The numeric representation of knowledge and logic—Two artificial intelligence applications in medical education

by W. D. Hagamen
M. Gardy

MEDCAT (medical diagnosis, consultation, and teaching) is a program that makes diagnoses from empiric data stored in patient records, explains its reasoning in response to questions (consultant mode), and uses its logical and communicative skills to instruct medical students in the proper approach to medical diagnosis (student mode). MEDCAT's reasoning can be modified by free-format discussion with physicians. CATS (computerized anatomical teaching system) is an entirely separate program designed to teach gross anatomy. Like MEDCAT, it has a consultant mode that the student may use to explore the program's reasoning, and a student mode in which the program takes the initiative. A prominent feature of CATS is its ability to discover meaningful general principles that reduce the need for memorization. Despite important differences in the subject matter, the data structure and code are very similar in the two programs. Both use a powerful natural-language interface that parses the input and generates the output.

Perhaps the most important characteristic of intelligent expert systems is the ability of the program to explain its reasoning. If a program were designed to print out medical diagnoses as a black box, this would neither impress nor be useful for physicians. However, if it could answer questions such as "Why did you make that diagnosis?" "What other diagnoses did you consider?" "Why did you consider them and why did you reject them?" "Why did you do that test and not do another one?" then it would begin to become believable. If the user could disagree with its conclusions and perhaps even win or lose the debate, the system would really appear to have some of the attributes by which we evaluate whether a person (and presumably a program) is, in fact, an expert. We like our experts to be articulate and responsive to our questions.

In the real world, few expert systems are competitive yet. Who among us would want the diagnostic and therapeutic decisions affecting our own health to be made by a computer? On the other hand, the techniques being developed are of theoretical interest, not only to computer people, but also to cognitive psychologists and physicians. Also, programs like this are fun to work on. One always has the feeling that, if one put in just a little more data and a little more logic, it really would be as good as its human counterpart.

One area in which such expert systems are very useful, but one that still has received little attention, is the realm of education. The abilities to understand what is being asked, to determine the answer by logical means, to explain the reasoning involved, to evaluate the reasoning of others, to learn from experience, and to discover general principles form the basis of good teaching. Sometimes the tendency of such programs to deal in generalities at the expense of details may be a liability in many real-world applications; in education, however, this becomes an asset.

This paper describes two programs used in the instruction of medical students. In a sense, internal medicine and gross anatomy represent opposite extremes within the spectrum of medical education. Medical diagnosis (MEDCAT)[1] is heavily dependent on deductive reasoning. Gross anatomy (CATS)[2] is a descriptive discipline that utilizes inductive reasoning to reduce the need for memorization by organizing things into logical groups. Despite these important differences, the data structure and the code are very similar. In both instances the basic data consist of (1) a character matrix containing the names of the things the program needs to talk about and (2) a three-row numeric matrix containing the subject matter logic.

The most important decision in designing an expert system may well be the choice of method for the internal representation of knowledge and logic.

The purpose here is to describe these two programs, to show how a numeric array can concisely store the knowledge and efficiently perform the reasoning required in these two diverse disciplines, and to describe a natural-language system that interfaces with this knowledge and logic base in a very natural way. An important feature that results from the interaction with the logical model is the ability of the language interpreter to resolve ambiguity based on the *context* of the discussion.

The knowledge and logic are stored external to the code. The programs are coded in APL and implemented on both 8086- (IBM PC/AT) and 68000-based microcomputers.

Representation of data

The most important decision in designing an expert system may well be the choice of method for the internal representation of knowledge and logic, such that comprehension, reasoning, and expression can interface with it in an intuitively natural way. The basic data in both CATS and MEDCAT consist of the names of things within their domains of knowledge (called *nodes*) and the relationships between these nodes (called *pointers*).

Node descriptors. The node descriptors are noun phrases, which may be single words or groups of words, and may include synonyms enclosed in parentheses. These are stored in a character matrix. The noun phrases are used by the program when it generates its narrative output. This character matrix is also used by the subroutine as it originally creates the vocabulary lists used in parsing the input. Consider the following node descriptors for MEDCAT and CATS, and note the presence of synonyms in parentheses:

MEDCAT

```
 84 CHOLESTATIC JAUNDICE (OBSTRUCTIVE)
144 DYSPNEA (SHORTNESS BREATHING DIFFICULTY)
252 LIGHT COLORED STOOLS (PALE CLAY)
342 PROTHROMBIN TIME
492 VITAMIN K REVERSAL
```

CATS

```
 20 AXILLARY ARTERY
 55 DEEP VOLAR BRANCH OF THE ULNAR NERVE
164 MEDIAL CORD
293 THIRD PALMAR INTEROSSEOUS (3 VENTRAL IO
I.O.)
318 ULNAR NERVE
```

Thus, for example, the program uses DYSPNEA when generating text for node 144, but it also recognizes "difficulty in breathing" and "shortness of breath" on the input side. The numbers represent the relative position of each node descriptor within the list, i.e., the row number for that element. The program uses this index number, not the character data, in its logical representation.

The numeric matrix. The relationships between nodes are stored in an integer matrix consisting of three rows. The first row contains the number of the node from which the pointer originates. The third row indicates the destination node. The middle row contains the signed integer value of the pointer.

MEDCAT and CATS differ in the way these pointer values are utilized. In MEDCAT they primarily indicate the *strength* of the relationship. High pointer values going to a node indicate strong reasons for considering a diagnosis or ordering a test. Negative pointers indicate reasons arguing against the same. In CATS pointer values indicate the *nature* of the relationship, such as "is a branch of," and "innervates."

From:	84	84	84	252	342
Value:	350	−1000	400	50	1
To:	171	226	239	84	492

The preceding table shows five selected columns from the pointer matrix for MEDCAT. The presence of CHOLESTATIC JAUNDICE (84), as an intermediate diagnosis, suggests the possibility of both EXTRAHEPATIC (171) and INTRAHEPATIC (239) OBSTRUCTION. It argues strongly against the presence of HEPATOCELLULAR JAUNDICE (226). If LIGHT COLORED STOOLS (252) were present, this would be additional evidence to support CHOLESTATIC JAUNDICE. The very weak pointer from elevated PROTHROMBIN TIME (342) to the VITAMIN K REVERSAL test (492) is a convention used to indicate reasons for doing a test. Such weak pointers do not influence the firing of a node because all empiric data are determined solely by the patient record. However, weak pointers do permit the program to determine the justification and proper sequence for doing tests. The following pointer values in CATS indicate the nature of the relationship:

From:	55	55	164	318
Value:	4	3	10	3
To:	293	318	20	164

The ULNAR NERVE (318) is a branch (3) of the MEDIAL CORD (164). If the syntax of a question were "What are the branches of the ulnar nerve?" the matrix would be interpreted in the opposite direction; i.e., the subject (ULNAR NERVE) would be sought in row 3 and the answer (DEEP VOLAR BRANCH) in row 1. The DEEP VOLAR BRANCH OF THE ULNAR NERVE (55) innervates (4) the THIRD PALMAR INTEROSSEOUS (293) (in addition to many other structures). If CATS needed to know which muscles the MEDIAL CORD innervates, it would make multiple passes through the pointer matrix. From this small subset of the matrix it can determine that the MEDIAL CORD (164) branches (3) into 318, which branches into 55, which innervates (4) 293. The MEDIAL CORD (164) is anterior (10) to the AXILLARY ARTERY (20).

Two sample dialogues

We now show selected examples from a question-and-answer sequence from both MEDCAT and CATS. They are presented in one place to give a feeling for the flow of dialogue. Numbers are attached so we may refer to specific examples in subsequent sections. For ease of reading, what the computer prints is shown in uppercase, and what the user types is in lowercase. Actually, all interactions are carried on in uppercase, which is the normal APL character set. These question-answer examples are referenced by number throughout this paper.

The following are examples from MEDCAT.

1. WHAT IS YOUR QUESTION?
   What was your diagnosis?
   ALCOHOLIC HEPATITIS AND ALCOHOLIC CIRRHOSIS.

2. WHAT IS YOUR QUESTION?
   What was the evidence for alcoholic cirrhosis?
   THE PRESENCE OF PAROTID GLAND ENLARGEMENT IS CONSISTENT WITH AND THE HISTORY OF EXCESSIVE ALCOHOL INTAKE AND THE PRESENCE OF CIRRHOSIS STRONGLY SUGGEST, ALCOHOLIC CIRRHOSIS.

3. WHAT IS YOUR QUESTION?
   How did you make the diagnosis of cirrhosis?
   THE ANOREXIA, NAUSEA AND VOMITING, THE PRESENCE OF PALMAR ERYTHEMA, GYNECOMASTIA AND ENLARGED LIVER, THE ELEVATED SERUM GLOBULIN AND PROTHROMBIN TIME AND THE SLIGHTLY ELEVATED SGPT AND SGOT ARE CONSISTENT WITH, THE PRESENCE OF SPIDER ANGIOMAS SUGGESTS AND CHRONIC HEPATOCELLULAR DISEASE AND THE COMBINATION OF HEPATIC ENCEPHALOPATHY AND ASCITES STRONGLY SUGGEST, CIRRHOSIS.

4. WHAT IS YOUR QUESTION?
   How elevated was the SGPT?
   THE SGPT WAS 85 U/L.

5. WHAT IS YOUR QUESTION?
   What about the GGTP?
   THE GGTP WAS 240 U/L.

6. WHAT IS YOUR QUESTION?
Did you think of cholestasis?
CHOLESTASIS WAS CONSIDERED, BUT REJECTED.

7. WHAT IS YOUR QUESTION?
Why did you reject it?
BECAUSE THE SLIGHTLY ELEVATED ALKALINE PHOSPHATASE, THE ABSENCE OF PRURITUS, THE DEPRESSED SERUM ALBUMIN AND THE ELEVATED SERUM GLOBULIN DO NOT SUPPORT CHOLESTASIS.

8. WHAT IS YOUR QUESTION?
What were the SGOT, total serum bilirubin and alk phos?
THE SGOT WAS 180 U/L.
THE TOTAL SERUM BILIRUBIN WAS 8 MG/DL.
THE ALKALINE PHOSPHATASE WAS 145 IU.

9. WHAT IS YOUR QUESTION?
Were there any complications of cirrhosis?
ASCITES AND HEPATIC ENCEPHALOPATHY.

10. WHAT IS YOUR QUESTION?
What if the alkaline phosphatase were markedly elevated?
THE DIAGNOSIS WOULD STILL BE ALCOHOLIC HEPATITIS AND ALCOHOLIC CIRRHOSIS.

11. WHAT IS YOUR QUESTION?
Was there any pain?
YES, RUQ PAIN WAS PRESENT.

12. WHAT IS YOUR QUESTION?
Tell me more about the pain.
IT WAS DULL, CONSTANT, GRADUAL IN ONSET AND DID NOT RADIATE THROUGH TO THE BACK.

13. WHAT IS YOUR QUESTION?
What indicated jaundice?
THE COMBINATION OF DARK URINE AND LIGHT COLORED STOOLS SUGGESTS, THE PRESENCE OF URINE BILIRUBIN STRONGLY SUGGESTS AND THE MARKEDLY ELEVATED TOTAL SERUM BILIRUBIN AND THE COMBINATION OF DERMAL AND SCLERAL ICTERUS CONFIRM, JAUNDICE.

The following are examples from CATS.

14. WHAT IS YOUR QUESTION?
What is the innervation of the biceps?
THE MUSCULOCUTANEOUS NERVE.

15. WHAT IS YOUR QUESTION?
What is its action?
IT FLEXES THE SHOULDER AND ELBOW AND SUPINATES THE FOREARM.

16. WHAT IS YOUR QUESTION?
What are the branches of the axillary artery?
THE AXILLARY ARTERY TERMINATES AS THE BRACHIAL ARTERY.

THE FIRST PART OF THE AXILLARY ARTERY BRANCHES INTO THE SUPREME THORACIC ARTERY.

THE SECOND PART OF THE AXILLARY ARTERY BRANCHES INTO THE LATERAL THORACIC AND THORACO-ACROMIAL ARTERIES.

THE THIRD PART OF THE AXILLARY ARTERY BRANCHES INTO THE ANTERIOR HUMERAL CIRCUMFLEX, POSTERIOR HUMERAL CIRCUMFLEX AND SUBSCAPULAR ARTERIES.

17. WHAT IS YOUR QUESTION?
What does the musculocutaneous innervate?
THE SKIN OF THE LATERAL FOREARM, LONG HEAD OF THE BICEPS, SHORT HEAD OF THE BICEPS, BRACHIALIS AND CORACOBRACHIALIS.

18. WHAT IS YOUR QUESTION?
What anastomoses would develop following an occlusion between the second and third parts of the axillary?
AN OCCLUSION BETWEEN THE SECOND PART OF THE AXILLARY ARTERY AND THE THIRD PART OF THE AXILLARY ARTERY WOULD PRODUCE THE FOLLOWING ANASTOMOSES:

THE DEEP BRANCH OF THE TRANSVERSE CERVICAL ARTERY TO THE LATISSIMUS DORSI TO THE THORACODORSAL ARTERY.

THE DEEP BRANCH OF THE TRANSVERSE CERVICAL ARTERY TO THE SUBSCAPULARIS TO THE SUBSCAPULAR ARTERY.

THE LATERAL THORACIC ARTERY TO THE SUBSCAPULARIS TO THE SUBSCAPULAR ARTERY.

THE SUPERFICIAL BRANCH OF THE TRANSVERSE CERVICAL ARTERY TO THE INFRASPINATUS TO THE CIRCUMFLEX SCAPULAR ARTERY.

THE SUBSCAPULAR ARTERY TO THE INFRASPINATUS TO THE CIRCUMFLEX SCAPULAR ARTERY.

THE THORACO-ACROMIAL ARTERY TO THE ANTERIOR PART OF THE DELTOID TO THE ANTERIOR HUMERAL CIRCUMFLEX ARTERY.

THE THORACO-ACROMIAL ARTERY TO THE ANTERIOR PART OF THE DELTOID TO THE POSTERIOR HUMERAL CIRCUMFLEX ARTERY.

THE THORACO-ACROMIAL ARTERY TO THE LATERAL PART OF THE DELTOID TO THE ASCENDING BRANCH OF THE DEEP BRACHIAL ARTERY.

THE THORACO-ACROMIAL ARTERY TO THE LATERAL PART OF THE DELTOID TO THE POSTERIOR HUMERAL CIRCUMFLEX ARTERY.

19. WHAT IS YOUR QUESTION?
Why does the pronator teres flex the elbow?
ALL MUSCLES THAT ORIGINATE FROM THE MEDIAL EPICONDYLE OF THE HUMERUS FLEX THE ELBOW JOINT.

ALL MUSCLES THAT PASS ANTERIOR TO THE ELBOW JOINT FLEX IT.

20. WHAT IS YOUR QUESTION?
What is the innervation of the first dorsal interosseous?
THE DEEP VOLAR BRANCH OF THE ULNAR NERVE.

21. WHAT IS YOUR QUESTION?
Why?
ALL MUSCLES THAT ARE PART OF THE HAND ARE INNERVATED BY THE DEEP VOLAR BRANCH OF THE ULNAR NERVE, EXCEPT THE THENAR GROUP, 1/2 THE LUMBRICALS, AND THE PALMARIS BREVIS.

ALL MUSCLES THAT INSERT ON THE LATERAL BAND ARE INNERVATED BY THE DEEP VOLAR BRANCH OF THE ULNAR NERVE, EXCEPT 1/2 THE LUMBRICALS.

ALL MUSCLES THAT PASS ANTERIOR TO THE METACARPOPHALANGEAL JOINTS ARE INNERVATED BY THE DEEP VOLAR BRANCH OF THE ULNAR NERVE, EXCEPT THE FLEXOR DIGITORUM PROFUNDUS, FLEXOR DIGITORUM SUPERFICIALIS, AND 1/2 THE LUMBRICALS.

ALL MUSCLES THAT EXTEND THE INTERPHALANGEAL JOINTS ARE INNERVATED BY THE DEEP VOLAR BRANCH OF THE ULNAR NERVE, EXCEPT 1/2 THE LUMBRICALS.

Understanding the input

As you read this paper you are somehow decoding the input and translating it into a form in which you can efficiently process and store it. Just how you do this is not entirely clear. However, certain features of the process are known. (1) You use certain syntactic clues (prepositions, pronouns, helping verbs, conjunctions, and word endings) to break the sentence into functionally useful parts. (2) You can readily recognize verb and noun phrases, even when you may not be familiar with some of the words. (3) The translation takes the form of ideas that are somewhat independent of the exact words that may be used, and these ideas, to the extent you can accept them, fit into predefined slots in your conceptual framework. (4) As the processing evolves, you develop certain expectations of what should come next, and you use this anticipation to decide which interpretations make sense. This eliminates ambiguities.

CATS and MEDCAT parse user questions by means that are nearly identical.

The program makes use of these four techniques.

Parsing the input. CATS and MEDCAT parse user questions by means that are nearly identical to each other. The input sentence is first broken up into its individual words. Each of these words is then compared against a vocabulary list. The vocabulary is created automatically by a utility program that extracts all the words that comprise the node descriptor matrix. It also records the row numbers (nodes) in which the word was found. This vocabulary is then catenated to a predefined list of (approximately 128) special words (syntactic markers). Each of these special words has an associated decimal value (called its word type) that reflects its role within our rules of discourse. These tagged words include interrogative pronouns, helping verbs, prepositions, pronouns, articles, conjunctions, and punctuation. Negating words, such as not, no, against, and reject, are given a negative sign (word type). Node descriptor words

A prime purpose of the parsing is to help identify which nodes the user is asking about.

not found in the list of special words are assigned a value of zero, as are words not found in the vocabulary.

Anyone who has read earlier descriptions of these programs may note that the program-defined vocabulary list is a recent addition. We used to search the entire node-descriptor list. There were three reasons for the change. (1) The list of node descriptors began to exceed the maximum size (32K) permitted for single data objects in our version of APL. Searching only the list of unique words circumvented this problem. (2) It also meant that we did not need to search the entire vocabulary, but could key on the initial character in each word. Thus we have an A-vocabulary, a B-vocabulary, and so forth. This slightly decreases the search time. (3) The most important reason relates to misspelled words. When a given word fails to match, we know which word is involved. Previously, such information was not available.

In any event, it is important to realize that these are not keywords entered by a programmer. They are the naturally occurring words from the node descriptors and are extracted automatically by the program, together with the numbers of the nodes to which they refer. There is no lexical information associated with any of these words. Their word types (0) are assigned by default. The short list of syntactic words, on the other hand, is entered by the programmer and does carry lexical information or nonzero word types (that is its purpose), and these words do not point to specific nodes.

Isolating the noun phrases. A prime purpose of the parsing is to help identify which nodes the user is asking about. Since several nodes may be referred to in a single question (e.g., Q8 and Q18), it is necessary to have a means of identifying individual noun phrases. The parser separates noun phrases by using word types that signal the beginning of a noun phrase (articles and determiners) or, more importantly, word types that flag the end of a noun phrase (such as punctuation and conjunctions). After the noun phrases have been identified, each one is pruned prior to node selection by eliminating all words with nonzero word types.

In the following example, those words found in the syntactic word list are underlined so that you may identify them. The numeric value assigned to each word as a result of this comparison is also shown. The three noun phrases are seen on lines 2–4. The first was separated from the second by the comma, and the second from the third by the AND. The only words involved in determining the node are those not underlined (word type 0). From this point forward all processing is done on this numeric version of the original input.

1. WHAT WERE	1.7	1.5		
2. THE SGOT ,	3	0	9	
3. TOTAL SERUM BILIRUBIN AND	0	0	0	9
4. ALK PHOS ?	0	0	9	

Only the first six characters of each input word are used for the initial comparison. This corrects for many typing errors and also adapts to the fact that students may be ignorant of, or careless about, the exact spelling of Latinized endings (e.g., "interosseous" versus "interosseus"). It may be surprising, but it is empirically true, that 96 percent of the total vocabulary of both MEDCAT and CATS is uniquely defined by the first six characters.

During the string search, input words are preceded by—but not followed by—a space. For this reason, "nucleosis" does not match "mononucleosis," but the abbreviation "mono" does match. This permits users to follow their natural inclination to abbreviate words (Q8).

The purpose of the question. A second reason for the parsing is to determine the nature of the question. This is accomplished by (1) finding the first word type with a floor of 1 (which means it is an interrogative), (2) finding the next word type that has a floor of either 0 (undefined word) or 1 (interrogative) or 7 (a known adjective), and (3) encoding the mantissas of these two word types into a single number (base 10). If the word is a tagged adjective, 0 is substituted for its mantissa. By this means "What were" became encoded as 75; "How (1.9) did (1.6)" becomes 96; and "How (1.9) elevated (7.0) was (1.5)" becomes 90. This encoding scheme provides much flexibility. "Is, are, was, and were," for example, all have the same word type (1.5) and are thereby equated, for our purposes. The encoding permits the program to branch selectively according to the purpose of the question, e.g., whether it is asking the reason for something ("Why" or "What made you"), the value of some empiric data ("What was"), the implications of empiric or diagnostic data ("What did"), or is a true/false-type question ("Was" or "Did"). It also permits the equating of encoded values. "How did" (96) and "How was" (95) are interpreted as "Why." Most other combinations with "How," such as "How elevated" (90) are treated as "What was." When the specific combinations of "How about" (92) or "What about" (72) occur, the program assumes that the question has the same intent (verb) as the previous one (Q4 and Q5).

MEDCAT does not need the exact form of the main verb to understand the question.

Verb analysis. CATS and MEDCAT differ in the way they use verbs. MEDCAT does not need the exact form of the main verb to understand the question. As discussed previously, the purpose of the question is defined by the encoded combination of interrogatives and helping verbs. MEDCAT sees only "What did the SGOT . . . ?" or "Why did you . . . hepatitis-B?" It does, however, make use of negating words, such as "reject, against, not." These have negative word types. The sign of the product of all such negative word types is taken as the sign of the question. If the question is posed in the affirmative, it looks for positive pointers and the absence of negative ones. If the question contains an odd number of negating words, it evaluates negative pointers and the absence of positive ones. During the learning mode, which is described later in this paper, MEDCAT relies on the form of the main verb to assign new pointer values. The quantitative value of the verb, for these purposes, is encoded in the decimal portion of its word type.

In CATS, the verbs represent the nature of the relationship. However, many of these relations are expressed as adjectivals or nominatives, rather than as verbs. At one time, a word may be used to express a relationship, but at another time, the same word may be used as a node descriptor. Fifty percent of the potential verbals exhibit this dual role. The program also must determine which row of the pointer matrix to search for the nodes, depending on their relation to the verb that is identified.

Identifying the functional verb. Potential verbs are flagged by having a word type with a floor of 2. Resolution is required when there is more than one such word in the input. The first test is whether the tagged word has any associated node pointers. Can it be part of a noun descriptor? If it has no node associations, it must be a verb. If the ambiguity remains, two additional tests may be done. (1) If the word is one of the adjectivals (ANTERIOR, POSTERIOR, LATERAL, MEDIAL, INFERIOR, SUPERIOR, DEEP, SUPERFICIAL, DISTAL, PROXIMAL), it must be followed by TO, or must be the terminal word, in order to be acting as a verb. (2) If it is an ambiguous nominative (BRANCH, PART), it must be preceded by a determiner (A, AN, THE), in order to be a verb. This takes advantage of the fact that these words are always preceded by an adjective when they occur in a noun descriptor (CLAVICULAR BRANCH OF, SECOND PART OF). Once the active verb has been identified, the word types of the competing verbals are converted to zero, and they become available to the parser as noun phrase constituents.

Determining the row numbers. Verbals encountered by the program include transitive verbs, intransitive verbs, predicate nominatives, and predicate adjectives. For this reason, it is not particularly useful to think in terms of subject, verb, and object. However, it should be apparent from the nature of the pointer matrices that all questions are *dyadic.* That is, they involve a *relation* between two things (or groups of things). One of these is stored in the first row of the matrix and the other in the third row. It is the program's task to determine which is which.

Some questions have the general form WHAT IS THE ORIGIN OF THE BICEPS? or WHAT IS THE CORACOID

PROCESS THE ORIGIN OF? Clearly, one of the two arguments (nodes) is given in the question. The other is unknown, but is tokenized by the place-holding function of the WHAT. This unknown represents the answer that the program must fill in. In true/false questions, both arguments are given, e.g., IS THE CORACOID PROCESS THE ORIGIN OF THE BICEPS? However, when attempting to answer this type of question, the program temporarily masks the left argument, processes the remainder as a WHAT-type question, and then compares its answer with the masked node. In both of these situations, the program needs to know which row contains the given node and which contains the answer.

This is performed by a simple APL function. We next describe the essential features of this function. However, a brief introductory discussion may help you evaluate what is to follow. (1) All verbals, except adjectivals, have an infinitive verb form. For example, ORIGINATE is the infinitive form of ORIGIN OF, and INNERVATE is the verb infinitive for INNERVATED BY. The node that is considered the subject of the verb infinitive is stored in row 1. (2) The adjectivals occur in pairs of logical opposites (e.g., ANTERIOR-POSTERIOR and MEDIAL-LATERAL). These have no verb infinitive equivalents. One half of each pair has a positive word type; its opposite is the negative of the same word type. If you think in terms of a declarative sentence (THE AXILLARY ARTERY IS ANTERIOR TO THE POSTERIOR CORD.), the left argument (AXILLARY ARTERY) of a positive adjectival (IS ANTERIOR TO) is in row 1, and the right argument (POSTERIOR CORD) is in row 3. If the adjectival has a negative word type (IS POSTERIOR TO), the relationship is reversed. In this way the absolute value of these word types performs both roles. There are no negative values in the pointer matrices.

The function that performs this task returns a value (1 or 3) to indicate the row number where the *answer*, as previously defined, is located. The process involves several steps and conditional tests. (1) If the verb is followed by a noun phrase, the answer is assumed to be in row 3; otherwise it is in row 1. (2) This decision is reversed if the verb either has a negative word type (IS POSTERIOR TO) or is the nominative form of an intransitive verb (e.g., ORIGIN OF, INSERTION OF, PART OF, BRANCH OF). The reason for this depends on the transitivity of the verb. In the nominative form of a transitive verb (THE MUSCULOCUTANEOUS NERVE IS THE INNERVATION OF THE BICEPS.), the structure doing the innervating appears before the verb. When the verb is intransitive (THE MUSCULOCUTANEOUS NERVE IS A BRANCH OF THE MEDIAL CORD.), the structure doing the branching follows the verb. (3) If the question contains a possessive, as indicated by the word types, the result is reversed again. ITS INNERVATION is equivalent to INNERVATION OF IT. (4) Passive voice, when present, causes an additional reversal. (5) A final reversal occurs if the verb is immediately followed by certain prepositions that have not already been factored in (e.g., BRANCH FROM versus BRANCH INTO). All decisions are based on the numeric (word type) form of the question that already exists.

If a system justifies its being termed "intelligent," it cannot impose any special rules of grammar or syntax on the user.

Contextual understanding. If a system justifies its being termed "intelligent," it cannot impose any special—even well-accepted—rules of grammar or syntax on the user. The program should be as understanding and flexible as its human counterpart. In order to do this, the program must be able to recognize and use the context of the dialogue. This has always been a glaring deficit of language processing programs. The ways in which this program uses context to help its understanding are illustrated in the following examples.

Role of individual words. Assume the following question: WHERE DOES THE EXTENSOR CARPI RADIALIS BREVIS ORIGINATE? The program extracts a four-word noun phrase, EXTENSOR CARPI RADIALIS BREVIS. Each of these words occurs as part of a number of different node descriptors. The node numbers associated with each word are shown as follows:

EXTENSOR	41 69 70 71 72 73 74 75 76 324 325 326 327 328
CARPI	69 70 71 87 88 315
RADIALIS	33 69 70 87 234
BREVIS	2 69 75 97 197

The only node number that all four words share in common is 69, which, not surprisingly, is the EXTENSOR CARPI RADIALIS BREVIS. The first word in the sequence, no matter which it happens to be, activates (brings to mind) a number of associations. Each succeeding word, by a process of logical anding, makes the meaning more specific. Thus, on the simplest level, the words in the noun phrase provide the context.

Role of the verb in pronoun reference. The pronoun reference seen in Q7, and used extensively, is achieved by storing the node number for the previous question. There is no ambiguity because, in MEDCAT, we are usually talking about one thing (node) at a time. In CATS, however, the discussion usually involves a relation between a subject and an object. To which should the pronoun refer? Consider the sequence in Q14 and Q15. "What is the innervation of the biceps?" THE MUSCULOCUTANEOUS NERVE. "What is its action?" Here the pronoun is interpreted to be referring to the subject of the previous question, not to the answer. However, if the second question had been "What else does it innervate?" both we and the program would have coupled it to the previous answer: THE MUSCULOCUTANEOUS NERVE. This is not dependent on grammatical rules. Rather, it is based on the semantic rule that nerves, not muscles, innervate things.

In this situation the verb provides the context. Every node has an associated *profile* (number) that indicates the type of structure (e.g., 1=artery, 2=bone, etc.). When the program searches the middle row of the pointer matrix for all instances of the number (3) representing the verb INNERVATES, it determines that all the node numbers in row 1 (representing the structure doing the innervating) have a profile of 6, which means they are nerves. We call this the *expected profile.* The BICEPS has a profile of 5 (muscle) and that of the MUSCULOCUTANEOUS NERVE is 6 (nerve). For this reason, the latter node is assigned to the pronoun.

Role of the verb in partial phrases. The profile and expected profiles have generic applications. In this regard, consider the question portion of Q16, which has nothing to do with pronoun reference: "What anastomoses would develop following an occlusion between the second and third parts of the axillary?" This is a very natural way to ask this question. Notice that the program extracts more information than was explicitly provided. It knows that the question refers to the AXILLARY ARTERY, as opposed to the AXILLARY NERVE or AXILLARY VEIN, because it is the only interpretation that would agree with the expected profile. Similar logic indicates that SECOND refers to SECOND PART OF THE AXILLARY ARTERY rather than the SECOND RIB or SECOND DORSAL INTEROSSEOUS.

After discarding the three nonzero word types (OF THE ?), the noun phrase contains three words: THIRD PARTS AXILLARY. The APL function that takes over the selection process from this point is called GNODE. Each word in the vocabulary that has a zero word type has an associated vector of numbers indicating all the node descriptors in which that word was found. The right argument of GNODE is all of these node numbers for each of the (three) words in this phrase. For this particular phrase there are 17 such numbers, which are given in the following list, together with the complete form of each node descriptor:

153 LOWER 2/3 OF ANTERIOR SURFACE OF HUMERUS (TWO THIRDS)
155 LUMBRICAL TO DIGIT 3 (FINGER THIRD)
294 THIRD PART OF AXILLARY ARTERY (3)
84 FIRST PART OF AXILLARY ARTERY (1)
251 SECOND PART OF AXILLARY ARTERY (2)
294 THIRD PART OF AXILLARY ARTERY (3)
19 AXILLA
20 AXILLARY ARTERY
21 AXILLARY GROUP OF MUSCLES
22 AXILLARY NERVE
23 AXILLARY VEIN
66 DORSAL SURFACE OF AXILLARY BORDER OF SCAPULA
84 FIRST PART OF AXILLARY ARTERY (1)
228 QUADRANGULAR SPACE OF AXILLA
251 SECOND PART OF AXILLARY ARTERY (2)
294 THIRD PART OF AXILLARY ARTERY (3)
307 TRIANGULAR SPACE OF AXILLA

The thing to notice here is that the first three node descriptors contain the word THIRD; the next three include PART; and the remaining eleven all represent AXILLA.

GNODE next eliminates all the elements of the list that do not have a node profile corresponding to ARTERY. The reason for this is that the verb in this case carries an expected profile of ARTERY, and only arteries anastomose in the program's experience. Notice that in the following shortened list all nodes contain the word ARTERY, even though this is not included in the user's question.

294 THIRD PART OF AXILLARY ARTERY (3)
84 FIRST PART OF AXILLARY ARTERY (1)
251 SECOND PART OF AXILLARY ARTERY (2)
294 THIRD PART OF AXILLARY ARTERY (3)
20 AXILLARY ARTERY
84 FIRST PART OF AXILLARY ARTERY (1)
251 SECOND PART OF AXILLARY ARTERY (2)
294 THIRD PART OF AXILLARY ARTERY (3)

Also note the redundancy here. Node 294 occurs three times, 84 and 251 each appear twice, and 20 is present only once. This means that node 294 is the closest match and is selected as the output of GNODE.

The program now looks at the next phrase (SECOND). The right argument for GNODE contains eight node numbers. There is no redundancy because there is only one word in the phrase.

154 LUMBRICAL TO DIGIT 2 (FINGER SECOND)
246 SECOND CERVICAL NERVE (C2)
247 SECOND DORSAL INTEROSSEOUS (2 IO)
248 SECOND FINGER (2 INDEX DIGIT)
249 SECOND METACARPAL (2)
250 SECOND PALMAR INTEROSSEOUS (2 VENTRAL IO)
251 SECOND PART OF AXILLARY ARTERY (2)
350 VENTRAL SURFACE OF SECOND METACARPAL (2 VOLAR ANTERIOR PALMAR)

Despite the fact that it must determine that SECOND uniquely defines SECOND PART OF THE AXILLARY ARTERY in the context of the total question, the process is even simpler than before. The only node number in the list that happens to have a profile number corresponding to ARTERY is 251.

Role of complementary phrases. As you and I read the original question, i.e., "What anastomoses would develop following an occlusion between the second and third parts of the axillary?" we also realize that SECOND referred to SECOND PART OF THE AXILLARY ARTERY by an entirely different means. We probably assume that the two phrases were parallel and that the missing parts of the partial phrase are to be found in the following one. The program also uses this logic when necessary.

Consider the question WHAT IS THE INNERVATION OF THE SECOND AND THIRD LUMBRICALS? Processing begins with the second of the two noun phrases resulting from the parsing (THIRD LUMBRICALS). The vector of node numbers presented to GNODE is as follows, wherein node 155 appears twice, indicating that it contains both words.

153 LOWER 2/3 OF ANTERIOR SURFACE OF HUMERUS (TWO THIRDS)
155 LUMBRICAL TO DIGIT 3 (THIRD)
294 THIRD PART OF AXILLARY ARTERY (3)
154 LUMBRICAL TO DIGIT 2 (SECOND)
155 LUMBRICAL TO DIGIT 3 (THIRD)
156 LUMBRICAL TO DIGIT 4 (FOURTH)
157 LUMBRICAL TO DIGIT 5 (FIFTH)
158 LUMBRICALS

The program next looks at SECOND. The list of competing node numbers is identical to what we showed earlier for SECOND (PART OF THE AXILLARY ARTERY). Here, applying the expected profile eliminates only node 246, because arteries (251), bone (249, 350), and skin (248), as well as muscle (247, 250, 154) are all innervated by nerves. We are still left with seven nodes from which the program must select one. There is no redundancy, because it was a single-word phrase.

The program does exactly what we would do, that is, it borrows the node pointers for the word LUMBRICALS from the previously determined node descriptor and thus uniquely identifies node 154. This is the reason it processes the phrases in reverse order. If the question had been WHAT INNERVATES THE SECOND AND FOURTH DORSAL INTEROSSEI? the program would have borrowed the two words DORSAL INTEROSSEI to select node 247.

Adjectives restricting the context. Some verbals (relational words) are generic, in the sense that they each refer to several specific subtypes. Attachments include both origins and insertions. Actions refer to flexion, extension, abduction, adduction, pronation, supination, external and internal rotation, etc. Relations include all the logical opposites described previously (cf. Verbal Analysis). These include anterior-posterior, medial-lateral, etc.

If the question is "What are the relations of the wrist?" for example, the program lists the anterior relations, and then the posterior relations, because these are different verbs. However, if the question is "What are the anterior relations of the wrist?" the program lists only the anterior relations. Any logical variation of this produces the expected results. For example, "What are the anterior and posterior relations of the wrist?" produces both answers. Effectively, this restricts the meaning of the predicate nominative ("are the relations of") to whatever adjectives modify "relations." If there are no modifiers, all relations are listed.

Prepositional phrases restricting the context. The domain of such verbs may also be limited by prepositional phrases. Consider the following sequence.

What are the actions of the biceps?
IT FLEXES THE SHOULDER AND ELBOW JOINTS AND SUPINATES THE FOREARM.

What are its actions on the forearm?
IT SUPINATES THE FOREARM.

"What are the actions of the biceps?" produces three verb-object pairings (flex-shoulder, flex-elbow, and supinate-forearm). Here the prepositional phrase has eliminated those actions on objects not included in the prepositional phrase ("on the forearm").

The matrix of nodes and pointers represents both the system's knowledge and its logic.

Contextual interpretation of spelling errors. The phrase EXTENSOR ACRPI RADIALIS BREVIS contains a single typographical error, the reversal of CA in CARPI. The program never becomes aware of this mistake, because the three remaining words uniquely define the intended node. ACRPI is not in the vocabulary and, for that reason, cannot have any node pointers.

If two words were misspelled (EXTENSOR ACRPI RDIALIS BREVIS), the ambiguity in this case exists between two nodes:

69 EXTENSOR CAPRI RADIALIS BREVIS (HAND SHORT)
75 EXTENSOR POLLICIS BREVIS (SHORT THUMB)

However, because the context has narrowed the ambiguity to these two nodes, the misspelled words (those with no node pointers) have only to be compared with the words in these two node descriptors. This is done by means of the closest character match, which is essentially an APL primitive function.

Certainly the task is simplified by the limited subject matter domain. On the other hand, anyone who has seen a 1500-page textbook of gross anatomy could hardly call CATS a trivial application. The ability to identify nodes despite the ambiguities in all the examples just shown, and to do so in a fraction of a second, represents an impressive facility for language comprehension.

Implementation of logic in CATS

The matrix of nodes and pointers represents both the system's knowledge and its logic. The knowledge (nodes and pointers) may be retrieved directly. In CATS the logic of this semantic net may also be traversed to yield new insights.

The process of answering questions is similar for CATS and MEDCAT in that it involves (1) determining the nature of the question, (2) determining which nodes are being asked about (both of which have been previously discussed), (3) finding these nodes in rows 1 and 3 of the pointer matrix, (4) finding the associated pointer values in row 2, and (5) enclosing the result in appropriate narrative text. The main difference between the two is that MEDCAT has a diagnostic phase. Because this is executed before the questioning begins, explaining its reasoning consists essentially of describing the output of this subroutine. CATS, on the other hand, must perform a logical analysis each time a question is asked.

CATS also differs from MEDCAT in that question answering often involves recursive searching of the matrix until a complete answer can be obtained, because only the most direct relationships are represented in the pointer matrix.

Direct search. For the simplest class of questions (Q14), the answer may be found after one pass. The search consists of the following steps. (1) The program finds the positions of the value representing "innervates" or "innervated by" in row 2. This happens to be the number 4. (2) It finds the positions in row 3 corresponding to the node number for the structure being innervated (biceps). (3) The logical anding of the binary results of these two searches is used to compress row 1, which contains the structures (nerves) doing the innervating. Question 15 also involves direct searching.

Recursive searching. Question 16 illustrates a second level of complexity. The relationship of "is a branch of" or "branches into" is indicated in row 2 by the value 3. The structure doing the branching is in row 3. The branches, therefore, are found in row 1. Recursion is necessitated because each branch may itself have branches. Searching continues until no additional branches are found. The cumulative result forms the basis for the answer.

Sequential searching. Question 17 implicitly involves more than one type of relationship. When one asks what a nerve innervates (an artery supplies, and a vein or lymphatic drains), it is assumed that this involves not only the structure mentioned, but also all of its branches as well. For this reason, the program must first determine the branches of the nerve or vessel, and then search the matrix for what each of them supplies.

A program that can reason represents the distinction between learning by rote and using reasoning to decrease memorization.

Significance of algorithmic searching. The fact that "innervates" or "supplies" implies "by the structure and its branches" is an example of some of the rules embedded in the discipline of gross anatomy. The finite list of such rules is included within the code as a series of algorithms that determine the sequence of searching. For example, if asked to plot collateral arterial pathways to bypass an occlusion (Q18), the program must go through the following steps. (1) It finds the direct branches of the artery proximal to the occlusion. (2) It then traces through the matrix to find the branches of these branches until it can determine the muscles they supply, since collateral circulation develops only in muscle. (3) It then determines what other arteries supply the same muscle. (4) It traces these back up to the artery distal to the occlusion. (5) Lastly, it generates a clear narrative description. It does all this with a reasonable response time of about three seconds.

The decision to handle complex questions in this manner involved a choice. Clearly, all the muscles innervated by a nerve could have been directly represented in the matrix, as could all potential collateral arterial pathways. The decision to have the program use reasoning to determine the answer each time is based on the following two considerations.

To do otherwise would greatly increase the size of the matrix. For a specific region of the body, such as the upper limb, there are approximately 300 nodes (named structures) and approximately 1000 pointers (columns) in the pointer matrix. These can be displayed on about three pages. However, when asked to print out all it knows about the upper limb, the program must type over 300 pages of text. This is comparable to (actually exceeds) the amount of detail in the standard unabridged textbooks. When viewed from the perspective of the subject matter expert who must enter the data, this represents a tremendous saving. These 3000 integers can be entered very quickly, whereas explicitly spelling out all their implications would have taken a great deal of time, effort, and thought.

More importantly, this method represents the distinction between learning by rote and using reasoning to decrease the amount of memorization. No experienced anatomist memorizes pathways for collateral circulation; an experienced anatomist figures them out when the need arises. Getting students to use the same approach and to develop the same skills is a primary task of good instruction.

"Why-type" questions. Gross anatomy is a descriptive science. Unlike medical diagnosis, it does not normally deal with cause and effect. Nevertheless, we encourage students to ask "Why is ... true?" CATS is programmed to answer this type of question, as we do, by trying to point out general principles. The purpose is to further reduce the need for memorization.

To answer why the pronator teres flexes the elbow (Q19), the program focuses on the answer to the previous question. It first searches the matrix for the verb (row 2) and object (row 3) used. In this example, using "flex" and "elbow," the program comes up with all the muscles that flex the elbow. It then uses these node numbers (muscles) to search again, one relation (verb) at a time, to determine the greatest generalization it can make. When it compares the muscles that arise from the medial epicondyle with those that flex the elbow, it finds a 100 percent correspondence and applies the "all" to the answer. If it cannot make a meaningful generalization by focusing on the verb and object, it tries the subject and verb (Q20 and Q21). Here it searches for "deep volar branch of the ulnar nerve" (row 1) and "innervates" (row 2) to find all the muscles innervated by the nerve. If it cannot find a complete correlation, it looks for "all but" where the first assertion is more than 50 percent true, and the exceptions are three or fewer in number. To state the generalization more concisely, and to reduce the number of exceptions to meet the requirements of the algorithm, the program tries to place the exceptions into groups. In Q20, six muscles become three exceptions by being expressed as "the thenar group" (three muscles), "½ the lumbricals" (two muscles), and "the palmaris brevis."

The extent of the ability of CATS to discover general principles is illustrated by the following data. For the upper limb, the program generated 846 unique questions. When asked "Why?" for the answers to each of these questions, it can give explanations for 677 (80 percent). The total number of explanations given is 1668, or an average of 2.46 explanations per question. Of these 1668 explanations, 308 are unique. Such redundancy is what one should expect from the algorithm. An explanation with the general format ALL THE ... THAT ... THE ... ARE ... applies equally to every member of the set referenced by ALL.

The attribute that most clearly distinguishes effective teachers in a subject such as anatomy is the ability to reduce the mass of material to a more manageable number of general principles. Acquiring this skill takes years of experience. When we originally started to implement CATS we thought of it as a very efficient method of storing gross anatomical knowledge and producing drill and practice sessions without the laborious authoring requirements of conventional CAI programs. The ability of CATS to generate meaningful general principles to explain the answers to questions in an average response time of less than three seconds was a very real and pleasant surprise. It is a feature that greatly enhances the teaching ability of the program.

Anatomical variations. People are not identical; we come in different sizes and shapes. For the most part, gross anatomy ignores these differences and emphasizes the attributes we share in common. We describe the norm, and students learn to accept variations as they occur.

On the other hand, we must encourage these future physicians and scientists to be critical observers. Variations do occur in their laboratory dissection. The left hepatic artery, for example, is usually a branch of the hepatic artery (for about 88 percent of people). In about 10 percent of cases, it arises from the superior mesenteric artery. This degree of detail is not included in students' textbooks. A conflict arises if this variation happens to appear in a particular dissection. We want students to believe what they see and not to be limited by what they may read.

It is important to us that CATS be able to go beyond the textbook and show the same flexibility and understanding that an experienced teacher exhibits.

It is important to us that CATS be able to go beyond the textbook and show the same flexibility and understanding that an experienced teacher exhibits. If the student asks "What is the left hepatic artery a branch of?" CATS responds HEPATIC ARTERY. However, if the student asks "Does it come from the superior mesenteric artery?" or in student mode tells us that it comes from the superior mesenteric artery, the program responds YES, IN 10 PERCENT OF CASES. USUALLY IT IS A BRANCH OF THE HEPATIC ARTERY. In other words, the program is as specific as the situation demands. Under most circumstances the program describes the norm, and percentages (which are not retained anyway) are not mentioned. However, when the variation is part of that student's experience (as indicated by the nature of the question), the program puts that experience into perspective.

This flexibility is achieved on a coding level by signed decimal pointer values. Pointers representing the norm have positive signs; those for variations have a negative sign. The pointer connecting HEPATIC ARTERY and LEFT HEPATIC ARTERY is 3.88; that between SUPERIOR MESENTERIC ARTERY and LEFT HEPATIC ARTERY is −3.1. To this extent, our description of the pointer matrix as an integer array is inaccurate. However, although the code is present, the actual data (decimal pointers) have only been entered for arteries in the abdomen and pelvis. These are the regions where such variations are most common and where accurate data may be obtained readily.

Implementation of logic in MEDCAT

The actual process of answering questions is simpler in MEDCAT than it is in CATS. However, before the questioning begins, the logic contained in the semantic net must first be executed to activate certain of the nodes and pointers and to inactivate others. This function is mediated by the diagnostic subroutine. In order to adequately describe the concepts involved, we need to introduce some additional numeric variables and a schematic diagram.

The schematic representation of medical logic. Figure 1 shows 8 nodes (boxes and circles) and 11 pointers (arrows) interconnecting them. These pointers may be either positive (facilitory) or negative (inhibitory). Facilitory pointers are shown as solid arrows, and inhibitory pointers represented by broken arrows. Each pointer has an integer value associated with it to indicate its strength.

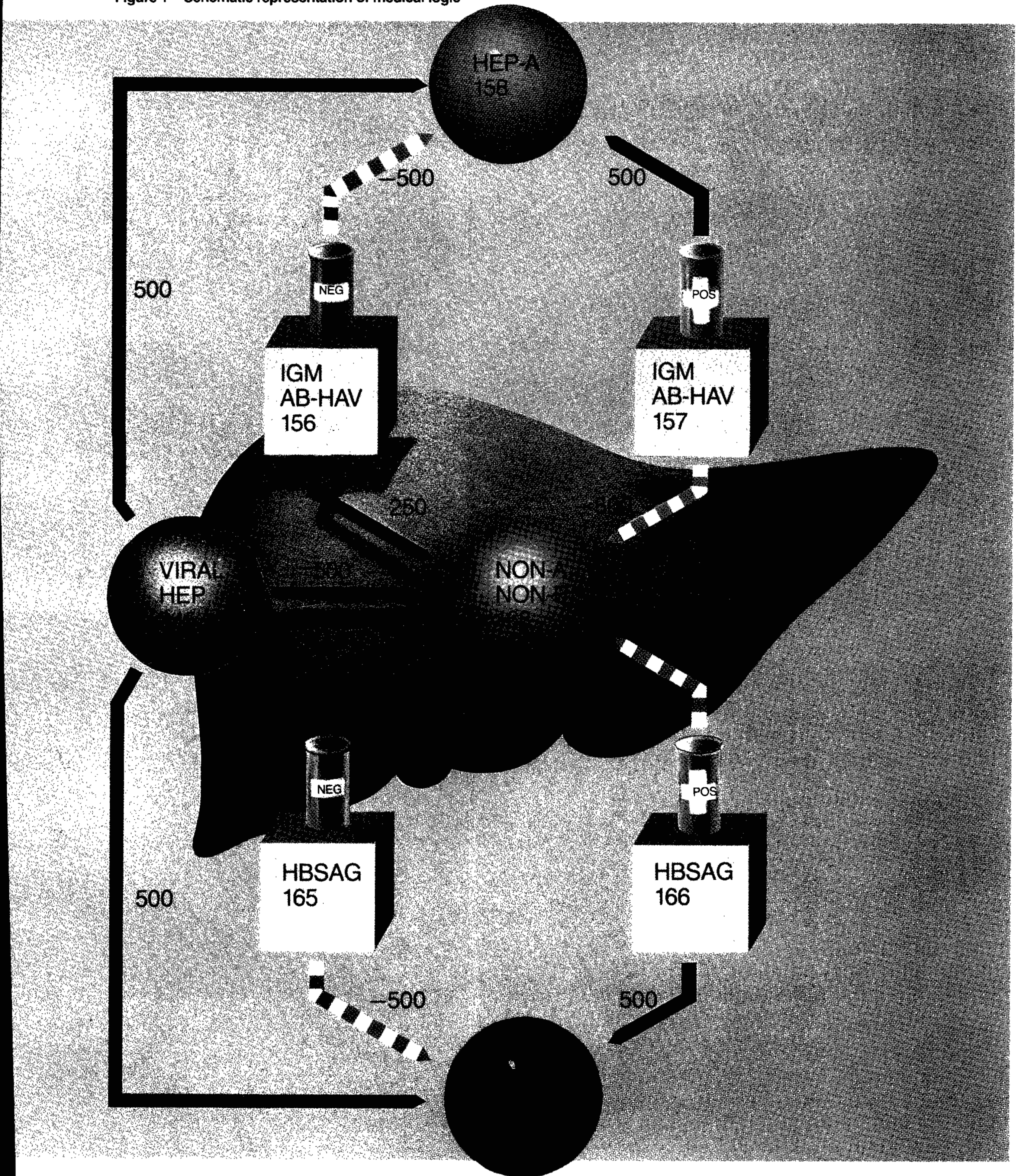
The complete form for each of these node descriptors used in Figure 1, together with its node number, is as follows:

156 IGM ANTIBODY TO HEPATITIS-A VIRUS (IGM AB-HAV)
157 IGM ANTIBODY TO HEPATITIS-A VIRUS (IGM AB-HAV)
158 HEPATITIS-A (ACUTE)
165 HEPATITIS-B SURFACE ANTIGEN (HBSAG)
166 HEPATITIS-B SURFACE ANTIGEN (HBSAG)
167 HEPATITIS-B (ACUTE)
168 HEPATITIS NON-A NON-B (ACUTE)
345 VIRAL HEPATITIS (ACUTE)

Numeric attributes. Five numeric attributes are shown for each of the nodes in Figure 1, together with their node numbers. Except for the profile, none of these have counterparts in CATS.

	156	157	158	165	166	167	168	345
PROFILE:	6	6	11	6	6	11	11	10
STIMULUS:	1	0	0	1	0	0	0	0
ADJECTIVE:	2	3	6	2	3	6	6	6
THRESHOLD:	100	100	500	100	100	500	500	500
PATIENT RECORD:	NEG	NEG		NEG	POS			

Figure 1 Schematic representation of medical logic



The *profile* of a node has already been described for its role in defining the context. More specifically, in MEDCAT, the first six types represent empiric data (1=chief complaint, 2=demographic data, 3=history, 4=symptoms, 5=signs, and 6=laboratory tests). Profile types 7–11 represent diagnoses with varying levels of specificity, 11 being the most specific. Nodes 156, 157, 165, and 166 are laboratory tests. The remaining are diagnostic nodes. The three with profiles of 11 are more specific than viral hepatitis (10); they are specific forms of viral hepatitis.

In Figure 1, four diagnostic nodes, together with their node numbers, are shown as circles, and four empiric nodes are shown as boxes. Empiric nodes always occur in sets of two or more, to reflect the range of possible values. We subdivide age, for example, into 11 ranges. The significance of a positive HBSAG node, and thus the group of pointers emanating from it, differs from that of a negative one. Only one node, positive or negative, in such a set may be on. If the test has not been performed, no nodes in that set are turned on. In the patient we will discuss, the IGM AB-HAV and the HBSAG are both negative. In contrast, all diagnostic nodes are unique; for instance, there is only one node with the descriptor HEPATITIS-B.

The *stimulus* is a binary vector that indicates which empiric nodes are on and which are off. If all the nodes in a set are off, the test was not done or the question was not asked of that patient. Diagnostic nodes all have stimulus values of 0. A physician postulates hypotheses and diagnoses from the empiric data available. Diagnostic nodes, therefore, must be turned on by execution of the program.

The *adjective* value of a node is an integer that is used to index a list of adjectives. This permits the program to generate responses containing phrases such as THE *POSITIVE* HEPATITIS-B SURFACE ANTIGEN, or THE *MARKEDLY ELEVATED* WHITE BLOOD CELL COUNT.

The *threshold* of a node is a positive integer. If the algebraic sum of all the active afferent (incoming) pointers to a diagnostic node equals or exceeds the threshold, the node is fired, and its efferent (outgoing) pointers become active.

The patient records are stored on disk in a manner similar to that of the automated medical record system developed for the New York Hospital.[3] The *patient record* for each patient contains the actual data for empiric nodes. The patient record data may be quantitative values for variables such as age, temperature, and total serum bilirubin, or they may be qualitative adjectives, such as those in the examples shown by nodes 156, 157, 165, and 166. The data are used to answer such questions as "What was the patient's temperature?" Diagnostic nodes have place holders but no values in the patient record.

It is very important for the program to examine the steps in the student's reasoning to make sure the appropriate questions were asked.

The threshold, stimulus, and patient record are specific to each patient. Thus their contents vary for each patient. The thresholds are calculated by the program.

The diagnostic process. Patient evaluation by a physician normally proceeds in a series of logical steps. The chief complaint, demographic data, current and past history, and symptoms are obtained by asking the patient questions. Signs are determined by physical examination. Some laboratory procedures are considered screening tests and are done more or less routinely; for example, there may be a chest film or urinalysis. Other tests are ordered only when the more preliminary examination suggests certain diagnostic hypotheses that require specific tests to confirm or exclude them.

It is also necessary to understand that there is a hierarchy of diagnoses with varying degrees of specificity. Three specific types of viral hepatitis are shown in Figure 1, and these have higher profile types (11) than viral hepatitis (10). Viral hepatitis is a subtype of the more general diagnosis of hepatitis (9). Hepatitis, in turn, is one of a number of manifestations of conjugated hyperbilirubinemia (8), but there are other types of hyperbilirubinemia that would point to entirely different parts of the diagnostic tree.

Except in those instances where some particular constellation of history, symptoms, and signs is so compelling as to strongly suggest a specific diagnosis, the physician's reasoning usually focuses first on the diagnostic levels with lesser specificity. If there is no evidence of increased bilirubin or other signs of liver disease, there is no need to pursue that branch of the tree. If such basic evidence is present, more specific questions will be asked and tests will be ordered to make the diagnosis more specific. Even when conditions do permit intuitive insight, prudence usually requires that one retrospectively check out the alternatives.

When MEDCAT is being used to teach medical students, it is very important for the program to examine the steps in the student's reasoning to make sure that the appropriate questions were asked, that there were valid reasons for ordering each test, and that the diagnosis was achieved for the right reasons rather than a lucky guess. MEDCAT has the ability to do this—to think like a physician—because the necessary logic is stored in the various numeric arrays already described.

On the other hand, when MEDCAT executes the diagnostic subroutine preliminary to discussing the patient with a physician or student, the process is quite different, and proceeds as follows.

1. All the empiric data are available to the program before it begins. The binary stimulus vector indicates all the findings that were present in that patient's medical record and which empiric nodes are active. Iteration within the program turns empiric nodes neither on nor off.
2. Pointers from these empiric nodes activate diagnostic nodes if their algebraic sum equals or exceeds the threshold of the node.
3. As may be seen from Figure 1, diagnostic nodes receive pointers from other diagnostic nodes as well as empiric nodes. Those diagnostic nodes that receive pointers solely from empiric nodes are turned on during the initial pass through the matrix, or they are not turned on at all. Diagnostic nodes that receive pointers from other diagnostic nodes are not activated until these precursor nodes are turned on. Such iterations continue until a steady state is achieved.

The basic iteration involves a single line of code:

```
L1:→L0×⍳∧/ST2=ST←STIM∨THRESH
 ≤ALGSUM←GALGSUM ◊ ST2←ST ◊ →L1
```

$ST2$ is initialized as 0. $STIM$ and $THRESH$ are the stimulus and threshold vectors and remain constant. $ALGSUM$ is a vector of the algebraic sums of the afferent pointers to each node and is calculated on each iteration by the subfunction $GALGSUM$. As may be seen, ST is a binary vector that results from the comparisons between $STIM$, $THRESH$, and $ALGSUM$. It represents the current state. If it is identical to the previous state ($ST2$), the iterative process is finished. If not identical, $ST2$ is respecified as ST and the calculation is repeated.

The result of this process is a list of all the active diagnostic nodes. The one (or ones) with the highest profile value are selected as the definitive diagnosis.

What was your diagnosis?
HEPATITIS NON-A NON-B.

Figure 1 shows the final phases of the diagnostic process in which the program has chosen HEPATITIS NON-A NON-B. Node 345 (VIRAL HEPATITIS) was activated during earlier iterations because the algebraic summation of its afferent pointers (not shown) exceeds its threshold. Nodes 156 and 165 were turned on by the stimulus. The algebraic sum of the active pointers from these three nodes to nodes 158 and 167 is 0. The summation at node 168 (HEPATITIS NON-A NON-B) is 1000, compared with its threshold of 500.

Why?
BECAUSE THE NEGATIVE IGM ANTIBODY TO HEPATITIS-A AND THE HEPATITIS-B SURFACE ANTIGEN AND THE DIAGNOSIS OF VIRAL HEPATITIS CONFIRMED HEPATITIS NON-A NON-B.

The "why" nature of the question causes it to look at the facilitory pointers to HEPATITIS NON-A NON-B and to enclose the node descriptors for the three nodes involved in appropriate narrative text. The word NEGATIVE is determined from the adjective vector. The verb CONFIRMED is dictated by the magnitude by which the algebraic summation exceeded the threshold.

Did you also think of hepatitis-A?
YES, HEPATITIS-A WAS CONSIDERED, BUT REJECTED.

What made you reject it?
THE NEGATIVE IGM ANTIBODY TO THE HEPATITIS-A VIRUS ARGUED STRONGLY AGAINST HEPATITIS-A.

The program says that HEPATITIS-A has been considered because there is at least one active facilitory pointer going to it. It says that HEPATITIS-A was rejected because the node was not turned on. The reason for rejection was obtained by looking at the active inhibitory pointer.

Why the IGM AB-HAV?
TO DETERMINE THE LIKELIHOOD OF HEPATITIS-A AND HEPATITIS NON-A NON-B.

When the "why" question refers to an empiric node, the program looks to the efferent pointers from the node to find out what decision (diagnostic node) these pointers would influence.

New knowledge and logic can be added to the program by free-format discussions.

Learning mode

New knowledge and logic can be added to the program by free-format discussions. This addresses the general problem of getting the most qualified experts to contribute to the logic and knowledge base. Experts usually are unaccustomed to analyzing the essential mechanisms that may underly their particular intellectual skills. They are experienced, however, in making decisions and justifying them once they are made. This is exactly what MEDCAT does—it engages experts on familiar intellectual turf. By asking questions such as "Why do you disagree?" or "Why do you think that?" or "Why did you order that test?" the program requires no more of physicians than what is expected of them on their teaching rounds.

In the following example the program has made an erroneous diagnosis and must correct itself and learn about a new disease and the basis for this new diagnosis. This further illustrates the utility of the method of data representation and the language capabilities of the system.

WHAT IS YOUR QUESTION?
What was your diagnosis?
HEPATITIS NON-A NON-B.
That's wrong.

WHY IS IT WRONG?
Because the positive beef red cell absorption test excludes it.

WHY DID YOU ORDER THE BEEF RED CELL ABSORPTION TEST?
Because the heterophil agglutinin test was positive.

WHY DID YOU ORDER THE HETEROPHIL AGGLUTININ TEST?
Because there were enlarged cervical lymph nodes and a flu-like syndrome.

The nodes referred to in this discussion already exist. The program is simply adding pointers to the pointer matrix. The physician's answer to why it was wrong caused a new (inhibitory) pointer to be added from an existing node (38) called BEEF RED CELL ABSORPTION to HEPATITIS NON-A NON-B (168). The program chose node 38 over the other node (37) in the same set, which has the same descriptor, because 38 had a stimulus value of 1. Verification was made by comparing its adjective value with the adjective used by the physician. The verb EXCLUDES caused the pointer to be sufficiently negative to outweigh all facilitory pointers. This was accomplished by adding a column (38 −1100 168) to the pointer matrix.

The reason given for ordering the BEEF ABSORPTION TEST added a pointer with a value of 1. Such weak pointers to empiric nodes affect only the conversational phases of the program and are discussed when we describe how the threshold is calculated.

Because the program now understands why its diagnosis was incorrect but does not know what the diagnosis should be, it asks.

WHAT IS THE DIAGNOSIS?
Infectious mononucleosis.

WHY DID YOU DIAGNOSE INFECTIOUS MONONUCLEOSIS?
Because the beef absorption and heterophil agglutinin tests were positive.

DO YOU MEAN BEEF RED CELL ABSORPTION TEST?
Yes.

WERE THERE ANY OTHER REASONS?

Enlarged cervical lymph nodes and flu-like syndrome.

ANY OTHERS?
No.

HOW IMPORTANT (ON A SCALE OF 1-5) IS THE BEEF RED CELL ABSORPTION TEST?
5.

INFECTIOUS MONONUCLEOSIS represents a new node and it is added to the list of node descriptors. It is a diagnostic node, and—because it is replacing HEPATITIS NON-A NON-B—it is assumed to have the same profile type. Because it is a diagnostic node, only place holders are required for its adjective type, stimulus value, and patient record. Its threshold, as with any node, is calculated prior to execution.

Four pointers must be added to INFECTIOUS MONONUCLEOSIS from the four nodes indicated. Because the physician did not use any quantitative verbs or adverb-verb combinations to indicate pointer values, these are explicitly requested. This process is repeated for the remaining pointers. If the verbs had been used, these values would have been inserted automatically (e.g., "excludes").

Note that the program recognized "beef absorption" as meaning BEEF RED CELL ABSORPTION. The reason it asked for confirmation is that here, in the learning mode, new nodes are being added. The program has no way of knowing that BEEF ABSORPTION is not a new node that differs from BEEF RED CELL ABSORPTION.

In order to test the results of this learning session, the diagnostic subroutine must be executed again.

WHAT IS YOUR QUESTION?
What was your diagnosis?
INFECTIOUS MONONUCLEOSIS.

Why?
BECAUSE THE ENLARGED CERVICAL LYMPH NODES AND FLU-LIKE SYNDROME SUGGEST, THE POSITIVE HETEROPHIL AGGLUTININ TEST STRONGLY SUGGESTS AND THE POSITIVE BEEF RED CELL ABSORPTION CONFIRMS, INFECTIOUS MONONUCLEOSIS.

Why did you order the beef absorption?
BECAUSE OF THE POSSIBILITY OF INFECTIOUS MONONUCLEOSIS, AND BECAUSE THE HETEROPHIL AGGLUTININ TEST WAS POSITIVE.

Why did you reject non-A non-B?
BECAUSE THE POSITIVE BEEF RED CELL ABSORPTION TEST EXCLUDES IT.

Evaluation of the logic. The program checks for obvious inconsistencies. It would reject BECAUSE THE HETEROPHIL TEST WAS NEGATIVE, because this disagrees with the patient record. Aside from conflicts with empiric data, however, the program accepts any line of reasoning, as long as it can be made consistent with what it already knows. The program stops asking questions and accepts the logic, when afferent pointers exist to each new diagnostic node.

Calculation of the threshold. The thresholds of new nodes do not have to be set by the physician. The threshold of all empiric nodes is set arbitrarily at 100. The values of afferent pointers to empiric nodes can be either -1 or $+1$. The algebraic sum of such pointers never exceeds the threshold of the empiric node and, therefore, cannot turn it on. Whether an empiric node is activated is determined solely by the patient data. This is an important theoretical point. Certainly one would not want the answer to a question or a test to be affirmative simply because the question was asked or the test was done.

These weak pointers are used when the program is generating answers to questions and when it is evaluating students' reasoning. One reason the BEEF RED CELL ABSORPTION test was ordered was because the HETEROPHIL AGGLUTININ test was positive, and this relationship was indicated by such a weak pointer. We are particularly concerned that students have adequate reasons for requesting laboratory tests. When requesting information from experts, the program asks for reasons for laboratory tests, although it does not ask for reasons for other types of empiric nodes.

The program calculates the thresholds for diagnostic nodes. The basic process consists of dividing the sum of positive afferent pointers to the node by two and using this or 500, whichever is greater, as the threshold. The sum is divided by two because we want to mimic varying degrees of certainty in a diagnosis. When the algebraic summation of afferent pointers to a node equals or barely exceeds the threshold, this is interpreted as a tentative diagnosis or working hypothesis. If it is more than twice the threshold, the diagnosis is confirmed.

From empiric nodes, only active pointers are used. This is necessary because each node in the set that represents a test has different pointer values. One would not want to sum these when only one of them can be relevant. It is also desirable because it permits the flexibility one sees in observing physicians at work. If a specific test or piece of information is not available, this usually should not rule out a diagnosis. A POSITIVE BEEF ABSORPTION TEST may confirm INFECTIOUS MONONUCLEOSIS, but the presumptive diagnosis can be made without it.

Afferent pointers from diagnostic nodes are counted, whether active or not, because there is no way of determining this before executing the diagnostic subroutine. Also, this is consistent with the role such intermediate diagnosis should play. If an intermediate diagnosis is missing, the more specific diagnosis should not be made.

Pointer precision. Our objective is to model medical diagnosis the way it is actually practiced. This largely rules out a statistical (Bayesian) approach. Physicians actually think, or at least explain their reasoning, in terms such as "is consistent with, suggests, supports, strongly suggests, confirms." On the other hand, the fact that the program is assigning quantitative pointer values based on these rough verbal approximations raises some important points.

Individual physicians use different terms to describe these relations. One person's "suggests" may be another's "supports." This goes beyond the role of synonyms. The same variation results when they are asked to rate the significance on a scale of 1–5. Reasonable consistency would be achieved if we went to three categories instead of five. We choose to use five mainly because it adds some richness to the conversational aspects of the program. In itself, this is not an important decision.

However, it is significant that this same group of experienced physicians will be almost unanimous in their final diagnostic evaluation, irrespective of the exact weighting of the individual components. The reasons for this depend on redundancy and a series of checks and balances; we have tried to build these into the program logic.

There is considerable convergence (redundancy) of afferent pointers at every decision node. Only half of these need be active to fire the node. If certain empiric data are not available, tentative intermediate diagnoses can still be made and these, in turn, trigger new hypotheses and indicate additional tests and questions.

If all the empiric data were permitted to converge (via pointers) on each definitive diagnostic node, not only would the semantic net become unwieldy and answers become too wordy (Q3), but also it would not work. The definition of intermediate nodes permits a logical anding and oring of the evidence. Considered in isolation, many signs and symptoms are so common and nonspecific as to have little diagnostic significance. However, certain combinations (anding) of these findings may have very specific implications. If one or more of the intermediate nodes, so defined, fail to be activated due to differences in weighting, the presence of a sufficient number (oring) of these still permits the reasoning to continue.

The program knows the purpose of the question and can take this into account when phrasing the response.

Generating the output

It has been said that the computer generation of text without the use of stored responses is more difficult than natural-language comprehension.[4] We have not experienced any such difficulty. There are several reasons for this.

The data structure, and therefore the answers, are neatly packaged into noun phrases and verb phrases. More specifically, the raw answers in all instances are columns of the pointer matrix. These columns, in turn, consist of nodes (rows 1 and 3) and pointers (row 2) that (integers) directly represent noun phrases and verb phrases.

The program knows the purpose of the question and can take this into account when phrasing the response. For this reason some answers are prefaced by BECAUSE . . . , IT IS . . . , OR NO, ...

With the simplest questions, formatting the answer may involve little more than retrieving the appropriate noun phrase or verb phrase. When multiple nodes and pointers are involved, proper sentence construction requires combining nodes into logical groups (phrases) according to their profile types and adjective values, and into clauses according to the associated verbs (pointer values).

Requests such as DESCRIBE THE COURSE OF . . . , DISCUSS THE RELATIONS OF . . . , or TELL ME ABOUT . . . elicit the most complex answers and necessitate the organization of sentences into paragraphs. The sequence of topics is largely determined by conventions specific to the subject matter disciplines (gross anatomy or internal medicine). The program recognizes these topics by the profile types of the nouns and a somewhat similar classification of verbs.

Given these components, it is relatively easy to program the output to mimic one's own linguistic style. The basic process in all these situations consists of taking these noun phrases, adjectives, and pointer values, and enclosing them in appropriate narrative prose. The exact form of the answer is determined by the nature of the question, pedagogical considerations, simple grammatical rules, and, most of all, a sense of how people involved in such discourses actually talk.

"What" questions. Questions that involve the retrieval of data from the patient record are answered as complete sentences (Q4, Q5, Q8). Many tests have a variety of names. SGPT, for example, may be referred to as "transaminase, alanine transferase, ALT, etc." Especially when dealing with students, including the name of the test in the answer avoids misunderstanding. It also aids in fostering a common (preferred) terminology.

"What was your diagnosis?" (Q1) or "What is the innervation of the biceps?" (Q14) are answered by simple noun phrases. There is no grammatical reason for this except that it is the way physicians speak, and to do otherwise would sound strange.

Certain generic types of questions in anatomy (actions, attachments, relations) require the appropriate verb in the answer (Q15). In all such instances the pronoun is substituted for the subject(s) when this does not lead to ambiguity. "What are the actions of the biceps and pronator teres?" would produce THEY FLEX THE ELBOW JOINT. THE BICEPS FLEXES THE SHOULDER AND SUPINATES THE FOREARM. THE PRONATOR TERES PRONATES THE FOREARM.

Whether a subject is plural or singular, for purposes of pronoun reference ("they" versus "it") and noun/verb agreement, is easily determined by the number of node numbers involved. We seldom need to contend with single nodes that are plural. In MEDCAT the rare exceptions are flagged as being plural, but even here this attribute is seldom called into play. The phrase "light-colored stools" is plural, for example. However, this usually is preceded by "the presence (or absence) of," which converts it to singular. "What do light-colored stools mean?" would elicit "They . . . ." In CATS the biceps muscle, for example, arises by two heads and at times may be referred to as though these were separate muscles (Q17). Conventional use, however, dictates "The biceps is . . . ."

The format of other answers is determined by their complexity and by the pedagogical points we want to make. There are many ways to ask about collateral circulation, for example. To eliminate any ambiguity between the program and the questioner, CATS paraphrases its interpretation of the question (Q18). Answers to this type of question are complex, involving lists of specific pathways. The format used is designed to make these relations as explicit and concise as possible. The name of the muscle where the anastomosis occurs is included, because this is the pedagogic point we are trying to make. The more detailed "branches of branches" type of explanation is not volunteered, because it would make the answers too wordy. When that type of answer is desired, we ask a different question.

The fact that the axillary artery happens to be subdivided into named parts triggers the type of answer shown (Q16). Suppose a student asks, "What are the branches of the brachial artery?" The program would answer THE BRACHIAL ARTERY BRANCHES INTO THE DEEP BRACHIAL, INFERIOR ULNAR COLLATERAL, RADIAL, SUPERIOR ULNAR COLLATERAL, AND ULNAR ARTERIES. The brachial does not have named parts. The branches are listed in the topographical (proximal to distal) sequence of their occurrence. Repetitions of the profile word ARTERY are compressed out of the object part of the answer and replaced by the cumulative plural form ARTERIES. This is simply a matter of preference, but mechanisms exist to implement almost any desired style.

"Why" questions. In MEDCAT, only when the question contains the word "why" do we preface the answer with BECAUSE (Q7). When alternative ways of asking the question are used (Q2, Q3, Q13), BECAUSE is not included. It is also omitted in CATS, where the answer represents a generalization, rather than cause and effect. Again, this is simply an attempt to mimic the way people actually speak.

The complex answers usually required here (Q2, Q3, Q7, Q13) are constructed of phrases, combinations of phrases, and independent clauses. In Q3, the three simple phrases ANOREXIA, NAUSEA, and VOMITING are taken directly from the node descriptor list. They are all symptoms (i.e., something the patient complains of); because the adjective descriptor for each symptom is empty, they are grouped together and prefaced by THE. PALMAR ERYTHEMA, GYNECOMASTIA, and ENLARGED LIVER are signs (i.e., something the physician observes) and have the same adjective value (PRESENCE OF). ENLARGED is part of the node descriptor and not in the list of adjective values. SERUM GLOBULIN, PROTHROMBIN TIME, SGPT, and SGOT are all laboratory findings. The first two and the last two are grouped together according to their adjectives. All of the nodes up to this point have the same range of pointer values (i.e., verb) and thus fit into the same clause.

Independent clauses are ordered according to the strength of the verb phrase (i.e., individual pointer values). The clause verb, e.g., IS CONSISTENT WITH, applies to each of the nodes in the clause, not to their combined weight. Similarly, the verb in the last clause, STRONGLY SUGGEST, refers only to the nodes in that clause and does not represent a summary statement. There are, of course, different degrees of certainty in diagnoses. This is determined by the degree by which the algebraic sum of all active pointers to the diagnostic node exceeds its threshold. This is expressed in answer to "What was your diagnosis?" by qualifiers such as THE *POSSIBILITY* OF CHOLECYSTITIS.

"True-false" questions. Decisions relating to true-false questions focus on when to respond with a simple "Yes" or "No" and when to provide additional information.

In CATS, students seldom ask true-false questions. When they do, it suggests either that they are lost or that they are checking on a previous answer that they may have misinterpreted. For this reason, the program tends to be as explicit as needed. (1) "Does the brachialis extend the elbow?"; NO, IT FLEXES THE ELBOW. This student is lost and is given the correct answer. (2) "Does the pronator teres pronate the forearm?"; YES, IT PRONATES THE FOREARM AND FLEXES THE ELBOW. He may have seen only one of these answers in a previous context and wants to check. (3) "Does the brachialis flex the elbow?"; YES. Here there is nothing on which to elaborate.

Questions about diagnostic nodes (Q6, Q9) are not prefaced by YES. If the answer were negative, it would be a simple NO. Questions about empiric nodes that have quantitative findings include this value in the answer. "Was the SGPT elevated?"; YES, THE SGPT WAS 85 U/L. As described earlier, each node has an associated element in the adjective vector. This is used when generating the output to have "elevated" or "slightly, moderately, or markedly elevated" most accurately reflects the actual data in that patient. However, when determining the validity of an assertion, as here, the program relies on the decimal portion of the word type for the adjective. All adjective phrases containing the word "elevated" have the same word type (7.5). "Had the patient recently eaten raw clams?"; THE HISTORY OF RECENT SHELLFISH INGESTION WAS NOT ASKED ABOUT. This illustrates the fact that such questions are not binary, but really admit of three classes of answers.

True-false assertions are often embedded, particularly in why-type questions. "Why does the brachialis extend the elbow?"; IT DOESN'T. IT FLEXES THE ELBOW. "Why did you order the total serum bilirubin?"; I DIDN'T. THE TOTAL SERUM BILIRUBIN WAS NOT MEASURED.

Silent nodes. In MEDCAT the significance of a particular combination of findings may be greater than the sum of the individual components (i.e., pointer values). This is the rationale behind the use of intermediate diagnoses. When considered individually, such symptoms as anorexia (i.e., loss of appetite), fatigue, headache, malaise, myalgias (i.e., muscle pains), and weakness are so common and nonspecific as to have little diagnostic significance. When enough of these are present in the same patient, however, they trigger the intermediate diagnosis of flu-like syndrome. Efferent pointers from this intermediate, named node suggest other more specific entities, including hepatitis. The values of these efferent pointers from such intermediate diagnostic nodes may be greater, and certainly more finely tuned, than if the intermediate node did not exist and the symptomatic pointers were simply summed. This strategy greatly reduces the size of the pointer matrix and makes the output that is generated more consistent with the way physicians actually think and speak.

Similar situations exist where there are no named intermediate entities. This is illustrated in Q3 and Q13. Dark urine may be a sign of jaundice. However, it can also be due to kidney disease, dehydration, or simply something you ate the previous night. Light-colored stools also are a common and, therefore, a nonspecific finding. However, the particular combination of the two would more strongly suggest the presence of jaundice. We call these *silent nodes*, because their node descriptors are represented by place holders that contain no text. On the input side, the program processes the individual component nodes and then checks to see whether they are part of any combinations and whether these combinations actually exist in the patient under discussion. On the output side, we add the phrase THE COMBINATION OF (Q3, Q13). The mechanism exists to handle any degree of pyramiding of such combinations, but more than a single level is seldom encountered.

Patriarch nodes. Nodes each of which represents a number of individual components are termed *patriarch nodes*. There are, for example, many types of pain, and each of these may have as many as 50 or more different attributes that determine the significance of the symptom (Q11 and Q12). The same is true of certain tests and procedures such as a chest film, sonogram, liver biopsy, and so forth. We refer to the top-level node as a *patriarch* and the attributes as *offspring nodes*. Each patriarch node is indicated by its node number in an additional integer vector that is equal in length to the total number of nodes. The offspring of this patriarch are flagged by the same absolute value but with a negative sign.

Unlike silent nodes, patriarch nodes do not conserve space in the pointer matrix. If a patriarch node sends a pointer to some other node, each of its offspring nodes may do so as well. The value of each pointer from a patriarch node is calculated as the sum of all the active offspring pointers to the same node.

The offspring pointers are not utilized in answering such questions as "Why did you consider . . . ?" nor are they utilized in the diagnostic subroutine, because they have negative node numbers. Questions like this are answered solely in terms of the patriarch nodes: BECAUSE OF THE RUQ PAIN, SONOGRAM AND LIVER BIOPSY. However, "Tell me more about the pain" (Q12) does elicit all the relevant components (offspring pointers) present in that patient.

The role of offspring nodes may be addressed directly, as in the question "What does the intensity suggest?" Here the fact that "intensity" refers to the RUQ pain is recognized by storing the number of the currently active patriarch node. This is similar to the method used for pronoun reference. The only advantage of the patriarch system is to make the conversational aspects of the program sound more natural. In our opinion, this fully justifies the additional code required.

Essay-type answers are elicited when the question contains no word types that define its purpose.

Essay-type answers. Essay-type answers are elicited when the question contains no word types that define its purpose, or when it contains certain higher-level word types that refer to more than one pointer type. Attachments, for example, include the origins and insertions of a muscle. Relations refer to those structures that are anterior, posterior, medial, lateral, inferior, superior, proximal, or distal to the structure being asked about. Actions or functions include such specific movements as flexion, extension, abduction, adduction, internal and external rotation, supination, pronation, etc.

Consider questions of the form "Describe the course of the ulnar nerve." Here "Describe" goes unmatched, but "course" triggers the activation of all pointers that indicate the beginning, end, relations, and branches of the ulnar nerve. The resulting sentences are arranged in a proximal-to-distal sequence. If the question were "Tell me about the biceps" or just "Biceps?" the lack of any purpose word type causes the program, either CATS or MEDCAT, to state in narrative form everything it knows about the subject. The answer would include the parts of the muscle as well as its attachments, actions, relations, and nerve and blood supply. These components would be arranged in the sequence that is used in conventional textbook descriptions.

The point to be made here is that even these complex answers are generated algorithmically. The sequence is specified as a hierarchy of pointer values. The components are obtained by traversing the pointer matrix. This ability to organize material into logical sequences is also used in the student mode when the program must choose a logical sequence of questions.

Student mode

Up to this point, we have described situations in which students ask the questions. This most clearly illustrates the relation between the data structure and the natural-language interface, unencumbered by the various strategies the program uses in deciding what to ask the student. We have tried to illustrate this with detailed examples. It is relatively easy to state the question, to show how these words are analyzed and mapped onto the abstract numeric model, how the problem is processed internally, and how the answer is translated back into narrative form. Both MEDCAT and CATS also have a student mode in which the program assumes the initiative. In a linguistic sense, analyzing a student's answer is much easier than understanding the question. The more interesting features of the student mode are related to such matters as how the program decides what to ask and when to ask it, how it evaluates the quality of the questions students ask of the patient or the validity of the reasons for ordering laboratory tests, and how it evaluates the student's total understanding of the patient and the disease independently of whether the diagnosis is correct.

The interesting features of the student mode are how the program decides what to ask and when to ask it.

It is not easy to illustrate these more global concepts with specific examples, because this requires long sequences of questions and answers. Instead, we shall describe the mechanisms in general terms and rely on the level of understanding of the system that the reader has acquired up to this point.

The student mode is very different from conventional computer-mediated tutorials.[5] There are no prestored questions or answers here. The program uses its own knowledge of the patient, its representation of medical logic, and the fact that it knows what information the student has obtained at any point to determine what questions to ask and when to ask them. Despite the highly structured manner in which the logic is represented within the program, it does not force the student to adhere rigidly to its own lines of reasoning. The program carries on the discussion in a flexible and individualized way, adapting itself to the student's knowledge of the patient and understanding of the problem. Regardless of the diagnostic approach taken by the student, the program is ultimately concerned with whether the student's inferences and conclusions drawn are sound.

MEDCAT. When MEDCAT is used to help medical students sharpen their diagnostic skills, the program plays dual roles: (1) It provides patient data as requested by the student; and (2) It functions as a tutor to evaluate and guide the student's reasoning.

The patient record is first read into memory and the diagnostic subroutine is executed, which takes about three seconds. This creates the vectors of active nodes and pointers specific to a patient that the program uses when discussing that patient. A brief narrative summary is then generated, including demographics and prominent presenting complaints. The student is then instructed to ask for any empiric patient data needed to make the diagnosis. This is an interactive process. The program may interrupt and ask the student questions under a variety of circumstances.

For example, when a student tries to obtain demographic, historic, or symptomatic information, he assumes that he is speaking directly to the patient. This permits the program to evaluate and comment on interviewing skills. A common error of beginning students is to ask questions of the following type that influence a patient's answer: "Did you have pain below the ribs on the right side?" At this point the program breaks and tells the student THAT IS A LEADING QUESTION. The program then requires the student to ask another question, which might take the following form.

Have you had any pain?
YES
Where was the pain?

Here the "patient" is permitted to answer. If the question the student asks the patient has a syntax similar to "Did you have" or "Was there," and if the question matches only one node in the list of node descriptors with profile types 2–4 (meaning it is patient-directed), it is treated as a leading question.

When conditions are such that an intermediate or definitive diagnosis can be made, the student is asked what that constellation of findings suggests.

If it matches on more than one, it is accepted. There is only one node descriptor that contains all the words "below pain ribs right side." There are many that contain only "pain." The patient gives an affirmative answer if it is true for any of the matches; otherwise the patient answers negatively. Questions with syntax such as "Where was the pain?" or "What was its quality?" are not considered leading questions, as in the patriarch nodes previously discussed.

Students are encouraged to ask such questions as "Tell me more about the pain." This is considered good interviewing technique, and the student gets a lot of information. The ability to answer such questions was one reason for developing the patriarch node concept.

When a student wants information about physical findings or laboratory data, he assumes he is speaking to the computer. As seen here, when the student requests a laboratory test that was not performed, the program asks why it was ordered.

What did the liver biopsy show?
LIVER BIOPSY WAS NOT DONE.
WHY DID YOU WANT THE LIVER BIOPSY?
Because I was considering malignancy.

These cases were carefully selected for teaching purposes. Therefore, if a test was not done, that fact indicates that the teaching physician believed that it was not warranted. This is important because many tests are expensive, and some tests involve risks to the patient. Students must be taught to justify their reasons for ordering such tests.

The program can evaluate the need for a test by looking at the status of the nodes and pointers involved. There are several diagnostic nodes that fall into the general category of malignancy involving the liver. If any of them have active afferent pointers, regardless of whether their algebraic sums exceed the threshold, this is an indication to order the test. Such is not the case in this patient. The program therefore follows this by asking what led the student to consider malignancy and continues until the misunderstanding has been resolved.

MEDCAT keeps a record of each question the student asks (i.e., empiric nodes asked about) and constantly compares the record with the information necessary to activate each diagnostic node. When conditions are such that an intermediate or definitive diagnosis can be made, the student is asked what that constellation of findings suggests or what competing hypotheses are ruled out by them.

At times, a student may not know what to ask next. He may ask for "help" by entering a null response. The program then looks at the questions the student has already asked and at the diagnostic implications (pointers) of each. The program selects the diagnostic node that has the most student questions directed toward it. The program then prompts the student to draw the correct inference by asking what additional information is needed to support that hypothesis or rule it out.

Sometimes a student asks questions in such a sequence that the diagnosis is arrived at very quickly. The program decides whether the diagnosis has been made too quickly by determining whether other competing lines of reasoning have been adequately ruled out. If these have not been ruled out, the program asks whether the student has considered these other possibilities and what evidence would be needed to confirm or exclude them. This is necessary, not only to ensure that the student reasons in a logical way, but also because a patient may have more than one diagnosis.

Other strategies may be employed that vary somewhat from these in detail, but they are all dependent on the facts that the program knows (1) which steps in the reasoning are essential to the diagnosis, (2)