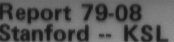

Report 79-08
Stanford -- KSL

Understanding Medical Jargon As If It Were A Natural Language.

Alain Bonnet,

Aug 1979

HPP-79-8

IJCAI BAOBAB-2

UNDERSTANDING MEDICAL JARGON AS IF IT WERE A NATURAL LANGUAGE

Alain Bonnet
Computer Science Department
Stanford University
Stanford California 94305

ABSTRACT

This paper presents BAOBAB-2, a computer program built around MYCIN [Shortliffe, 1974] 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 these structures 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 are described. Issues relating to selection and use of these schemas by the program during interpretation of the summaries are discussed.

Key-words: natural-language comprehension, discourse structure, schemas, knowledge-based systems.

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. Computer facilities were provided by the SUMEX-AIM facility at Stanford University under National Institute of Health grant RR-00785.

1 Introduction

Early works on memory [Bartlett, 1932] and psychological experiments [Collins et al., 1978] 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 is that, if the story adequately fits the model people have in mind, the easier the story will be understood.

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 must be linked by some kinds of causal-consequences relationships, chronological orderings... Flashbacks are not contradictory with coherence but they certainly make the text more difficult to comprehend. The models seem to be structured in terms of topic. Consequently, an important problem to face is recognizing the different topics and deciding when a "shift of topic" occurs.

2 Related works and goals

AI research recently explored strategies to recognize shifts of topic occurring during dialogs or written texts. One issue faced in doing so can be the necessity to narrow down the space of possible referents of a linguistic object by focusing at different levels of detail. Thus, [Grosz, 1977] studies focus of attention to resolve definite noun phrases in task-oriented dialogs. The task structure is represented in a goal tree. The appropriate referent of "bolt" in "give me the bolt" can be determined by looking through the tree. [Bullwinkle, 1977] is concerned with pronoun references and other kinds of anaphoras occurring in dialogs about meeting scheduling. [Rosenberg, 1977] studies how themes were linked through references in newspaper articles.

The concern expressed by [Charniak, 1978] is how to choose among competing frames in order "to avoid death by combinatorial explosion." His fear for such an explosion is mainly motivated by the large amount of inferences possibly drawn if all the possible frames were activated whereas some of them might rule out others, thus enabling the space of possible inferences to be pruned. The issue of choosing the top frames is not faced here.

All the works mentioned so far (and this one too) have a common feature: they do not interpret sentences in isolation, but rather use a discourse structure. BAOBAB-2 also

explores issues of (1) what constitutes a model 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 facilitates the consideration of 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.

The domain of application is the medical summaries for which processing so far [Sager, 1978] has mainly consisted of filling in formatted grids and has not exhibited any interactive behavior. The program objective is to understand a summary 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, which guides the comprehension. This model consists of a set of related schemas 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 speaker is a physician who is to describe the status of a patient. The hearer is the program, which uses both its medical knowledge and its model of the usual description of a medical case. The program must interpret the dialogue or the text and produce an internal structure usable by MYCIN, which will then attempt 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. This decreases her/his frustration at having to wait until asked 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; (d) a generator of English used to display in a non ambiguous fashion what has been understood. The generator was previously existent in MYCIN and has already been described in [Shortliffe, 1974] The main emphasis here will be put on the third feature. 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 on to schemas that constitute the model. Matching a schema can be "discontinuous", that is, two episodes referring to the

same schema are not necessarily juxtaposed (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. These texts are usually well structured or at least coherent, that is redundancies can appear but discrepancies are rather rare (if there are any, they must be detected); expectations are usually satisfied. "Deixis" issues, that is, how the speaker is related to the story, are by and large simple, "I" clearly refers to the physician, "she" or "he" to the patient or another physician.

4 Schemas and their relations.

The MYCIN parameter structure has been taken as a basis for the schemas. Attributes that are not used by the consultation program have been added in collaboration with physicians. Further development of MYCIN might lead to taking them into consideration, but in order to do so, a connection with the present MYCIN's knowledge would be necessary. So far, the extension has been oriented toward the comprehension of summaries dealing with meningitis cases.

A schema is a frame-like structure [Minsky, 1975] Information associated with slots are expected values, default values, attached procedures. It is currently a subset of possibilities that can be found in frame representation languages such as KRL [Bobrow and

Winograd, 1977] Thus, attributes relating to the same topic are gathered into schemas. There is some overlap between them, such as "weight" which can account for the identification of the patient as well as data of a physical exam.

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

Figure 1: Part of a schema.

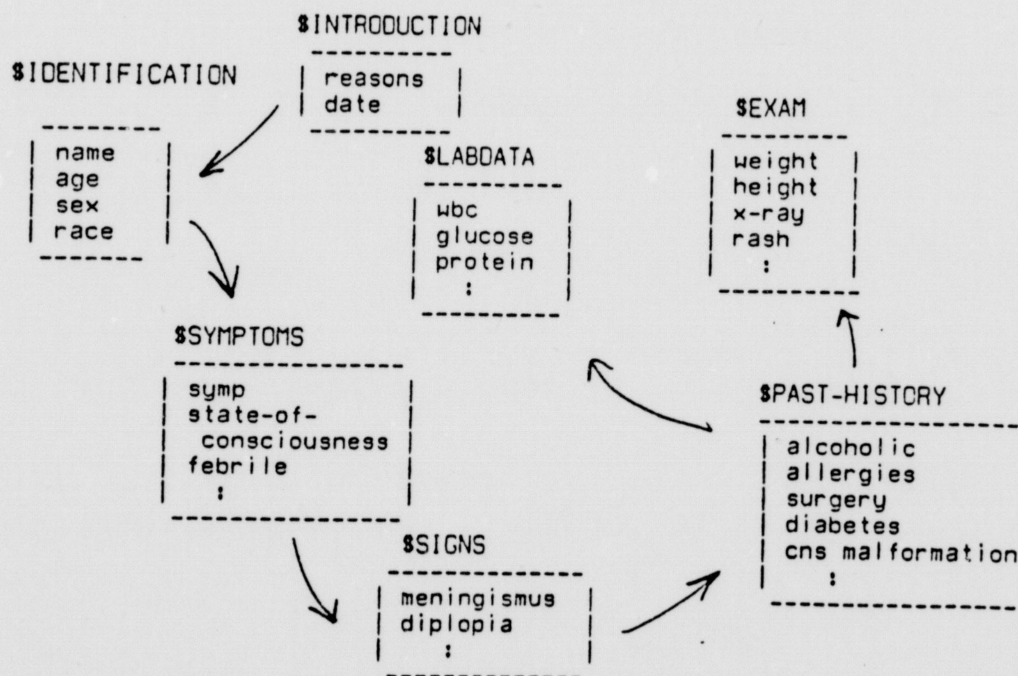

Figure 2: Set of schemas
Arrows indicate sequencing preferences.

Example of the internal representation of an episode:

"... the temperature went up to 103 and he
observed weakness in his legs."

This statement is an episode that corresponds to the $SYMPTOMS schema. Two internal clauses are built out of it. The first requires an inference in order to be used by MYCIN. This inference is performed by the attached procedure INFERFEVER (Figure 1.):

(TEMP 103) ----> (FEBRILE YES).

The second is straightforward: (SYMP WEAKLEGS). Note that if no state of consciousness is mentioned, it will be inferred that "the patient is alert" (by default). Furthermore, at least one symptom should be given somewhere.

5 The grammar.

In a technical domain where specialists write for specialists, terseness of style is so widespread ("T 101.4 rectal") that a syntactic parsing does not provide enough additional information to justify its utilization to comprehend texts in such a domain. [1] Instead, a computer program can use a semantically oriented grammar. This makes the parsing process unambiguous and therefore very efficient. Justifications can be found in [Burton, 1976].

The parser uses a context-free augmented grammar, "augmented" having the same meaning as in "augmented transition network" [Woods, 1970]. A grammar rule specifies the syntax, a semantic verification of the parse tree resulting from the syntactic component and a response expression leading to build one or several clauses. The grammar is an extension of the one described in [Bonnet, 1978]. It is divided into specific rules and non specific rules .

Specific grammar rules are associated with the slots of schemas and describe the

---

[1] This mainly occurs because subtleties gained from syntax are not represented, given the overall purpose of the system.

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. Whenever a specific rule is successful, the input is already semantically interpreted. Consequently, building the corresponding internal clause is fairly straightforward.

Non specific grammar rules use general concepts such as <attribute>, <object>, <value>, commonly used to represent knowledge in systems. This kind of rule is general enough to be used in other domains, but once the syntax has been recognized, they do need a semantic check in order to make sure that, say, values and attributes fit together, hence the importance of the "augmentation" of the context-free grammar.

Specific grammar rules enable the system to recognize very peculiar constructions. For example, 120/98 and 98F do not belong to well-known systactic classes but have to be recognized as values for blood pressure and temperature. Grammar rules such as:

```
<VITAL> -----> <BP> <HIGH/LOW>
<VITAL> -----> <TEMP> <TEMPNUM> / <TEMP> <NUM> (DEGREES)
```

are used to parse " *BP 130/94*" *or* "*T 98F* ". The category <TEMPNUM> has an attached procedure, a specific piece of code that recognizes F as Farenheit, detaches it from 98, verifies that 98 is a reasonable value for a temperature, and finally returns 98 degrees as the value of the temperature.

6 Strategies to detect new episodes.

It seems to me that a language describing choices between frames and therefore frame-shifts strategies, should include an attempt to answer the following questions.

How is a schema foc[illegible]ed, confirmed, abandoned? What are the links between them such as exclusive or sequencing relations?

6.1 Suggest vs. confirm

[Bullwinkle, 1977] 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.

This phenomenom is handled in a different fashion here. 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 suggests "csf results" that are not given (weak clue). In example 2, a detail of csf result (strong clue) is given directly; in other words, the physician jumps into detail and the frame is directly confirmed.

6.2 Top-down vs. bottom-up

Sometimes the schema is explicitly announced, as in "and now the 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* frame. These are

indeed very simple indices. Research on more sophisticated methods for recognizing the relevant schema, such as discrimination nets, have been suggested in [Charniak, 1978].

6.3 Termination conditions

A simple case in which a schema can be terminated is when all its slots have been filled. This is an ideal situation that does not occur very often. Another case is when the intervention of a schema implies that another schema is out of focus. This could be the result of chronological succession, but not necessarily. In general, this phenomenon occurs when the speaker actually starts the plot after setting the personnages of the story. There is no standard way to decide when the setting is finished. However, as soon as the story actually starts, the setting could be closed and possibly completed with default values or with the answers to questions about whatever was not clear or omitted. 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.

6.4 Termination actions

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; 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 with regard to the next schema likely to appear is made. This is possible 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.

6.5 Schema-grammar links

Grammar rules are ranked in order of decreasing likelihood of focus, that is, the specific rules associated with the current schema are tried first, and then the ones associated with its preferred follower, etc. The so-called general rules are the last tried. A schema shift thus leads to re-ordering the specific rules. Whenever an input has been parsed, the shift strategies are invoked. Therefore, the more structured the text is, the less shifts are necessary, and the easier is the interpretation process.

6.6 Comparison

Other methods have been proposed to take advantage of coherent structures of texts. Story grammars suggested by [Rumelhart, 1975] entail too rigid constraints with regard to the categories (episodes) intervening within the rules. For example, a given action is not necessarily followed by its consequence (though it would be very convenient for comprehension), as an adjective is followed by a noun. A flashback make's understanding more difficult but cannot be considered as anomalous (ungrammatical). The model-based approach enables more flexibility with regard to schema sequencing because it merely proposes a partial ordering. Bottom-up and top-down strategies can thus be mixed.

7 Level of comprehension

The combined action of the grammar and the model composed of related schemas provides the system with a certain level of comprehension that can be characterized by the following features:

Making explicit what is implicit by using default values, for example, the date of a culture will be assumed to be the date of admission if not specified.

Clarifying vague pieces of information: "The patient drinks 6 cans of beer every morning " will lead to ask "*is the patient alcoholic?*" since the system has evidently no knowledge about alcohols.

Ignoring things: Usually this includes any information which is not handled by MYCIN. It mainly occurs when pieces of text does not mention any attribute, value or clue indicating a link with a frame, for example "the patient's father is a drapery hanger".

Understanding by using context: "P 65" by itself has no obvious meaning but appearing with results of physical exam, the meaning "the pulse is 65" is clear.

Drawing inferences: Many inferences result from running procedures attached with slots. A first kind of inference is from a numeric value to adequate conceptual information usable by the system, for example, "the temperature of the patient is 39 degrees" will imply that "the patient is febrile".

Detecting inconsistencies: "Weight 255 kgms" will cause the program to prompt: "do you really mean that the weight of the patient is 255 kgms?"

The program also deals with time inferences like in "two days after the admission" which is interpreted as "two days after the date of admission".

A limited set of references (and co-references) can be understood, mainly references by value as below in (1), references via simple inferences as in (2), or contrastive references as in (3).

1) *The patient had malaise and chills. Two days after the symptoms worsened*

2) *The temperature went up to 38 degrees... the fever stopped.*

3) *Cultures were drawn from urine and blood. The first culture turned positive, the second culture was negative.*

8 Direction for future work

For the time being, the grammar is not very large (about 200 rules); Only seven schemas have been implemented. In order to be able to parse efficiently more complicated texts involving symptoms which might imply different infections (with interactions between them), a computer program will need more sophisticated clues to determine which schema is the most appropriate to apply in priority. Furthermore, certain concepts are currently ignored because their relevance to medical knowledge is not always straightforward. For example, an infection acquired by a member of the patient's family or even the patient's occupation could give hints to the physician during the early stage of the diagnostic formation.

9 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 them. The relationships could be rather loose but the more constrained they are, the better this feature would work.

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.

10 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. An actual summary written by a physician has been chosen, slightly simplified and broken down into inputs and given as a dialog to facilitate the interactions. 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 $.

```
** CASE 45                                                  ($introduction)
I am ready for case 45
** REASONS FOR ADMISSION: CHILLS HEADACHE MALAISE      (sets up expectations)
So I will expect chill headache malaise as values for symptoms
```