIFICATION schema (name, age, sex, race), as it is very unlikely that the speaker will give the sex of the patient in the middle of the description of the symptoms. This is due to the highly-constrained nature of the domain. When a schema is terminated, the program infers all the default values of the unfilled slots. It also checks whether the expectations set during the story have been fulfilled. These actions can be performed only when a shift has been detected or at the end of the dialogue; otherwise, the program might ask too early about information that the user would indeed give later. In the case where a schema has been exhausted (all its slots filled), an a priori choice is made with regard to the next schema likely to appear. This is performed by using a *preferably-followed-by* pointer that, in the absence of a bottom-up (data-driven) trigger for the next schema, decides in a top-down fashion which one is the most probable at a given point.

Schema-grammar links: The grammar used to parse the text is divided into specific rules and non-specific rules. Specific grammar rules are associated with particular slots in the schemas and describe the way they could be mentioned at the surface level. Categories used in the rules are things like <patient>, <sign>, <diagnosis>. This link between the grammar and the schemas provides a means to try in priority those grammar rules that are appropriate to the schema in focus. The

notion of grammar rules in focus can be viewed as an extension of Grosz's [6] notion of focusing mentioned above. A schema shift thus leads to re-ordering the specific rules.

5 Conclusion.

The strategies outlined above could be applied to a broad range of structured texts. The approach rests on the assumption that their scenario can be seen as sequences of episodes identifiable by the program, in order to be integrated into appropriate schemas. Therefore, clustering attributes into frame-like structures must make sense in the domain of application. The episodes could simultaneously refer to several schemas, that is, the schemas associated could have slots in common. Furthermore, it should be possible to define partial-ordering links between schemas. Therefore, the more structured the text is, the less unexpected shifts are necessary, and the easier is the interpretation process.

Expert systems usually need some kind of understanding to communicate in natural jargon with various users (expert, consultant, student). The technique described here (breaking down the knowledge into schemas that should correspond to different pieces of texts, associating semantic grammar rules with the schemas, and using strategies for recognizing episode shifts) should be applicable in such domains.

6 Example of a session

The physician types the input in upper case after the double asterisk; BAOBAB-2's responses are in lower case. Generation of what the program has understood is somewhat "heavy" for the main purpose is to be unambiguous. Some passages have been cut off for lack of space. A few comments have been inserted on the right for sake of illustration. Names of schemas start with a \$.

**** REASONS FOR ADMISSION: CHILLS HEADACHE MALAISE**
(sets up expectations)

Ok, I will expect these symptoms
**** DATE OF ADMISSION: DECEMBER 25 1978**
My understanding is that:
the patient was admitted at the hospital 4 days ago
**** PAT IS A 26 YEAR OLD CAUCASIAN FEMALE**
My understanding is that:
the age of the patient is 26 years
the race of the patient is Caucasian
the sex of the patient is female
the name of the patient is Pat
**** ONE DAY PRIOR TO ADMISSION SHE EXPERIENCED CHILLS AND HEADACHE**

(shift to \$symptoms)
My understanding is that: (time inferences)
Chill is one of the symptoms that the patient had 5 days ago, headache is one of the symptoms that the patient had 5 days ago
**** ONE DAY AFTER SHE HAD A MALAISE**

My understanding is that:
malaise is one of the symptoms that the patient had 4 days ago

**** A LUMBAR PUNCTURE WAS PERFORMED**
My understanding is that:
a lumbar puncture has been performed on the patient
**** SOCIAL HISTORY: OCCASIONAL ALCOHOL USE**
Do you suspect recent alcoholic history in Pat?

(clarification)
**** NO**
My understanding is that:
the patient is not an alcoholic
**** THE SKIN IS NORMAL**
(leads to a clause via inference)

My understanding is that: the patient does not have a rash or cutaneous lesions
**** NECK: MARKED NUCHAL RIGIDITY**
My understanding is that:
Meningismus is one of the signs that the patient showed (clause via inference)

**** CSF RESULTS** (sets up a context)
Ok.
**** WBC 9500** (understood in context)

My understanding is that:
the WBC from the CSF is 9500
etc.
****WBC FROM CBC: 235**
My understanding is that:
The white count from the patient's peripheral cbc (in thousands) is 235
**** THE VALUE OF WBC IS 225**
(corrects a mistyping)
Instead of 235?

**** YES**
My new understanding is that:
The white count from the patient's peripheral cbc (in thousands) is 225
etc.

7 Acknowledgments

I wish to thank James Bennett, Barbara Grosz, Candy Sidner and Steven Rosenberg for their helpful comments on earlier drafts of this paper.

- [1] Bartlett F., Remembering, a study in experimental and social psychology, Cambridge University press, 1932.
- [2] Bonnet A., Baobab, a parser for a rule-based system using a semantic grammar, Stanford Computer Science Dept, Report STAN-CS-78-668, September 1978.
- [3] Bonnet A., Schema-shift strategies for understanding structured texts in Natural Language, Stanford Technical report, 1979.
- [4] Charniak E., With a spoon in hand, it must be the eating frame, Proc. of Tinlap-2, Urbana-Champaign, Illinois, 1978.
- [5] Collins A., Brown J. S., Larkin K., Inference in text understanding, in Spiro, Bruce and Brewer, Theoretical Issues in reading comprehension, Hillsdale, N. J Lawrence Erlbaum Associates, 1978.
- [6] Grosz B., The representation and use of focus in a system for understanding dialogs, Proc. of 5th IJcai, Cambridge Mass, August 1977.
- [7] Minsky M., A framework for representing knowledge, In P. Winston (ed.) The Psychology of Computer vision. New York McGraw-Hill, 1975.
- [8] Sager N., Natural language information formatting, the automatic conversion of texts in a structured data base, Advances in computers vol.17, 1978.
- [9] Schank R., Abelson R., Scripts, plans and knowledge, in Advance papers of the 4th IJCAI, Tbilisi, USSR, 1975, pp 151-158.
- [10] Schank, R., Abelson, R., Scripts, plans, goals and understanding, an inquiry into human knowledge structures, Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1977.
- [11] Shortliffe E., MYCIN, a rule-based computer program for advising physicians regarding antimicrobial therapy selection, Ph. D thesis, October 1974.
- [12] Sidner C., A computational model of co-reference comprehension in English, Ph. D. Thesis, MIT, Cambridge, Mass, 1979.