Research Paper ■

Representing Clinical Guidelines in GLIF:

Individual and Collaborative Expertise

VIMLA L. PATEL, PhD, VANESSA G. ALLEN, BA, JOSÉ F. AROCHA, PhD,
EDWARD H. SHORTLIFFE, MD, PhD

A b s t r a c t   **Objective:** An evaluation of the cognitive processes used in the translation of a clinical guideline from text into an encoded form so that it can be shared among medical institutions.

Design: A comparative study at three sites regarding the generation of individual and collaborative representations of a guideline for the management of encephalopathy using the GuideLine Interchange Format (GLIF) developed by members of the InterMed Collaboratory.

Measurements: Using theories and methods of cognitive science, the study involves a detailed analysis of the cognitive processes used in generating representations in GLIF. The resulting process-outcome measures are used to compare subjects with various types of computer science or clinical expertise and from different institutions.

Results: Consistent with prior studies of text comprehension and expertise, the variability in strategies was found to be dependent on the degree of prior experience and knowledge of the domain. Differing both in content and structure, the representations developed by physicians were found to have additional information and organization not explicitly stated in the guidelines, reflecting the physicians' understanding of the underlying pathophysiology. The computer scientists developed more literal representations of the guideline; additions were mostly limited to specifications mandated by the logic of GLIF itself. Collaboration between physicians and computer scientists resulted in consistent representations that were more than the sum of the separate parts, in that both domain-specific knowledge of medicine and generic knowledge of guideline structure were seamlessly integrated.

Conclusion: Because of the variable construction of guideline representations, understanding the processes and limitations involved in their generation is important in developing strategies to construct shared representations that are both accurate and efficient. The encoded guidelines developed by teams that include both clinicians and experts in computer-based representations are preferable to those developed by individuals of either type working alone.

■ **JAMIA.** 1998;5:467–483.

In response to increasing economic pressures and a demand for a reduction in practice variation, there has been a growing emphasis on the production of computer-based clinical guidelines to support medical practice. The challenges associated with encoding clinical guidelines and sharing them among institutions with disparate computing environments have accordingly been among the focused research areas

for participants in the InterMed Collaboratory,[1,2] a collaborative partnership among investigators from Columbia, Harvard, and Stanford universities, with evaluation support by researchers from Cognitive Studies in Medicine, a part of the Centre for Medical Education at McGill University. Such collaboration among geographically distributed organizations with different goals and cultures provides significant challenges. The principal mandate for InterMed's participants has been to join in the development of shared infrastructural software, tools, and system components that will facilitate and support the development of diverse, institution-specific applications. The standardized implementation and sharing of clinical guidelines is one such effort, with broad potential applicability in local patient-care settings supported by the collaborating groups and, in time, for distribution to the general medical informatics community.

One experimental question, underlying all that InterMed has set out to achieve, is whether modern communication technologies can effectively bridge such cultural and geographic gaps, allowing the development of shared visions and cooperative activities so that the results are greater than any one group could have accomplished on its own. In another report we summarized the InterMed philosophy and mission, described our progress over three years of collaborative activities, and presented study results regarding the nature of the evolving collaborative processes.[2] As described in that paper, a major effort in the past two years of InterMed collaboration has been the development of a common language for representing clinical guidelines, the GuideLine Interchange Format (GLIF). A discussion of the syntax of GLIF, including a description of its use in encoding a variety of guidelines, is available elsewhere[3] and will not be included in this report. The advantages of using a common format such as GLIF include support for the collaborative development of guidelines, minimization of duplication by facilitating the sharing of guidelines among institutions, and enhanced mechanisms for updating the guidelines as medical advances occur.

The burgeoning interest in clinical guidelines has tended to focus on the creation of the guidelines through a professional consensus process, generally guided by relevant articles from the clinical literature. The successful dissemination of those guidelines, once created, has been limited, depending largely on information published in monographs or articles and the assumption that clinicians will read such information and incorporate it into their own practices. Because it is unrealistic to expect clinicians to read and utilize all the published guidelines that are available, several organizations and investigators have begun to look to automated methods of delivering guidelines to practitioners when the guideline is most relevant to the care of patients—i.e., while a patient is being seen by a clinician. Such approaches generally require the integration of a guideline's logic with a clinical information system for electronic medical record that can compare a given patient's situation with the logic of the guideline, checking to see if it should be triggered. The Arden Syntax[4] was developed in part to facilitate the sharing of decision logic among institutions, but Arden is not generally adequate for encoding complex temporal decision logic involving coordination among multiple medical logic modules (MLMs). While temporal logic is handled within MLMs, it is this coordination among MLMs that is needed to represent the sort of logic typical in complex clinical guidelines. Because there have been no robust standards for encoding complex guidelines or for defining underlying terminology, it has been difficult to share anything other than the text versions of guidelines among institutions, making it necessary for each organization to translate guidelines of interest into formats that are compatible with their local information systems and terminologies. It is these problems with the sharing of guidelines that the InterMed Collaboratory is attempting to address.

Within the collaboratory, the development of GLIF representations begins (Figure 1) with the authoring of the clinical guidelines by professional organizations such as the American College of Physicians, federal agencies such as the Agency for Health Care Policy and Research, or individual provider institutions such as Columbia Presbyterian Medical Center and Massachusetts General Hospital. InterMed's goal is to make available to health care organizations the GLIF representations of such authoritative clinical guidelines so that they may be locally adapted for use with local medical information systems and then presented to practitioners as they care for patients. However, as shown in Figure 1, there are many translation steps that occur before the GLIF-mediated guideline can be made available to a clinician. In broad terms, these steps can be categorized by the location at which such translations occur:

- *Generation of the paper guidelines at authoring institutions*. This step involves evidence-based consensus development, generally through examination of reports in the scientific literature.

- *Translation of the paper guidelines into GLIF representations at InterMed sites*. This can be done either

Figure 1 The process of shared guideline development as envisioned by the InterMed Collaboratory.
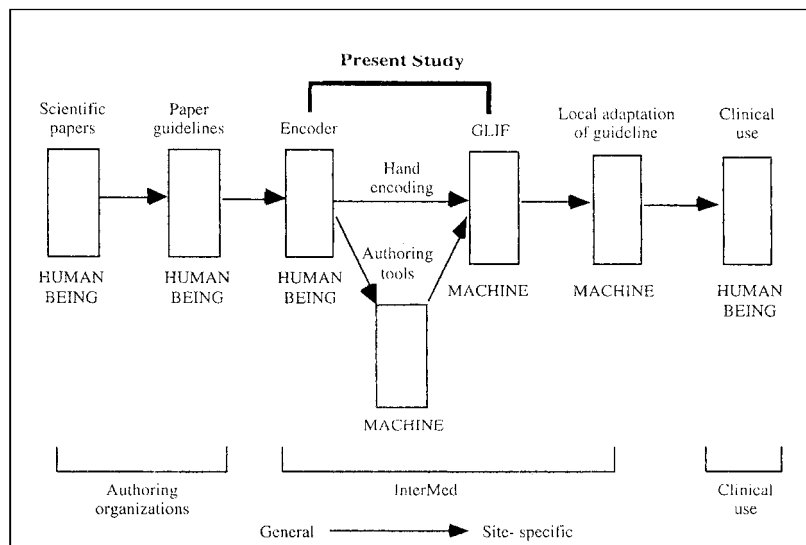
manually or by use of Stanford's Protégé systems[5] or Harvard's Geode system,[6] two separately developed diagrammatic authoring tools for generating GLIF encodings of clinical guidelines.

■ *Implementation of the GLIF representations within the clinical institution's application system.* Such adaptation may be largely automated, depending on the nature of the local system and possible institutional requirements for changes to the recommendations that a guideline may generate. It is unlikely that any organization will accept a generic guideline without at least some minor local modifications.

■ *A clinician's interpretation of the guideline as it is represented in the guideline applications,* which in the case of InterMed's ongoing efforts is intended to occur using the existing clinical information systems at Massachusetts General Hospital and Columbia Presbyterian Medical Center.

In this paper we examine in detail the second step in this process—the mechanisms and cognitive processes by which individuals analyze a text protocol document and convert it into an encoded representation of the guideline's logic. Our goal in such a study is to understand the kinds of expertise that are required by individuals who do such translations, to gain insight into the reproducibility of the process, and to develop criteria to assist in the design and refinement of authoring environments (such as the Protégé[5] and Geode[6] systems). We envision a day when computer-based authoring tools may be used not only to encode written guidelines but to create new guidelines, allowing domain experts to follow a rigorous

process of defining the events and decisions that must be considered for a given clinical situation. This would allow the merging of steps 1 and 2 in the four-step process outlined in Figure 1 and is one of the longer range goals of the InterMed work.

Theoretic Framework

The term "representation" is central to cognitive science.[7] Representations can be either internal (cognitive) or external (physical).[8,9] We use the term "external representation" to refer to the physical representation, such as a written clinical guideline or the printout of a computer program, and the term "internal representation" to refer to the mental "image" in a person's head. Thus, external representations are physical symbols and physical constraints of those symbols in the external world (e.g., spatial relations between symbols), whereas internal representations are knowledge stored in human working memory, and therefore are limited by the capacity of that memory. The development of an internal representation is limited by declarative knowledge—i.e., factual as opposed to procedural knowledge of a domain. A problem solver must develop a mental representation of the problem in working memory, which depicts how that person sees the problem—i.e., an individualized interpretation of the external information. Given the importance of external and internal representations for understanding the process of interpretation and translation of clinical guidelines into GLIF, we will discuss some concepts relevant to the generation of representations.

Understanding Guidelines Expressed as Text

In the present study, we have used cognitive theories to investigate the processes of representation and interpretation during the GLIF translation process (i.e., during the process of encoding a text guideline into GLIF, as depicted in Figure 1). The primary focus is the extent to which GLIF can be used for developing shared understanding of generic clinical guidelines and subsequently translated into site-specific guidelines that can be accurately and efficiently used by practitioners. Assuming that the criteria of accuracy and efficiency are satisfied by the original written guideline—from which we began our study—the adequacy of the GLIF representation depends on two main considerations that should be traded off: 1) the equivalence of the original guideline and the development of representations using GLIF, and 2) the flexibility of the GLIF representations for use in institutional settings with varying goals, priorities, cultures, and practical constraints.

The first consideration stems from the fact that standardized care can be compromised by differences in the way clinical guidelines are understood. Because of this, one of the central objectives of developing a shared computational language is the ability to construct representations that are equivalent to the original text guideline. There are two senses in which the original guideline and the GLIF-encoded representation can be said to be equivalent[9]: informationally and computationally. Informational equivalence refers to the idea that all the information that can be inferred from one guideline can also be inferred from the other. The original text guideline and its GLIF representation are thus informationally equivalent if they contain the same concepts and relations. On the other hand, two representations are computationally equivalent when an inference that can be drawn "quickly and easily" from one representation can also be drawn quickly and easily from the other. That inferences can be drawn "quickly and easily" means that the same cognitive operations (e.g., the same inferences) are performed when interpreting the two guidelines. If a relation is explicitly given in the original guideline and yet is implicit in the GLIF-encoded guideline, the two forms are not computationally equivalent; they require different cognitive steps and operations—mere retrieval or recall of information on the one hand, and a more complex reasoning on the other. As a result, it is possible that different inferences would be drawn when two representations are not computationally equivalent. Both informational and computational equivalence can be assessed by using methods of semantic analysis, as will be shown below.

A second consideration that underlies the development of GLIF-encoded representations is flexibility.[10] A flexible representation is generic enough to be applicable to different sites or institutions while allowing easy application to any specific site. This flexibility facilitates the generation of site-specific guidelines from generic, or site-independent, guidelines, which have become more important as researchers communicate and collaborate from various geographic locations. People across the world are now capable of accessing information, communicating, and engaging in collaborative projects through electronic media, such as e-mail and the World Wide Web. Although this opens the door for collaboration that was never possible before, there is, of course, no guarantee that these collaborative efforts will be successful. For communication, and therefore collaboration, to be successful, the information has to be interpreted in accordance with its intended meaning. However, information can be interpreted at different levels of abstraction, leading to different representations. There is also the danger that the shared information may be too general to be used in any particular context. The clinical guidelines being addressed by the InterMed collaborators are intended to be used across multiple institutions for a variety of purposes, including teaching, screening and disease prevention, treatment and management of patients, and generating reminders or alerts through event monitors that are triggered by data entered into a patient's record. A guideline that is developed for one of these purposes might not be easily adapted to fulfill a different purpose. Even if the general purpose of the guideline is the same, it must be possible to accommodate the particular priorities and institutional constraints at each site.

In response to this need for flexibility, GLIF was designed so that it would be implementation neutral.[3] That is, GLIF was designed so that it is sufficiently general to be used for a wide variety of purposes in a variety of settings. However, its flexibility must be balanced by the inclusion of details necessary for informational and computational equivalence. Too much information will limit the flexibility of the guideline for use across institutions, hindering the efficiency that such a project was designed to provide. In contrast, too little information will increase the possibility of alternative, possibly erroneous, or even dangerous interpretations. In fact, earlier evaluations of the GLIF-encoded representations of the flu vaccine and the breast-mass workup guidelines showed that different recommendations would be given on the basis of the same clinical case, depending on whose GLIF encoding of the guideline was followed.[3] Re-

flecting the attempt to reach the fine balance between equivalence and flexibility, such studies have led to extensions and clarifications of GLIF, which remains an evolving language, as more experience is gained with its use. The difference in recommendations reflects a lack of informational and computational equivalence, which can be investigated by focusing on the cognitive processes underlying the development of GLIF-encoded guidelines (examples of GLIF encoding can be found in Ohno-Machado et al.[3]). By investigating these processes, we have sought to describe some of the potential risks to the suitability and efficiency of this project and the processes through which some of these risks might be prevented. Before describing our study methodology and results, we review in the next section our motivating theoretic framework concerning the construction of internal representations in the process of interpreting written information, such as written guidelines.

Construction of Internal Representations

The process of interpretation involves the encoding of information present in external stimuli (e.g., a text) and the construction of a mental image of those stimuli. In this process, a distinction is made between what is written (the *text base*) and what we interpret (the *situation model*). What is written, the text base, can be viewed as the literal representation of the information present in the external stimuli, whereas what we interpret, the situation model, consists of the inferences we make from the text[11] and is dependent on our prior knowledge and experience in the text's domain. When interpreting a text, a person processes what he or she reads in working memory (i.e., memories that are currently active) while retrieving from long-term memory prior knowledge that gives meaning to what is read. The internal representation that is constructed results from the interaction between what is "out there" in the world (e.g., a clinical guideline) and what we think "in here" in our long-term memory (i.e., as affected by our prior domain knowledge). As a physician is presented with a clinical guideline, his or her internal representation of the procedure to be followed is based on domain knowledge and clinical experience. In other words, we use what we know to interpret what is before us.

Differences in the construction of the situation models are supported by research in medical expertise.[12] Many studies of medical expertise, which focus on differences in subject-matter knowledge in medical tasks, have shown that people differ in the nature and form of the mental representations they build from clinical information.[12,13] Experts represent medical in-

formation at a high level of abstraction, whereas novices represent it at a lower level.[13–17] For instance, a novice may recall literal information or make inferences that consist of simple operations, such as interpreting "temperature of 41°C" as "fever." In contrast, an expert can interpret clinical information involving longer chains of inferences and reasoning, for instance, encoding "BP 200/120 on left arm and 110/70 on right arm" as "aortic dissecting aneurysm."[12]

These research results suggest that, depending on their level of experience (e.g., years of practice and familiarity with particular diseases) and prior knowledge, different physicians are likely to interpret the same clinical guideline in different ways. If two external guideline representations are the same (i.e., they are informationally and computationally equivalent), their text bases should be equivalent, given that they reflect the literal meaning of the guideline. However, the preservation of the text base alone is not sufficient to maintain this equivalence in the construction of the GLIF-encoded representation. The problem is that the situation model provides the context for interpreting the guideline in a clinically meaningful way. Having the appropriate situation model makes it possible to overcome errors, fill gaps, "disambiguate" procedures or temporal sequences, and reorganize the information, both in constructing external GLIF-encoded representations of the original text and when using the guideline in clinical practice. Thus, because their situation models are different, expert physicians and less expert physicians will provide different GLIF representations on the basis of the same information. A possible solution would be to develop guidelines where participants with different backgrounds and expertise collaborate in the guideline development and translation process, sharing and clarifying concepts and issues, on the basis of their knowledge of and experience in their domains of expertise. Not only would expertise be shared in collaborative development, but as experts communicate their knowledge to those with less medical expertise, it may lead them to make their implicit knowledge more explicit. Collaborating to develop a shared understanding, however, involves some additional aspects that are not apparent in the individual interpretation process; one important aspect, the negotiation of "shared" situation models, is introduced below.

The Development of Shared Representations: Collaboration and Cooperation

Collaboration in scientific and technologic fields has engendered recent interest in many different disciplines, including education,[18,19] information science,[20]

social psychology,[21] sociology,[22] management,[23] and artificial intelligence.[24] This growing interest has been motivated particularly by the increasing need to share resources among different research laboratories[25] and by the possibility of collaboration among distant people, facilitated by advances in computer-based communications and networking.[26]

Collaboration refers to a process whereby different people have a commitment to pursuing a common goal[24,27] and where different "parties who see different aspects of a problem can constructively explore their differences and search for solutions that go beyond their own limited vision of what is possible.[28] For collaboration to be successful, different people should have a "shared" situation model of what the collaboration is about. This includes an overall idea of the task to accomplish and how it should be carried out. However, conflicts may arise in building the shared model as individuals bring different knowledge and experience to the task. Although it may be assumed that conflict interferes with the building of a shared model, the negotiation of conflict may actually be beneficial as experts contribute the relevant knowledge as they explain and justify their individual interpretations.[29–31] Through this process, ambiguities can be revealed and questioned by those with less knowledge and experience of the domain, and yet explicated by experts, who might otherwise access this knowledge implicitly. Furthermore, with the collaboration of experts in different domains of expertise—in this case physicians and computer scientists—this dialectic clarifies the knowledge of both the medical concepts contained in the clinical guidelines and the nature of the computer language used to represent such guidelines. The shared situation model that results includes explicit knowledge from the various domains of expertise.

The theoretic framework outlined above suggests some possible hypotheses about the likely outcome of the study. According to comprehension theory, it is expected that different situation models will be constructed, especially among individuals with different domains of expertise. Given their knowledge of the domain, it is likely that physicians will add implicit procedures, concepts, and relations from the biomedical and clinical domains as they construct GLIF representations. In contrast, computer scientists are likely to limit themselves to the medical concepts that are given by the original guideline. Not only the types of knowledge used (e.g., explicit, implicit; biomedical, clinical) are likely to differ, but also their organization. Previous studies in medical expertise have shown that experts organize their representations in a top-down

fashion, paralleling the hierarchic ordering of medical knowledge.[32] With prior knowledge and experience of medicine, experts are able to chunk clinical data into higher-level concepts, which subsume lower-level concepts. In contrast, without much knowledge of the underlying relations between the concepts, nonexperts must rely not only on the surface information in the text but also on the organization that is imposed by the text. The hierarchic organization of their knowledge allows experts to make rapid decisions without having to consider many alternatives.[33] Therefore, when interpreting a guideline, experts are more likely to include only the necessary steps, without having to consider all alternatives.

Methodology

The present study was designed by InterMed's evaluation team from McGill University to examine a subsection of InterMed's guideline translation process, namely, the development of GLIF representations of an existing text version of guidelines by subjects at Columbia, Harvard, and Stanford Universities. This examination requires a detailed characterization of the knowledge structures and the reasoning processes involved in translating the text guideline into GLIF. Our characterization is founded on a cognitive-science approach to the study of human thinking and behavior. Research in the cognitive and social sciences most often utilizes theoretic frameworks and models that do not provide precise predictions. This research capitalizes on the detailed investigation of processes looking for general trends in the existence of certain psychological phenomena (such as directionality of reasoning). Generalizability here is the second stage of the study, where these phenomena are tested under various conditions of applicability (e.g., directionality of reasoning under emergency and primary care conditions). This is unlike epidemiologic research, in which generalizability is a primary concern and the phenomenon under investigation may not be true. These two methods can be viewed as complementary means of investigation. Our research utilizes a strategy[34] characterized by the detailed analysis of a few subjects and the mapping of the perceptual and cognitive steps underlying observable behavior, focusing on individuals' specific organization of knowledge structures and sequences of reasoning steps. Our main source of data consists of verbal protocols.[35]

The approach developed in this paper reaches these goals through the following methodologic steps. First, a cognitive task analysis is carried out by specifying the knowledge and information processing capabilities humans have and how these are used in a specific

task, with particular focus on the goals, the underlying assumptions, the domain knowledge, and the reasoning steps necessary to perform the task. A task analysis is a job description but in terms of psychological demands that the job makes on the person. Second, after this task analysis has been completed and used for the development of a detailed, predictive model, an empirical investigation of the subjects is carried out. Third, the behavior of the subjects is analyzed and comparted with the results from the task analysis, so that inferences can be made relating the observed behavior to the theoretic model. The extent to which the observed data fit the theoretic model is taken as evidence of the psychological validity of the model. We have explored this approach in the study of medical problem solving, comprehension, and reasoning in laboratory-based psychological investigations of clinical performance.[13] Subsequently, we have engaged in field research to test, refine, and apply our theories and methods to real-world contexts (e.g., medical and surgical intensive care units) and problems.[36] The theory has informed practice, and in turn we have gained substantial insight from practice to inform and extend our cognitive theories.[37]

Applying this cognitive science approach to the investigation of the translation process, we will focus on the development of internal representations intermediate to the original paper-based guidelines and GLIF-encoded representations. Internal representations are predicted to be the main source of variation as individuals contextualize the given information at this point of the translation process. More specifically, it is expected that internal representations will diverge from the original external representation (paper-based guideline) with the adding of details, the reorganization of information, and the resolution of ambiguities such as temporal relations.[38] Because of differences in expertise among individuals, we also predict that internal representations will vary among individuals; the nature of the details added, the way the original representation is reorganized, and the ambiguities that are resolved are expected to be a function of the information stored in long-term memory (which is built up from prior knowledge and experiences). Our aim is to study these internal representations developed in working memory and the relationship they have to expertise. Coinciding with this, we hope to facilitate the standardization of the GLIF encoding process and suggest ways to increase the accuracy of the GLIF representations for clinical use.

As internal representations are not directly accessible for analysis, we propose an alternative way of investigating the differences that emerge in the translation

of paper-based guidelines to GLIF-encoded representations. We infer the content and structure of the internal representations on the basis of 1) participants' verbalized interpretation of the original external representation (the paper form of the guideline) to the final form of external representation (the GLIF-encoded representation), and (2) the final GLIF representation itself. Furthermore, we compare these interpretation processes and final products among individuals of different domains of expertise. Through analysis of interpretation processes and the GLIF-encoded representations, differences that emerge that are analogous to the predicted differences in the construction of internal representations may be attributed to the internal processes themselves.

Our inquiry into the interpretation of paper-based guidelines and translation into GLIF representations involved participants at three InterMed sites. More specifically, our study focused on the developers and implementers at Columbia, Harvard, and Stanford universities, who were asked to generate GLIF representations in their natural, everyday work environment. We asked these computer scientists, physicians, and medical informaticians to translate a paper guideline (text and flowchart) into GLIF format. In addition, to examine one possible solution to guideline variability, some participants were asked to develop GLIF representations in collaboration with others. A clinical guideline for the management of encephalopathy[39] was selected for two reasons: it was easy to access, and it was not familiar to any of the participants, who had worked on other previous InterMed guideline translations.

The subjects who were observed during the construction process included: 1) at Columbia, a physician/medical informatician (denoted CPMI, for Columbia physician and medical informatician); 2) at Harvard, a team interaction among members of the Brigham and Women's Hospital, including both computer scientists and a physician (denoted HIN, for Harvard interaction), and 3) at Stanford, independent encodings by two computer scientists (denoted SCS1 and SCS2, for Stanford computer scientist 1 and 2), by a physician (SP, for Stanford physician), and then by the physician interacting with one of the computer scientists (SIN, or Stanford interaction). All subjects were members of the InterMed Collaboratory, except for the Stanford physician who agreed to participate without any prior knowledge of GLIF or the details of the InterMed project. We consider this "opportunistic" research where, in the midst of observing a subject encode a guideline, we realized that some medically related questions needed to be clarified. We took this
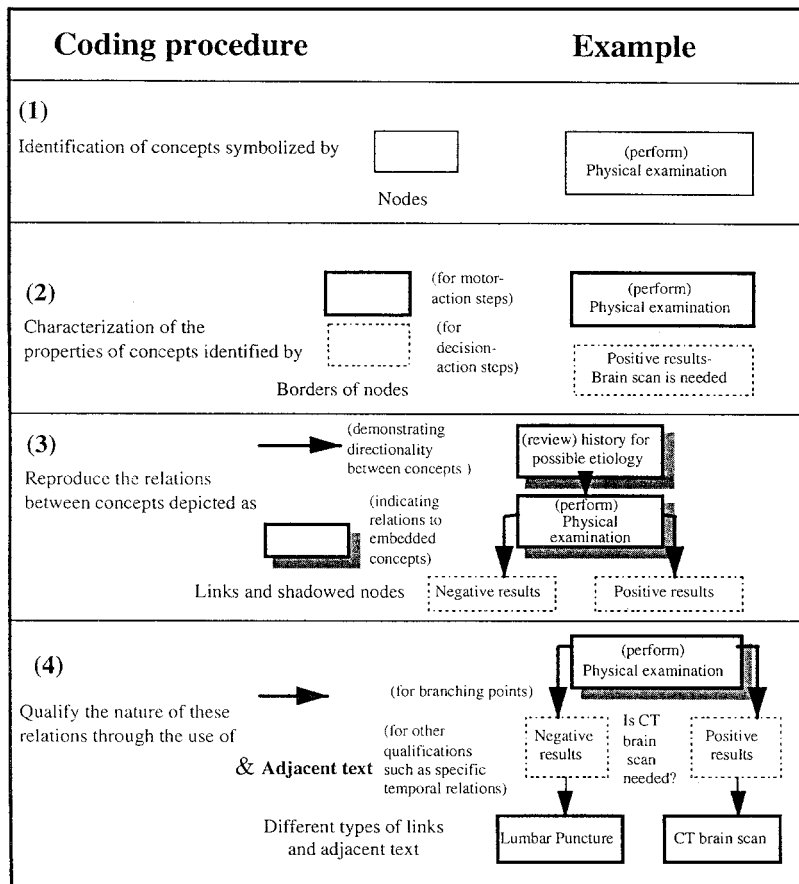
Figure 2 Outline of coding procedure of individual verbal protocols.

opportunity to approach a physician to work in collaboration with the subject (computer scientists) to complete the encoding.

The data were collected over a period of two weeks. At each of the three sites, each subject was asked to read the original guideline and think aloud as he or she encoded the natural language into another language using GLIF, making notes if he or she wished. It is assumed that these data reflect information that individuals have in working memory during problem solving.[35] At the sites where two or more people were involved in the construction of the GLIF representation, the interactive dialogue of the guideline translation process was also recorded. Written notes, diagrams, and computer screen snapshots that the participants constructed were also collected for analysis. The participants were asked not to discuss any part of this project until all the data were collected. Compliance with this request was verified (since we had access to most InterMed communications, including e-mails, conference calls, and telephone communication).

The data derived from individual participants' pro-

tocols were transcribed and analyzed using methods built of propositional and semantic representations.[12] Based on these methods, coding of the protocols involved an analysis of the following four elements (presented schematically in Figure 2 and illustrated by the analysis of one individual's protocol in Figure 3):

- The concepts described in the protocols (e.g., encephalopathy and physical examination),

- The properties of these concepts, such as whether a step is a motor-action or a decision-action step (e.g., the performance of a physical examination and the consideration of computed tomographic brain scan),

- The relations among concepts (e.g., that the administration of drugs is related to both the history review for possible etiology and the laboratory findings for blood samples), and

- The nature of these relations, including whether concepts are to follow one another in time or whether they are subcomponents of other concepts (e.g., the administration of drugs without delay, or

different physical tests organized as subcomponents of physical examination).

These concepts and relations were presented as a semantic network. Within these networks, the concepts are presented as nodes and relations between concepts are depicted either as links or as embedded structures (for hierarchically organized concepts). Declarative information that could not be captured simply as nodes or links, such as questions of ambiguity, temporal relations, and organizational themes, is added as adjacent notes.

Differences in the inclusion of details, organization of concepts, and treatment of temporal relations were the primary focus of our comparison. However, in order to do justice to the complex processes involved in the translation process, we supplemented our analysis of declarative and procedural information with a detailed description of how each GLIF representation was developed over time. This descriptive account was used as further verification of the role of internal representations and long-term memory; through investigating the manner in which the task was approached and completed, describing the degree of reliance on the original paper-based guideline, and discussing possible ambiguities (whether or not they were resolved), the analysis captured details that also support the role of expertise in the construction of internal representations. For example, although an individual may be unaware of how to represent a certain relation in GLIF, the awareness of a potential problem may be a sign of increased expertise.

In the analysis of the collaborative construction of GLIF representations, we required additional methods for examining the individual construction process. To examine the flow and management of information, the addition of knowledge, and the coordination of the participants within the dialogue, a descriptive coding scheme was developed that involved parsing the dialogue into goal-based episodes, in which a goal is defined as a specific *point* (change in topic) that is introduced. Each occurrence of a *change* in goal was coded as a new episode. Each episode was coded for the following:

- Number of conversational turns (defined as uninterrupted input in the dialogue)

- Number of participants

- Type of goal (management, purpose, representation, summary, and definition)

- Type of exchange ("question and answer," offering of alternative, clarification, summary, differing op-
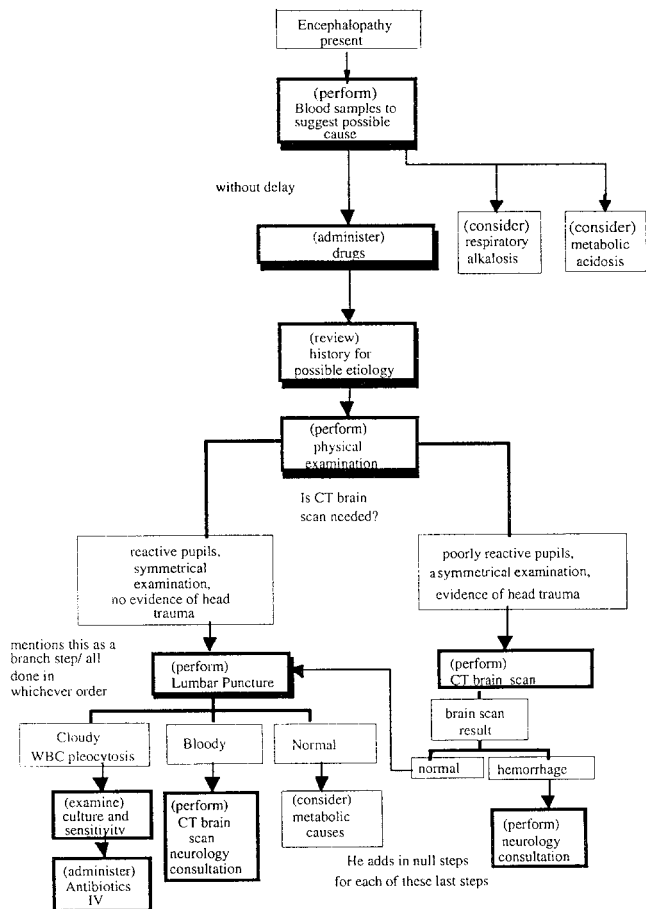

Figure 3 A semantic network of one participant's protocol while he developed a GLIF representation of the encephalopathy guideline.

tions, and addition of information)

- Type of knowledge added

- Evidence in support of a position

- Conclusion of the episode

Illustrating this coding scheme, an excerpt from the coding of the dialogue during the collaborative translation process at Harvard is given in the appendix. It presents the analysis of each episode in terms of the goal, the number of conversational turns, the participants, and the type of exchange. We investigated the processes of interaction and the collaborative development of GLIF representations, including the management and elicitation of information through strategies such as "question and answer," and the specific similarities and differences between individual and group representations.
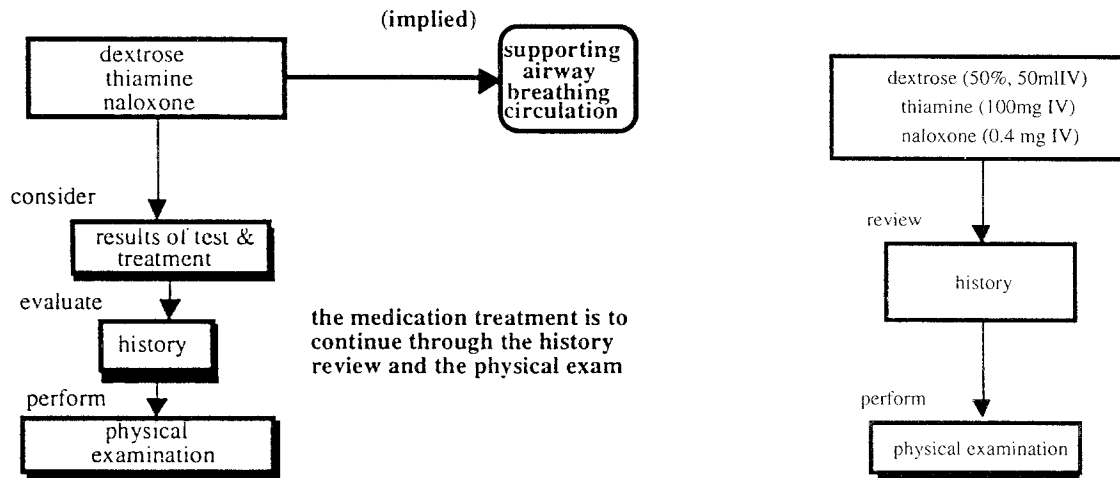
Figure 4 Semantic network of the protocols generated by a physician (*left*) and a computer scientist (*right*) at Stanford while they constructed GLIF guideline representations showing different levels of detail.

Results

Analysis of the protocols generated during the development of GLIF representations of the encephalopathy guideline shows differences in structure and content as a function of domain expertise. Individual representations were constrained by implicit and explicit knowledge of the domain resulting in differences in the level of detail provided (text and situation knowledge); the organization of knowledge, such that different reasoning patterns were generated; and the knowledge of the procedural nature of the domain of practice to represent temporal sequencing of events. We also found that differences emerged as a result of GLIF development (technical environment), demonstrating external as well as internal constraints on the translation process. Examination of the processes of generating collaborative representations demonstrated some of the strategies used to communicate and negotiate expert knowledge during the translation process, where domain knowledge was made explicit through dialogue with others less experienced in the medical and computer science domains. In reporting the results, a summary statement is provided at the beginning of each subsection, followed by illustrative examples to convey how these conclusions were drawn. This particular strategy—unlike the traditional approach, in which methods, results, and discussion sections are almost always presented sequentially—is designed to help the reader by providing an organization for walking through the complex and detailed analysis. In this approach, the summary statements are used as "advance organizers."

Level of Detail

The present analysis of the protocols generated during the development of GLIF representations revealed differences in the representation of detail, especially between the physician and the computer scientists. Illustrating these differences, Figure 4 gives fragments of the semantic networks of the protocols generated by the physician and one of the computer scientists. The Stanford physician (SP) supplements the guideline by adding the procedures of checking for supporting airway, breathing, and circulation, and by clarifying the temporal relations between steps (Figure 4, *left*). This added detail is not included in the GLIF representation developed by the computer scientists. For example, SCS1 keeps the procedures simple, adding little knowledge to the text base (Fig 4, *right*). Without the same degree of experience in the medical domain, the computer scientists were found to rely more heavily on concepts that are explicitly represented in the text. However, some of the computer scientists added details relating to their area of expertise such as the location of branching and null steps. These findings are consistent with our predictions based on the theories of text comprehension and expertise, in that the subjects constructed situation models that match their own experience in their domains. A consequence of the differences in added detail is that the physician's GLIF representation (Figure 4, *left*) gives a deeper explanation of the process (explanatory model), which in turn leads to a clearer outline of the task (performance model). With limited contextualization, the explanatory model developed
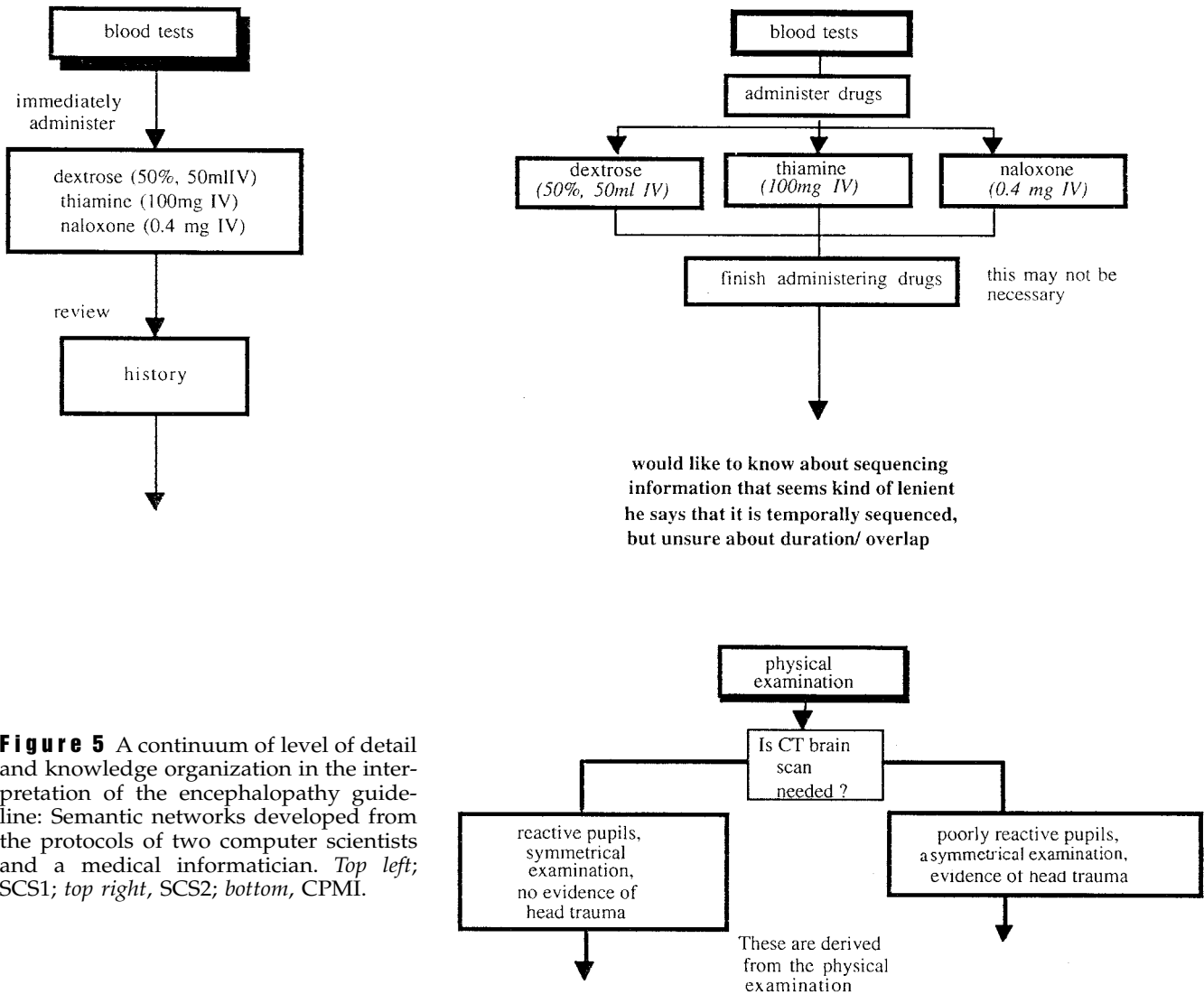
Figure 5 A continuum of level of detail and knowledge organization in the interpretation of the encephalopathy guideline: Semantic networks developed from the protocols of two computer scientists and a medical informatician. *Top left*; SCS1; *top right*, SCS2; *bottom*, CPMI.

by the computer scientists (Figure 4, *right*) is incomplete, leading to possible implications for the use of these external representations in a clinical setting. The pattern of results reported in the illustrative example of the computer scientist was similar to the pattern of the two other computer scientists working individually.

Knowledge Organization

The GLIF-generated representations developed by the physician and the computer scientists also show differences in the organization of medical concepts (Figure 5). Illustrating these differences, the SP's representation shows increased organization of the concepts with higher-level, compact macrostructures, coherent relations among concepts, and a high degree of differentiation between levels of granularity (e.g., pro-

cedure types and symptoms). In contrast, the representations developed by the computer scientists are bottom-up, with lower-level concepts dominating, and not hierarchically organized. For example, SCS1 represents symptoms such as "reactive pupils," "asymmetrical examination," and "no evidence of head trauma" as one step, and "review of history" as another. These organizational differences are supported by the fact that medical knowledge is organized in a hierarchy, with clinical information at the top level followed by pathophysiologic and basic sciences at lower levels.

Organizational differences were also found in the degree of branching. Analysis of the protocols revealed that computer scientists developed a highly branched representation enabling parallel actions and the inclusion of Boolean conditions, whereas the physician en-
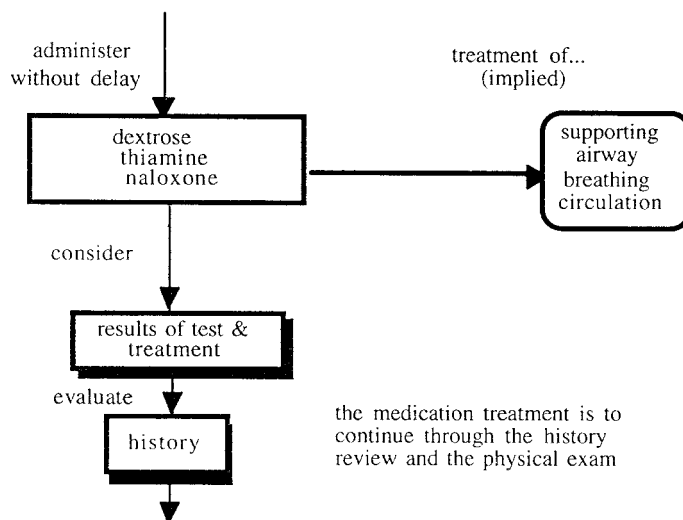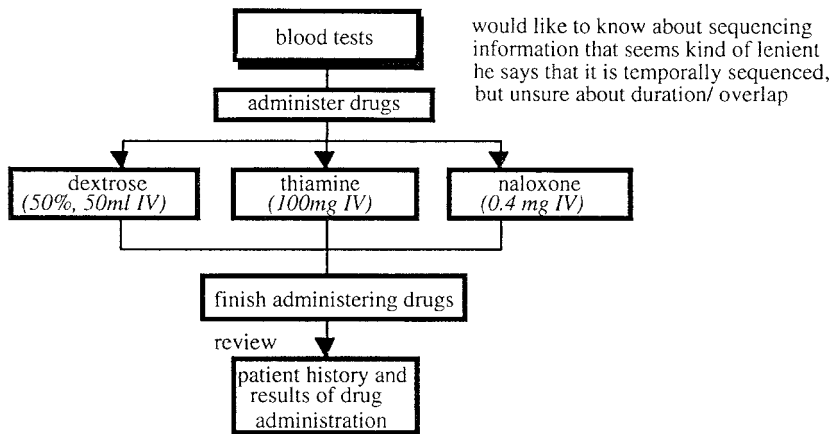
Figure 6 The differences in the interpretation of temporal relations between steps as illustrated by the GLIF representations developed by a computer scientist and a physician at Stanford. *Top*, SCS2; *bottom*, SP.

coded the guideline as a linear sequence of actions. Rather than considering several options in parallel, the physician recognized a pattern leading to linear decision making that is appropriate for action in urgent situations. In contrast, as the goal of computer scientists is to include the various options of the clinical procedure as accurately as possible, the use of branches is an appropriate method. A similar pattern to the organization and branching demonstrated by the illustrative example was found in the analysis of the protocols of the two other computer scientists working alone.

Consistent with the theoretic rationale provided earlier, differences in GLIF representations emerged among the computer scientists as a function of their knowledge and experience in the medical domain. Illustrating one side of this continuum, with little experience in the medical domain, SCS1 focused on the procedural component of the guideline and did not adequately capture declarative knowledge, such as the components of the history review and the time frame of the various steps; or reorganize the concepts as presented in the original flowchart (Figure 5, *top left*). At the other extreme, CPMI (a medical doctor and a computer scientist) added declarative information (Figure 5, *bottom*) but did not reorganize the information presented in the original guideline. Intermediate between these two extremes, SCS2 noted the need to consult a physician for this information (Figure 5, *top right*). His intermediate knowledge of medical issues was reflected by his sensitivity to the ambiguities of the original guideline, identifying the kinds of details that would be required for a more complete GLIF representation.

Temporal Relations

Differences were found in the encoding of temporal information. Illustrating this, SCS1 was unable to include accurate temporal relations between steps. More experienced in medicine, SCS2 did not rely on the guideline to determine how the steps are to be followed and instead requested consultation with a physician. (Figure 6, *top*). At the other extreme, the physician added specific temporal relations (Figure 6, *bottom*). This variation is supported by the theory of the processing of temporal information. Without knowledge of a domain, temporal relations can be inferred through the order of presentation and spatial relations between steps in a text or a flowchart. However, because of the implicit nature of temporal information, if these inferences are made without prior experience in the domain, they may be misleading.[38] Since many relations are difficult to include in text form without the addition of extensive detail, as when describing the degree of overlap between procedural steps, variation would be predicted as a function of knowledge of the practical domain. Like the other computer scientists, the medical information from Columbia (CPMI) also did not explicitly clarify the temporal relations in the GLIF encoding.

Constraints of the Computer Environment

Although not part of the original hypotheses of our study, differences were also found to result from the computer environment that was used for developing GLIF representations; the two computer environments used for encoding the encephalopathy guideline were GLIF itself and Protégé—an intermediate encoding tool. GLIF is a declarative format, in which the representation is given as propositions. Illustrating the impact of developing representations directly in GLIF, CPMI considers what aspects should be represented as conditional steps, Boolean conditions, and the temporal relations within branch steps such as the order in which the results of the lumbar puncture should be considered (Figure 7, *top*). Also, as it is not highlighted through the use of this declarative format, he does not consider the temporal relation between the administration of drugs and the steps that follow the administration step.

Illustrating the differences in computerized guideline representations developed with the use of Protégé, a diagramatic tool used to facilitate the encoding process that is to be subsequently translated to GLIF, SCS2's Protégé representation failed to capture one of the critical links leading from the results of the physical exam to the lumbar puncture (Figure 7, *bottom*). Although the graphic Protégé representation can be supplemented with embedded declarative informa-

tion, in this example the diagrammatic nature of Protégé's flowcharting scheme appeared to make it more difficult to capture multiple links.

Development of Cooperative and Collaborative Representation

Differences were found between the GLIF-generated representations developed by individuals and those developed through interaction with others. Not only did the collective representations lead to the addition of information from each participant's domain of expertise, interacting with those who were less familiar with the concepts and relations led the experts to articulate implicit information. Thus, many details that were made explicit in the collaborative representations remained implicit in the representations that the experts developed on their own.

Accordingly, in a collective GLIF representation developed for the present analysis, one computer scientist at Stanford supplemented his representation through questions and answers with a physician (SIN). This cooperative effort resulted in the clarification of ambiguities, such as the temporal relation between the drug administration step and the steps that follow it; the addition of further information, such as signs of systemic infection; and the clarification of some implicit strategies, such as the consideration of differential diagnoses during the history-taking part of the guideline. However, the general organization of the GLIF representation that emerged from this dialogue did not change much from the representation developed by a computer scientist on his own; the computer scientist clarified points at which he was unsure what to do, but knowledge was otherwise not shared much between him and the clinician.

The dialogue between two computer scientists and the physician at Brigham and Women's Hospital at Harvard was a collaborative process that resulted in a GLIF representation that reflected the expertise of each participant (HIN). In this interaction, each of the three participants added to the overall representation by answering questions, offering clarifications, providing new information, and contributing alternative interpretations. For example, the physician added information, such as which blood gas tests would have to be performed, and offered suggestions for cases that were not included in the guideline, such as whether a lumbar puncture would be performed given signs of laryngitis. The computer scientist with more knowledge about GLIF discussed the temporal relations that could be captured in GLIF. The second computer scientist, who has more experience in the
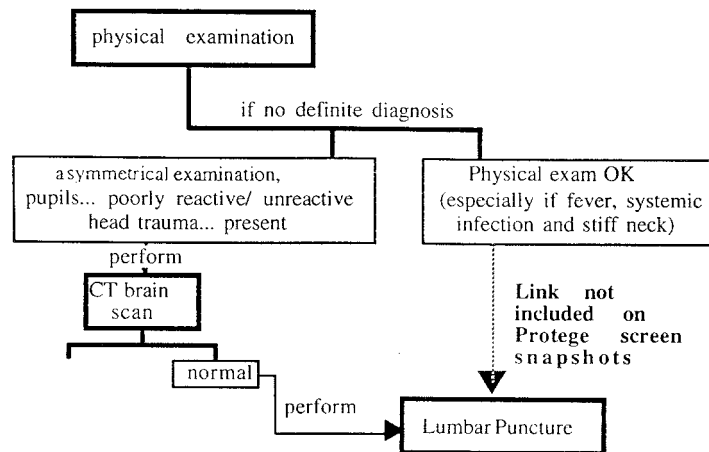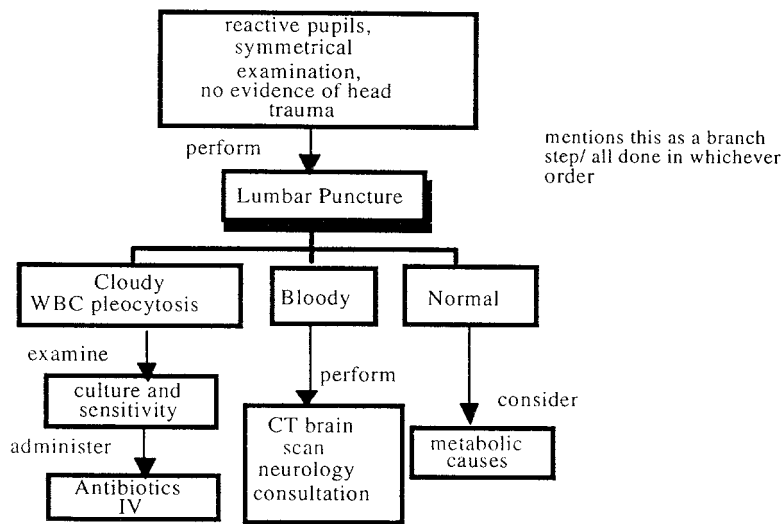
Figure 7 Selected parts of semantic networks developed from protocols while participants constructed external representations in different computer environments. *Top*, In GLIF (CPMI); *bottom*, in Protégé (SCS2).

medical domain, clarified the timing of the steps and added parameters for determining whether a patient has metabolic acidosis. This interaction led to the development of a shared collaborative GLIF-generated representation, in which each participant contributed information that the other participants either lacked or about which they did not have enough expertise.

A more detailed analysis of HIN demonstrates some of the strategies that were used in the construction of this collaborative GLIF representation. Although "question and answer" was the main mechanism for adding information to the representation, the discussion of differing or conflicting opinions about the representation also played an important role (Table 1). On average, the discussion of conflicting views in the interpretation process involved more conversational

turns than the episodes that centered on questions and answers. Yet through elaborating the reasons why the conflicting individuals should interpret the guideline in a particular way, the participants discussed the underlying knowledge that would otherwise have remained implicit.

As the interaction involved coordinated effort among the participants, much of the discussion was devoted to the management of the collaborators in the developmental process, such as the coordination of participants with respect to the part of the guideline that was being translated at any particular time, rather than focusing solely on the representational task. Accordingly, in addition to requiring the development of shared mental models of the goal of the collaborative enterprise, some of the exchange also focused on the

Table 1 ■

Characteristics of Exchanges in Dialogue

	Question and Answer	Alternative	Clarification	Summary	Differing Opinions	Addition
Total number of episodes	32.00	8.00	10.00	1.00	11.00	2.00
Mean number of conversation turns	9.16	13.50	9.40	3.00	14.10	20.50

management of the dialogue, the purpose of the task at hand, and the limitations of developing the GLIF representation.

Some differences emerged from a comparison of the GLIF-generated representation developed by the collaborative effort (HIN) and the one developed by an individual physician (SP). Although the physician included many details, some remaining ambiguities were identified by individuals with less experience in the medical domain who noted that key concepts were not included in the physician's representation. Thus, through the collaborative effort, those aspects that might be improperly interpreted by those with less experience in the domain can be identified and addressed in the translation process. The GLIF representations should capture only information that is necessary to ensure safety of the guideline representation in clinical use, and not so much information as would hinder the flexibility of the GLIF representation for efficient use across institutional settings. Examination of the collaborative representation of the encephalopathy guideline in GLIF suggests that the process of question and answer and the negotiation of differing opinions help provide just that information in the final text base of the GLIF representation. This occurs even though it involves the added task of managing and coordinating the mental models of individuals with different knowledge and experience. Negotiation itself leads to a representation that can be safely and efficiently used by individuals with different levels and domains of expertise, priorities, goals, and institutional constraints.

There is a tradeoff between including generic information and encoding the underlying situation model. The first task must be context free, whereas the second must be specific to the domain. The inclusion of both generic and specific information as multiple layers appears to be important in the translation process. The physician provides context to the GLIF representations, and those with less expertise narrow down the contextual information in the final representation. The result is to include only what is necessary for the accuracy of the final GLIF representation in its clinical

use, helping to ensure reusability with minimal local adaptation.

Discussion

By investigating the cognitive processes that underlie the development of individual internal representations, human–machine interaction, and collaboration, we have sought to describe some of the potential risks to the accuracy and efficiency of guideline encoding and identify the processes through which some of these risks might be prevented. The problems arise because of a number of factors. The first is that experts interpret information differently from nonexperts. We found many differences in the way that the physician encoded the encephalopathy information compared with computer scientists. When solving familiar real-life problems, experts often skip steps during the inferential process,[12,40] whereas trainees (e.g., residents) may depend on details that appear unnecessary to experts. Furthermore, there is assumed knowledge and understanding that do not have to be made explicit when one physician talks with another physician. The author of the original encephalopathy guideline seemingly wrote it to communicate with other physicians who had similar knowledge and accordingly left out a number of "obvious" steps leading to therapeutic and management decisions.

The second problem arises when guideline information is being translated into a structured computer language. Here, every step has to be clearly stated in the procedure if the rules are to be executed properly. Thus, the missing information may be inserted by a computer scientist working alone, which is how inaccurate inferences can be generated. However, when computer scientists work with physicians, the ability of each to detect errors is enhanced by their differences in perspective, through a process of negotiation. If a relationship is explicitly represented in the original guideline but is implicit in the GLIF representation, the two forms are not informationally and computationally equivalent. Without computational equivalence, inferences from the GLIF representation

will not be made as quickly and as easily as from the original paper guideline. Without informational equivalence, it is also unlikely that the same inferences would be derived from both the GLIF representation and the paper guideline.

Given the pattern of results from the detailed analysis of the protocols of the computer scientists and the physician, it appears that the computer scientists focus on surface information when given a text description of the clinical problem. The physician, familiar with this clinical problem, includes the underlying clinical model in his representation, which goes beyond the surface features of the task. These findings are consistent with the well-known theoretic perspective on text comprehension discussed earlier, whereby persons unfamiliar with a domain focus on the surface cues whereas those familiar with the domain are likely to make inferences based on a deeper understanding or underlying model.[11] Investigation of the process by which different people negotiate a shared understanding of a common GLIF-encoded representation points to a solution to the problem of multiple potentially conflicting models. One advantage of GLIF is that it seeks to satisfy the criterion of generality; that is, the guidelines represented in GLIF are intended to be used across different contexts, settings, and times. This flexibility requires a relatively abstract level of detail within the guidelines. If, on the other hand, temporal relations were given explicitly, the number of detailed specifications required at each step would increase and significantly constrain the general applicability of the GLIF-encoded guideline representations. Similarly, domain experts will constrain the generality by focusing on specific details, as shown by the physician in this study who constrains the generality by making the GLIF-encoded representations situation-specific.

The cooperative and collaborative efforts we studied have revealed some of the processes involved in developing a shared guideline representation and suggest ways to develop more complete and accurate GLIF representations for future use. The results suggest that the developmental process for a GLIF-encoded representation must involve active cooperation between participants with different perspectives at each stage in the guideline's translation. In this particular study, the different perspectives are provided by the physician (clinical domain knowledge) and computer scientists (knowledge of GLIF encoding). Although our studies identify robust phenomena, our conclusions have limited generalizability. Further studies to include different subjects and various conditions of applications would be a natural extension

of the reported investigation. An additional issue is whether some of the benefits of collaborative encoding of guidelines can be achieved through the enhancement of authoring tools which, as they act as partners in the encoding process, will probably be used for our future GLIF encoding work. Given the collaboratory's efforts to meet the challenges of resolving guideline ambiguity, GLIF and its use offer intriguing new tools for clinical practice and research.

References ∎

1. Shortliffe EH, Barnett GO, Cimino JJ, Greenes RA, Huff SH, Patel VL. Collaborative medical informatics research using the Internet and the World Wide Web. Proc AMIA Annu Fall Symp. 1996;125–9.
2. Shortliffe EH, Patel VL, Cimino JJ, Barnett GO, Greenes RA. A study of collaboration among medical informatics laboratories. Artif Intell Med. 1998;12(2):97–123.
3. Ohno-Machado L, Gennari JH, Murphy S, et al. The GuideLine Interchange Format: a model for representing guidelines. J Am Med Inform Assoc. 1998;5:357–72.
4. Hripcsak G, Ludemann P, Pryor TA, Wigertz OB, Clayton PD. Rationale for the Arden Syntax. Comput Biomed Res. 1994;27:291–324.
5. Musen MA, Gennari JH, Eriksson H, Tu SW, Puerta AR. PROTEGE II: computer support for development of intelligent systems from libraries of components. Medinfo. 1995:766–70.
6. Stoufflet P, Deibel RA, Lee D, Traum JH, Greenes RA. Integrating guidelines and information support into clinical workflow in a distributed environment: data structures and knowledge representation for GEODE-CM. Program and abstracts of the AMIA Spring Congress; June 5–8, 1996; Kansas City, Missouri; p. 109.
7. Palmer SE. Fundamental aspects of cognitive representation. In: Rosch E, Lloyd BB (eds): Cognition and Categorization. Hillsdale, New Jersey: Lawrence Erlbaum, 1978: 250–303.
8. Zhang J. The nature of external representations in problem solving. Cognitive Sci. 1997;21:179–217.
9. Larkin JH, Simon HA. Why a diagram is (sometimes) worth ten thousand words. Cognitive Sci. 1987;11(1):65–99.
10. Fridsma DB, Gennari JH, Musen MA. Making generic guidelines site-specific. Proc AMIA Annu Fall Symp. 1996: 597–601.
11. van Dijk TA, Kintsch W. Strategies of discourse comprehension. New York: Academic Press, 1983.
12. Patel VL, Groen GJ, Knowledge-based solution strategies in medical reasoning. Cognitive Sci. 1986;10:91–116.
13. Patel VL, Arocha JF, Kaufman DR. Diagnostic reasoning and expertise. The Psychology of Learning and Motivation: Advances in Research and Theory. 1994;31:137–252.
14. Elstein AS, Shulman L, Sprafka SA. Medical Problem Solving: An Analysis of Clinical Reasoning. Cambridge, Mass.: Harvard University Press, 1978.

15. Feltovich PJ, Johnson PE, Moller JH, Swanson DB. LCS: The role and development of medical knowledge in diagnostic expertise. In: Clancey WJ, Shortliffe EH (eds). Readings in Medical Artificial Intelligence: The First Decade. Reading, Mass.: Addison-Wesley, 1984:275–319.

16. Patel VL, Groen GJ, Arocha JF. Medical expertise as a function of task difficulty. Memory and Cognition. 1990;18(4): 394–406.

17. Patel VL, Groen GJ. The general and specific nature of medical expertise: a critical look. In: Ericsson A, Smith J (eds). Toward a General Theory of Expertise: Prospects and Limits. New York: Cambridge University Press, 1991:93–125.

18. Crook C. Computers and the Collaborative Experience of Learning. New York: Routledge, 1996.

19. Edelson DC, Pea RD, Gomez LM. The Collaboratory Notebook. Communications of the ACM. 1996;39(4):32–3.

20. Ras ZW. Collaboration control in distributed knowledge-based systems. Info Sci. 1997;96(3–4):193–205.

21. Murphy SE. Collaboration and human performance. Performance and Instruction. 1995;34(7):20–2.

22. McClelland K. The collective control of perceptions: toward a person-centered sociology. Paper presented at the Annual Meeting of the Control Systems Group; July 1996; Flagstaff, Arizona.

23. Inkpen AC. Creating knowledge through collaboration. California Management Review. 1996;39(1):123.

24. Grosz BJ. Collaborative systems. Artif Intell Magazine. 1996; 17(2):67–85.

25. Kouzes RT, Myers JD, Wulf WA. Collaboratories: doing science on the Internet. IEEE Computer. 1996;229:40–6.

26. Patel VL. Individual to collaborative cognition: a paradigm shift? Artif Intell Med. 1998;12(2):93–6.

27. Smith JB. Collective Intelligence in Computer-Based Collaboration. Hillsdale, N.J.: Lawrence Erlbaum, 1994.

28. Gray B. Collaborating: Finding Common Ground for Multiparty Problems. San Francisco, Calif. Jossey-Bass, 1989.

29. Suthers D, Toth E, Weiner A. An integrated approach to implementing collaborative inquiry in the classroom. Computer-Supported Collaborative Learning (CSCL'97), Toronto, Ontario, 1997.

30. Stein NL, Miller CA. The development of memory and reasoning skill in argumentative contexts: evaluating, explaining, and generating evidence. In: Glaser R (ed). Advances in Instructional Psychology. Vol 4. Hillsdale, N.J.: Lawrence Erlbaum, 1993.

31. Okada T, Simon HA. Collaborative discovery in a scientific domain. Cognitive Sci. 1997;21(2):109–46.

32. Coughlin LD, Patel VL. Processing of critical information by physicians and medical students. J Med Educ. 1987(62): 818.

33. Klein GA, Calderwood R. Decision models: some lessons from the field. IEE Trans Systems Man Cybernetics. 1991; 21(5):1018–26.

34. Runkel PJ. Casting Nets and Testing Specimens: Two Grand Methods of Psychology. Westport, Conn.: Praeger, 1990.

35. Ericsson A, Simon HA. Protocol Analysis: Verbal Reports as Data. Cambridge, Mass: MIT Press, 1993.

36. Patel VL, Kaufman DR, Magder SR. The acquisition of medical expertise in complex dynamic environments. In: Ericsson KA (ed). The Road to Excellence: The Acquisition of Expert Performance in the Arts and Sciences, Sports and Games. Hillsdale, N.J.: Lawrence Erlbaum, 1996:127–65.

37. Patel VL, Kaufman DR. Medical informatics and the science of cognition. J Am Med Inform Assoc, in press.

38. Shahar Y, Musen MA. Knowledge-based temporal abstraction in clinical domains. Artif Intell Med. 1996;8(3):267–98.

39. Greenberg DA. Encephalopathy. In: Don H (ed). Decision Making in Critical Care. 2nd ed. St. Louis, Mo.: Mosby, 1985: 6–7.

40. Patel VL, Ramoni MF. Cognitive models of directional inference in expert medical reasoning. In: Feltovich PJ, Ford KM, Hoffman RR (eds). Expertise in Context: Human and Machine. Cambridge, Mass.: MIT Press, 1997:67–99.

APPENDIX

An Excerpt from the Coding of the Harvard Dialogue

Episode 2
  Goal: To determine the purpose of the guideline
  Number of conversational turns: 4
  Number of participants: 2
  Type of goal: Purpose of the guideline
  Type of exchange: Clarification
  Knowledge added: None
  Support: From text
  End state: Goal fulfulled

Episode 3
  Goal: Determining how GLIF encoding will be followed after section B
  Number of conversational turns: 11
  Number of participants: 2
  Type of goal: Representation
  Type of exchange: Difference of opinion, question and answer

  Knowledge added: None
  Support: None
  End State: Resolved goal as turning this into GLIF, even though problems

Episode 4
  Goal: The entry condition for the guideline
  Number of conversational turns: 4
  Number of participants: 2
  Type of goal: Representation
  Type of exchange: Question and answer
  Knowledge added: A separate component to define the entry condition (C#2)
  Support: None
  End state: Determined the entry condition/Developed a plan for implementation