16

Experience Using EMYCIN

James S. Bennett and Robert S. Engelmore

The development of expert systems is plagued with a well-known and crucial bottleneck: in order for these systems to perform at all the domain-specific knowledge must be engineered into a form that can be embedded in the program. Advances in understanding and overcoming this knowledge acquisition bottleneck rest on an analysis of both the process and the product of our current, rather informal interactions with experts. To this end the purpose and structure of two quite dissimilar rule-based systems are reviewed. Both systems were constructed using the EMYCIN system after interviewing an expert. The first, SACON (Bennett et al., 1978), is meant to assist an engineer in selecting a method to perform a structural analysis; the second, CLOT (Bennett and Goldman, 1980), is meant to assist a physician in determining the presence of a blood clotting disorder.

The presentation of the details of these two systems is meant to accomplish two functions. The first is to provide an indication of the scope and content of these rule-based systems. The reader need not have any knowledge of the specific application domain; the chapter will present the major steps and types of inferences drawn by these consultants. This conceptual framework, what we term the *inference structure,* forms the basis for the expert's organization of the domain expertise and, hence, the basis for successful acquisition of the knowledge base and its continued maintenance. The second purpose of this chapter is to indicate the general form and function of these inference structures.

We first present the motivations and major concepts of both the SACON and CLOT systems. A final section then summarizes a number of observations about the knowledge acquisition process and the applicability of EMYCIN to these tasks. This chapter thus shows how the knowledge acquisition ideas from Chapter 9 and the EMYCIN framework from Chapter 15 have been used in domains other than infectious disease.

This chapter is a shortened and edited version of a paper appearing in *Pergamon-Infotech state of the art report on machine intelligence.* Maidenhead, Berkshire, U.K.: Infotech Ltd., 1981.

16.1 SACON: A Consultant for Structural Analysis

SACON (Structural Analysis CONsultant) was developed to advise nonexpert engineers in the use of a general-purpose computer program for structural analysis. The automated consultant was constructed using the EMYCIN system. Through a substitution of structural engineering knowledge for the medical knowledge, the program was converted easily from the domain of infectious diseases to the domain of structural analysis.

The purpose of a SACON consultation is to provide advice to a structural engineer regarding the use of a structural analysis program called MARC (MARC Corporation, 1976). The MARC program uses finite-element analysis techniques to simulate the mechanical behavior of objects, for example, the metal fatigue of an airplane wing. Engineers typically know *what* they want the MARC program to do—e.g., examine the behavior of a specific structure under expected loading conditions—but do not know *how* the simulation program should be set up to do it. The MARC program offers a large (and, to the novice, bewildering) choice of analysis methods, material properties, and geometries that may be used to model the structure of interest. From these options the user must learn to select an appropriate subset of methods that will simulate the correct physical behavior, preserve the desired accuracy, and minimize the (typically large) computational cost. A year of experience with the program is required to learn how to use all of MARC's options proficiently. The goal of the automated consultant is to bridge this "what-to-how" gap, by recommending an analysis strategy. This advice can then be used to direct the MARC user in the choice of specific input data—e.g., numerical methods and material properties. Typical structures that can be analyzed by both SACON and MARC include aircraft wings, reactor pressure vessels, rocket motor casings, bridges, and buildings.

16.1.1 The SACON Knowledge Base

The objective of a SACON consultation is to identify an *analysis strategy* for a particular structural analysis problem. The engineer can then implement this strategy, using the MARC program, to simulate the behavior of the structure. This section introduces the mathematical and physical concepts used by the consultant when characterizing the structure and recommending an analysis strategy.

An analysis strategy consists of an *analysis class* and a number of associated *analysis recommendations.* Analysis classes characterize the complexity of modeling the structure and the ability to analyze the material behaviors of the structure. Currently, 36 analysis classes are considered;

among them are Nonlinear Geometry Crack Growth, Nonlinear Geometry Stress Margin, Bifurcation, Material Instability, Inelastic Stiffness Degradation, Linear Analysis, and No Analysis. The analysis recommendations advise the engineer on specific features of the MARC program that should be activated when performing the actual structural analysis. (The example consultation in Figure 16-3 concludes with nine such recommendations.)

To determine the appropriate analysis strategy, SACON infers the critical material stress and deflection behaviors of a structure under a number of loading conditions. Among the material stress behaviors inferred by SACON are Yielding Collapse, Cracking Potential, Fatigue, and Material Instabilities; material deflection behaviors inferred by SACON are Excessive Deflection, Flexibility Changes, Incremental Strain Failure, Buckling, and Load Path Bifurcation.

Using SACON, the engineer decomposes the structure into one or more *substructures* and provides the data describing the materials, the general geometries, and the boundary conditions for each of these substructures. A substructure is a geometrically contiguous region of the structure composed of a single material, such as high-strength aluminum or structural steel, and having a specified set of kinematic boundary conditions. A structure may be subdivided by the structural engineer in a number of different ways; the decomposition is chosen that best reveals the worst-case material behaviors of the structure.

For each substructure, SACON estimates a numeric *total loading* from one or more *loadings.* Each loading applied to a substructure represents one of the typical mechanical forces on the substructure during its working life. Loadings might, for example, include loadings experienced during various maneuvers, such as braking and banking for planes, or, for buildings, loadings caused by natural phenomena, such as earthquakes and windstorms. Each loading is in turn composed of a number of point or distributed *load components.*

Given the descriptions of the component substructures and the descriptions of the loadings applied to each substructure, the consultant estimates stresses and deflections for each substructure using a number of simple *mathematical models.* The behaviors of the complete structure are found by determining the sum of the peak relative stress and deflection behaviors of all the substructures. Based on these peak responses (essentially the worst-case behaviors exhibited by the structure), its knowledge of available analysis types, and the tolerable analysis error, SACON recommends an analysis strategy. Figure 16-1 illustrates the basic types of inferences drawn by SACON during a consultation.

Judgmental knowledge for the domain, and about the structural analysis task in particular, is represented in EMYCIN in the form of production rules. An example of a rule, which provides the transition from simple numeric estimates of stress magnitudes to symbolic characterizations of stress behaviors for a substructure, is illustrated in Figure 16-2.

One major feature of EMYCIN that was not used in this task was the

Analysis Strategy of the Structure
↑
Worst-Case Stress and Deflection
Behaviors of the Structure
↑
Symbolic Stress and Deflection
Behaviors of Each Substructure
↑
Composite Numeric Stress and Deflection
Estimations of Each Loading
↑
Numeric Stress and Deflection
Magnitudes of Each Load Component

FIGURE 16-1 Inference structure during a SACON consultation. The user specifies loading and substructure descriptions that the system uses to infer material behaviors and, finally, an analysis strategy.

certainty factor mechanism—i.e., the ability to draw inferences using uncertain knowledge. The present consultation strategy and the associated mathematical models were designed to estimate extreme loading conditions, from which SACON concludes the appropriate analysis class. Consequently, by using a conservative model, the rules, though inexact, are sufficiently accurate for predicting response bounds with certainty.

The existing knowledge base is able to select from among 36 nonlinear analysis strategies. If nonlinear analysis is not indicated by the response estimates, the consultation recommends linear analysis. In addition, if relative stress and displacement estimates are low (less than 5% of critical values), the consultation indicates that no analysis is required. The knowledge base consists of 170 rules and about 140 consultation parameters. A

```
RULE050

IF:  1) The material composing the sub-structure is one of: the metals, and
     2) The analysis error (in percent) that is tolerable is between 5 and 30, and
     3) The non-dimensional stress of the sub-structure is greater than .9, and
     4) The number of cycles the loading is to be applied is between 1000 and 10000
THEN:  It is definite (1.0) that fatigue is one of the stress behavior phenomena in the sub-structure

PREMISE:  ($AND (SAME CNTXT MATERIAL (LISTOF METALS))
                (BETWEEN* CNTXT ERROR 5 30)
                (GREATER* CNTXT ND-STRESS .9)
                (BETWEEN* CNTXT CYCLES 1000 10000))
ACTION:  (CONCLUDE CNTXT SS-STRESS FATIGUE TALLY 1.0)
```

FIGURE 16-2 A sample rule from SACON, showing both the English and LISP forms.

typical consultation (2 substructures, 3 loadings, 3 load components) requires about 25 minutes in an interactive session.

The transcript in Figure 16-3 illustrates SACON's interactions with an engineer and SACON's recommendation of an analysis strategy. The engineer's responses to questions follow the prompt (**) and are printed in bold capitals.

July 26, 1978 10:54AM in <SACON>MARC.SAV;80709 by SACON>

Hello Dr. Melosh,

This consultation concerns the 747 wing. The purpose of the consultation is to demonstrate the general features of a structural analysis consultation on a simple problem. Thus the consultation involves only one substructure: The total wing.

26-Jul-78 10:54:44
[consultation of 13-Jul-78 1:55PM].
--------STRUCTURE-28--------
1) What is the name of STRUCTURE-28?
** **BOEING 747 WING**
2) Assuming that your characterization of the boeing 747 wing in terms of its geometry, material characteristics, and boundary conditions are accurate, how much error (in percent) is tolerable for the analysis?
** **10**
3) Do you want to examine the responses of the boeing 747 wing, to evaluate its instability, or both?
** **EXAMINE-RESPONSES**
4) Does the boeing 747 wing have any time dependent terms in its equations of equilibrium?
** **NO**
5) Is heating or cooling of interest in determining structure responses?
** **NO**
6) Have you decomposed the boeing 747 wing into a number of potential sub-structures which are relevant to this analysis?
** **YES**

--------SUB-STRUCTURE-1--------
7) What is the name of SUB-STRUCTURE-1?
** **TOTAL WING**
8) What is the material composing most of the total wing (SUB-STRUCTURE-1)?
** **HIGH-STRENGTH-ALUMINUM**
9) Enter the average longest dimension of the total wing (SUB-STRUCTURE-1) in inches (or <number> CMS):
** **1420**
10) Enter the weight of the total wing (SUB-STRUCTURE-1) in pounds (or <number> KGS):
** **46000**
11) Is the total wing (SUB-STRUCTURE-1) thin-walled or solid?
** **THIN-WALLED**
12) What is the configuration of support for the total wing (SUB-STRUCTURE-1)?
** **ONE-SIDE**
13) Is the total wing (SUB-STRUCTURE-1) to be modelled as a 1, 2, or 3 dimensional system?
** **2**
14) Would you describe the construction of the total wing (SUB-STRUCTURE-1) to be truss and/or frame elements (NETWORK), sheets and/or blocks of material (CONTINUUM), or a combination of both?
** **CONTINUUM**
15) Is the overall geometry of the total wing (SUB-STRUCTURE-1) curved or planar?
** **PLANAR**

FIGURE 16-3 A SACON consultation.

The first loading applied to the total wing (SUB-STRUCTURE-1) will be called:

--------LOADING-1--------
16) What is the name of LOADING-1?
** **FLIGHT**
17) Enter the number of loading cycles to be applied:
** **20000**
The first load component associated with LOADING-1 will be called:

--------LOAD-COMPONENT-1--------
18) Would you describe LOAD-COMPONENT-1 as being DISTRIBUTED over most of the sub-structure or as acting at a POINT of the sub-structure?
** **DISTRIBUTED**
19) Which surface of the sub-structure does LOAD-COMPONENT-1 act NORMAL to? (If more than one surface, you should consider the loading as two or more component loadings)
** **WIDTH-LENGTH**
20) Enter the average wall thickness of the total wing (SUB-STRUCTURE-1) in inches (or <number> CMS):
** **.31**
21) Enter the average depth of the total wing (SUB-STRUCTURE-1) in inches (or <number> CMS):
** **41**
22) Enter the magnitude of the distributed load (in psi):
** **1**
23) Are there any other load components associated with LOADING-1?
** **NO**
24) Are there any other loading conditions associated with the total wing (SUB-STRUCTURE-1)?
** **NO**
25) Are there any other sub-structures of the boeing 747 wing relevant to this analysis?
** **NO**
26) Do the supports of the boeing 747 wing involve Coulomb friction, nonlinear springs, and/or gapping?
** **NO**

The following analysis classes are relevant to the analysis of your structure:
1) general-inelastic

The following recommendations apply to this case:

Activate incremental stress—incremental strain analysis.

Model nonlinear stress-strain relation of the material.

Solution will be based on a mix of gradient and Newton methods.

Logic to scan peak stress at each step and evaluate fatigue integrity should be used.

Logic to scan stresses, smooth, and compare with allowable stresses (with appropriate safety factors) should be used.

Logic to scan deflections, calculate relative values, and compare with code limits, should be called upon.

Cumulative strain damage should be calculated.

Analysis should include two or more load cycles (if cyclic) with extrapolation for strain accumulation.

Shakedown extrapolation logic should be used.

A single cycle of loading is sufficient for the analysis.

Do you wish advice on another structure?
** **NO**

16.2 CLOT: A Consultant for Bleeding Disorders

In a different, and in some ways more standard, application of EMYCIN, we have recently developed a prototype of a consultant called CLOT, which advises physicians on the presence and types of disorders of the human coagulation system. CLOT was constructed by augmenting the EMYCIN system with domain-specific knowledge about bleeding disorders encoded as production rules. Section 16.3 describes the general structure of the CLOT knowledge base.

Our primary intent in constructing CLOT was to explore knowledge acquisition techniques that might be useful during the initial phases of knowledge base specification. Thus we sought to determine the primary inference structures and preliminary medical concepts that a consultant might require. We acquired the initial medical expertise for CLOT from a third-year medical student within a brief amount of time. This expertise has not yet been refined by an acknowledged expert physician. We conjecture that with these structures now in place the arduous task of detailing the knowledge required for truly expert performance can proceed at a more rapid pace. However, we have not had the opportunity to test this conjecture (cf. Mulsant and Servan-Schreiber, 1984).

16.3 The CLOT Knowledge Base

The primary objective of a CLOT consultation is to identify the presence and type of bleeding defect in a patient. If a defect is diagnosed, the consultant attempts to refine its diagnosis by identifying the specific conditions or syndromes in the patient and their plausible causes. These refined diagnoses can then be used by the physician to evaluate the patient's clinical status and to suggest possible therapies. At present, CLOT makes no attempt to recommend such therapies. This section briefly introduces the physiological basis and inference structure used by the consultant when characterizing the bleeding defect of the patient.

There are two major types of bleeding disorders, corresponding to defects in the two component subsystems of the human coagulation system. The first subsystem, termed the platelet-vascular system, is composed of the blood vessels and a component of the blood, the platelets. Upon sustaining an injury, the blood vessels constrict, reducing the flow of blood to the injured area. This vasoconstriction in turn activates the platelets, causing them to adhere to one another and form a simple, temporary "plug," or thrombus. This thrombus is at last reinforced by fibrin, a protein resulting from a complex, multienzyme pathway, the second component subsystem of the coagulation system. Fibrin converts the initial platelet plug into the more permanent clot with which most people are familiar. A defect in either the platelet-vascular or the coagulation (enzymatic) subsystem can cause prolonged and uncontrollable bleeds. For example, the familiar "bleeder's" disease (hemophilia) is the result of a missing or altered enzyme in the coagulation system, which inhibits the formation of fibrin and hence of the final clot.

CLOT was designed to be used eventually by a physician attending a patient with a potential bleeding problem. The system assumes that the physician has access to the necessary laboratory tests and the patient's medical history. CLOT attempts to diagnose the bleeding defect by identifying which of the two coagulation subsystems might be defective. This inference is based first on clinical evidence and then, independently, on the laboratory findings. Finally, if these independent conclusions are mutually consistent, an overall estimation of the defect is deduced and reported.

The consultation begins with the collection of standard demographic data about the patient (name, age, sex, and race) followed by a review of the clinical, qualitative evidence for a bleeding disorder. The physician is asked to describe an episode of bleeding in terms of its location, whether its onset was immediate or prolonged, and whether the physician feels the amount of bleeding was disproportionate for its type. Other factors such as the spontaneity of the bleeding, its response to applied pressure, and its persistence (duration) are also requested. These data are supplemented with facts from the patient's background and medical history to provide an estimate of the significance of the episode. These factors are then used to provide suggestive, but not definitive, evidence for the presence of a bleeding defect. This suggestive, rather than diagnostic, expertise was encoded using EMYCIN's certainty factor mechanism. Each rule mentions a key clinical parameter whose presence or absence contributes to the final, overall certainty of a particular bleeding disorder. (See Figure 16-4.)

The clinical description of the bleeding episode is followed by a report of the coagulation-screen test results. These six standard, quantitative measurements made of the patient's blood sample are used to determine if the blood clots abnormally. If the patient's blood does clot abnormally, CLOT attempts to infer what segment of the enzymatic pathway might be impaired and what platelet dysfunction might be present.

Finally, if clinical and laboratory evidence independently produce a mutually consistent estimation of the defect type, the case data and the intermediate inferences about the significance and possible causes of the bleed combine to produce a refined diagnosis for the patient. Currently, for patients experiencing a significant bleed, these conclusions include specific enzyme deficiencies, von Willebrand's syndrome, Kallikrein defects, thrombocytopenia, and thrombocytosis.

```
RULE025

IF:   1) Bleeding-history is one of the reasons for this consultation,
      2) There is an episode of significant bleeding in the patient,
      3) Coagulation-defect is one of the bleeding disorders in the patient,
      4) The defective coagulation pathway of the patient is intrinsic, and
      5) There are not factors which interfere with the patient's normal bleeding
THEN:  It is definite (1.0) that the following is one of the bleeding diagnoses of the patient: The
       patient has one or more of the following conditions: Hemophilia A, von Willebrand's
       syndrome, an IX, XI, or XII deficiency, or a high molecular weight Kallikrein defect.

PREMISE:  ($AND (SAME CNTXT REASON BLEEDING-HISTORY)
                (SAME CNTXT SIGBLD)
                (SAME CNTXT FINALDEF COAGULATION-DEFECT)
                (SAME CNTXT DEFPATH INTRINSIC)
                (NOTSAME CNTXT INTERFERENCE))
ACTION:  (CONCLUDETEXT CNTXT DX (TEXT DXHEMOPHILIA) TALLY 1000)
```

FIGURE 16-4 A sample rule from CLOT. RULE025 partly determines the final diagnosis.

Figure 16-5 indicates the major inferences drawn by CLOT during a consultation.

The current knowledge base is able to select from among 13 refined diagnoses and can indicate that the patient is a normal bleeder. CLOT is also able to detect possibly false negative test results due to various interfering factors such as recent exercise, pregnancy, presence of chronic illnesses, and the patient's use of drugs known to alter the clotting process. The knowledge base consists of 63 rules and 50 consultation parameters. A typical consultation requires about five minutes of interaction. A sample consultation is shown in Figure 16-6. The transcript in Figure 16-6 illustrates CLOT's interactions with a physician about a patient and CLOT's final diagnosis. The physician's responses follow the prompt (**) and are printed in bold capitals.

Final Diagnosis

↑

Final, Consistent Estimation
of the Bleeding Defect

↑

Estimation of Enzymatic and Platelet
Defects on the Basis of the Coagulation Battery

↑

Initial Estimation of Bleeding Defect Type
Based on Clinical Evidence

FIGURE 16-5 Inference structure during a CLOT consultation.

--------PATIENT-110--------
1) Name:
** **Pt110**
2) Age:
** **90 YEARS**
3) Sex:
** **FEMALE**
4) Race:
** **CAUCASIAN**
5) Please indicate your reason(s) for requesting this consultation:
** **BLEEDING-HISTORY**
6) What type of bleeding describes Pt110's most recent episode of bleeding?
** **HEMARTHROSIS**
7) Is there a history of a genetic bleeding disorder in Pt110's family?
** **YES**
8) Was the onset of the bleed immediate or delayed?
** **DELAYED**
9) BT:
** **5 MINUTES**
10) PT:
** **13**
11) PTT:
** **50**
12) TT:
** **15**
13) FSF:
** **NORMAL**
14) Has Pt110 recently exercised?
** **NO**
15) Is Pt110 currently receiving any of the following drugs: ASA, Heparin, Coumarin, oral-contraceptives, Ephedrine, Epinephrine, ADH?
** **NO**
16) Is Pt110 diagnosed as having cirrhosis, collagen disease, cancer, or any chronic disease?
** **NO**

Conclusions: the blood disorders of Pt110 are as follows:
COAGULATION-DEFECT (.97)
Conclusions: the statements about the consistency of the case data and CLOT's interpretation are as follows:

Both clinical and lab data are internally consistent and there is overall, consistent interpretation of the blood disorder.

Conclusions: the bleeding diagnoses of Pt110 are as follows:

The patient has one or more of the following conditions:

Hemophilia A, von Willebrand's syndrome, an IX, XI, or XII deficiency, or a high molecular weight Kallikrein defect. (.97)

FIGURE 16-6 Transcript of a CLOT consultation.

16.4 EMYCIN as a Knowledge Representation Vehicle

We did not find the representation formalism of EMYCIN to be a hindrance to either the formulation of the knowledge by the expert or its eventual implementation in either program. In fact, the simplicity of using and explaining both EMYCIN's rule-based formalism and its backward-chaining control structure actually facilitated the rapid development of the knowledge base during the early stages of the consultant's design. Moreover, the control structure, like the rule-based formalism, seemed to impose a salutary discipline on the expert during the formulation of the knowledge base.

The development of SACON was a major test of the domain-independence of the EMYCIN system. Previous applications using EMYCIN had been primarily medical, with the consultations focusing on the diagnosis and prescription of therapy for a patient. Structural analysis, with its emphasis on structures and loadings, allowed us to detect the small number of places where this medical bias had unduly influenced the system design, notably in the text strings used for prompting and giving advice.

Both the MARC expert and the medical student found that their knowledge was easily cast into the rule-based formalism and that the existing predicate functions and context-tree mechanism provided sufficient expressive power to capture the task of advising their respective clients. The existing interactive facilities for performing explanation, question answering, and consultation were found to be well developed and were used directly by our application. None of these features required any significant reprogramming.

EMYCIN provides many tools to aid the knowledge engineer during the process of embedding the expertise into the system. During the construction of CLOT we found that the knowledge acquisition tools in EMYCIN had substantially improved since the construction of SACON. These facilities now perform a large amount of useful checking and default specification when specifying an initial knowledge base. In particular, a new facility had been implemented that provides assistance during the specification of the context tree. This facility eliminates a substantial amount of user effort by setting up the multitude of data structures for each context and ensuring their mutual consistency. Furthermore, the facility for acquiring clinical parameters of a context now performs a significant amount of prompting and value checking on the basis of a simple parameter classification scheme; we found these facilities very useful.

We made extensive use of the ARL (Abbreviated Rule Language) facility when acquiring the rules for CLOT. Designed to capitalize on the stereotypically terse expression of rule clauses by experts, ARL reduces the amount of typing time and, again, ensures that the correct forms are used when specifying both the antecedent and consequent parts of a rule. For example, when specifying the CLOT rule shown in Figure 16-4, the medical student engaged in the interaction shown in Figure 16-7. The user's input follows a colon or a question mark.

In addition to ARL, EMYCIN's rule-subsumption checker also proved very useful during the specification of larger rule sets in the system. This checker analyzes each new rule for possible syntactic subsumptions, or equivalences with the premise clauses of the other rules. We found that,

```
Enter Parms, Rules, Save changes, or Go? Rules
Rule number of NEW: NEW
RULE025
PREMISE: (REASON = BLEEDING, SIGBLD, FINALDEF = COAGULATION,
          DEFPATH = INTRINSIC ~ INTERFERENCE)
RULE025
ACTION: (DX = DXHEMOPHILIA)
BLEEDING → BLEEDING-HISTORY? Yes
COAGULATION → COAGULATION-DEFECT? Yes
Translate, No further changes, or prop name:
```

FIGURE 16-7 Interaction with EMYCIN, using the Abbreviated Rule Language (ARL) to specify the CLOT rule shown in Figure 16-4.

for the larger rule sets, the checker detected these inconsistencies, due to either typing mistakes or actual errors in the rule base logic, and provided a graceful method for dealing with them. Together these facilities contributed to the ease and remarkable rapidity of construction of this consultant. For further details on the design and operation of these aids, see van Melle (1980).

16.5 Observations About Knowledge Acquisition

To bring the SACON program to its present level of performance, we estimate that two person-months of the expert's time were required to explicate the consultation task and formulate the knowledge base, and about the same amount of time was required to implement and test the rules. This estimate does not include the time devoted to meetings, problem formulation, demonstrations, and report writing. For the first 170 rules in the knowledge base, we estimate the average time for formulating and implementing a rule was about four hours. The marginal time for a new rule is about two hours.[1]

The construction of CLOT required approximately three days, divided as follows. The first day was spent discussing the major medical concepts, clinical setting, and diagnostic strategies that were appropriate for this consultant. At the end of this period, the major subtasks of the consultant had been sketched, and a large portion of the clinical parameters the consultant would request of the physician had been mentioned. The following

[1]These estimates represent a simple average that held during the initial construction of these projects. They do not reflect the wide variation in the amount of effort spent defining rules versus the other knowledge base development tasks that occurred over that time period.

two days were spent detailing aspects of the parameters and rules that the EMYCIN system required (i.e., specifying expected values, allowable ranges on numeric parameters, question formats, etc.) and entering these details into the system itself. We may approximate the average cost of formulating and implementing a rule in such a system based on the number of person-hours spent in construction versus the number of rules specified. CLOT required about 60 person-hours to specify 60 rules yielding a rate of 1 person-hour per rule. The marginal cost for a new rule is expected to be similar.

Our experience explicating these rule bases provided an opportunity to make some observations about the process of knowledge acquisition for consultation systems. Although these observations were made with respect to the development of SACON and CLOT, other knowledge-based consultation systems have demonstrated similar processes and interactions.

Our principal observation is that the knowledge acquisition process is composed of three major phases. These phases are characterized strongly by the types of interaction that occur between expert and knowledge engineer and by the type of knowledge that is being explicated and transferred between the participants during these interactions. At present only a small fraction of these interactions can be held directly with the knowledge-based system itself (Davis, 1976; 1977), and research continues to expand the knowledge acquisition expertise of these systems.

16.5.1 The Beginning Phase

The beginning phase of the knowledge formalization process is characterized by the expert's ignorance of knowledge-based systems and unfamiliarity with the process of explicitly describing exactly what he or she knows and does. At the same time, the knowledge engineers are notably ignorant about the application domain and clumsily seek, by analogy, to characterize the possible consultation tasks that could be performed (i.e., "Well, in MYCIN we did this . . . . ").

During the initial weeks of effort, the domain expert learns what tools are available for representing the knowledge, and the knowledge engineer becomes familiar with the important concepts of the domain. During this period, the two formulate a taxonomy of the potential consultation areas for the application of the domain and the types of advice that could be given. Typically, a small fragment of the complete spectrum of consultation tasks is selected to be developed during the following phases of the knowledge acquisition effort. For example, the MYCIN project began by limiting the domain of expertise to the diagnosis and prescription of therapy for bacteremia (blood infections); SACON is currently restricted to determining analysis strategies for structures exhibiting nonlinear, nonthermal, time-independent material behaviors.

Having decided on the subdomain that is to be developed and the type of advice that is to be tendered, the team next identifies the major factors (parameters) and reasoning steps (rules) that will be used to characterize the object of the consultation (be it patient or airplane wing) and to recommend any advice. This forms the inference structure of the consultant.

16.5.2 The Middle Phase

After this initial conceptual groundwork is laid, work proceeds to detailing the reasoning chains and developing the major rule sets in the system. During the development of these rule sets, the amount of domain vocabulary, expressed as contexts, parameters, and values, increases substantially. Enough knowledge is explicated during this middle phase to advise a large number of common cases.

While developing these systems, we profited by "hand-simulating" any proposed rules and parameter additions. In particular, major advances in building the structural analysis knowledge base came when the knowledge engineer would "play EMYCIN" with the expert. During the sessions the knowledge engineer would prompt the expert for tasks that needed to be performed. By simulating the backward-chaining manner of EMYCIN, we asked, as was necessary, for rules to infer the parameter values, "fired" these rules, and thus defined a large amount of the parameter, object, and rule space used during the present consultations. This process of simulating the EMYCIN system also helped the expert learn how the program worked in detail, which in turn helped him develop more rules and parameters.

16.5.3 The Final Phase

Finally, when the knowledge base is substantially complete, the system designers concentrate on *debugging* the existing rule base. This process typically involves the addition of single rules to handle obscure cases and might involve the introduction of new parameters. However, the major structure of the knowledge base remains intact (at least for this subdomain), and interactions with the expert involve relatively small changes. (Chapters 8 and 9 describe debugging and refining a knowledge base that is nearly complete.)

The initial development of the knowledge base is greatly facilitated when the knowledge engineering team elicits a well-specified consultation goal for the system as well as an inference structure such as that depicted in Figure 16-1. Without these conceptual structures to give direction to the knowledge explication process, a confused and unusable web of facts typically issues from the expert. We speculate that the value of these organizational structures is not restricted to the production system methodology. They seem to be employed whenever human experts attempt to formalize their knowledge in any representation formalism, be it production rules, predicate calculus, frames, etc. Indeed, when difficulties arise in building a usable knowledge base, we suspect that the trouble is as likely to come from a poor choice of inference structure as from the choice of any particular representation scheme.

The inference structure is a form of *meta-knowledge,* i.e., knowledge about the structure and use of the domain expertise (see Part Nine). Our experience shows that this meta-knowledge should be elicited and discussed early in the knowledge acquisition process, in order to insure that a sufficient knowledge base is acquired to complete a line of reasoning, and to reduce the time and cost of system development. Also, Chapter 29 discusses the need to explain such meta-level knowledge.

Making the inference structure an explicit part of the program would assist the explanation, tutoring, and further acquisition of the knowledge base. Several researchers, including Swartout (1981) and Clancey (1979b), have employed portions of the inference structure to guide both the design and tutoring of a knowledge-based system. The success of this work supports the hypothesis that the inference structure will play a critical role in the development of new knowledge-based consultation systems.