Exploiting Linguistic and Societal Metaphors
for Knowledge Acquisition

Stephen Michael Griffin

M.Phil. Thesis

Submitted to the University of Surrey in partial fulfilment of
requirements for the degree of Master of Philosophy

Department of Mathematical and Computing Sciences
University of Surrey
Guildford
Surrey

July 1997
ABSTRACT

Our inter-disciplinary research examines new approaches to knowledge acquisition through the exploitation of linguistic and societal metaphors. We argue that conventional knowledge acquisition relies too heavily on a psychological metaphor, and that this is insufficient in broad domains, where geographical and political issues make the expertise more socially situated, because it lacks input from the society in which the knowledge exists. We attempt to provide a methodology which captures this input by introducing a Domain Interface Group to support the knowledge engineer in his/her tasks. This presents a changing role for the knowledge engineer to primarily that of a group facilitator, and we suggest guidelines for brainstorming sessions to facilitate consensus decision making. We advocate the continued use of expert interviews, but suggest ways to improve their productivity. In particular, we attempt to alleviate reductive bias through the use and understanding of domain specific terminology and lexical semantics, during all domain communication and particularly during knowledge acquisition from text. We situate our work in the constructivist modelling paradigm and describe mediating representations which emphasize the importance of human comprehension of the model, for the knowledge engineer, the expert and the end user, above programming considerations. We have undertaken an evaluation of our methodology and an audit of a resulting paper knowledge base, and present the results in an attempt to prove the efficiency, effectiveness and accuracy of our approach.
I would like to thank Khurshid Ahmad for his enthusiasm and support as my supervisor over the past six years. Our working relationship has always been pleasurable and rewarding, for one of us at least, and I hope that this does not signify its end.

I would like to acknowledge the many members of the AI group with whom I have had the pleasure to work over the years. I must also thank Professor Steve Schuman and Professor Trevor Sweeting for providing the supportive environment of the Mathematical and Computing Sciences Department. I particularly wish to acknowledge Paul Holmes-Higgin, Andrea Davies, Steve Hook and Peter and Caroline Jones for the provision of a most relaxed and enjoyable atmosphere in which to work in the past, and with whom I still look forward to spending time after so many years, albeit in a more social environment nowadays. I would also like to thank Sarah Lewis for taking the time and interest to help with the proof-reading of this manuscript, and numerous colleagues at PA Consulting Group for their constructive comments and enthusiastic support during the gruelling months of writing-up.

I wish to say a special thank you to Anthea Jeffcoat, without whose help, patience, understanding and moral support, the preparation of this thesis could not have been completed on time. Now what are we going to do with our evenings, weekends, holidays ...?

Above all, I wish to thank my parents, for supporting me throughout my education and beyond. For their confidence in my ability, their faith in my future, and their continued love and support, this thesis is dedicated to them.
# CONTENTS

LIST OF FIGURES iv
LIST OF TABLES vi

1. INTRODUCTION 1

1.1 MOTIVATION 1
1.2 BACKGROUND 2
1.3 OUTLINE OF THE THESIS 7

2. RECENT TRENDS IN KNOWLEDGE ACQUISITION 17

2.1 A CHANGE IN APPROACH 17
2.2 EXPERTS AND EXPERTISE 19
2.3 A CONSTRUCTIVIST METAPHOR 23
2.4 INTERVIEWING 26
   2.4.1 Descriptive elicitation 27
   2.4.2 Structured Expansion 28
   2.4.3 Scripting 29
   2.4.4 Validation 30
   2.4.5 Performance 31
   2.4.6 Tools 33
2.5 BRAINSTORMING 33
2.6 MEDIATING REPRESENTATIONS 34
2.7 LANGUAGE (TERMINOLOGY AND TEXT ANALYSIS) 40
2.8 EXPERT SYSTEM ‘GANGS’: SOCIETAL INPUT FROM THE DOMAIN COMMUNITY? 42
2.9 AUTOMATED KNOWLEDGE ACQUISITION TOOLS 45
3. EVOLUTION OF OUR METHODOLOGY

3.1 AI RESEARCH AT THE UNIVERSITY OF SURREY

3.2 OUR CASE STUDIES

3.3 SOCIOLOGY: A NEWMETAPHOR FOR KNOWLEDGE ACQUISITION?
  3.3.1 Society
  3.3.2 Group
  3.3.3 Peer group
  3.3.4 Group Dynamics
  3.3.5 Consensus
  3.3.6 Language

3.4 LANGUAGE AS A KNOWLEDGE ACQUISITION METAPHOR
  3.4.1 Understanding and using specialist language
  3.4.2 Improvements in text analysis and its affect on knowledge acquisition
  3.4.3 Automatic elicitation of propositions from text
  3.4.4 Benefits of the Linguistic Approach

3.5 ENHANCING A CLASSICAL APPROACH WITH SOCIETAL AND LINGUISTIC METAPHORS

4. A PSYCHOLOGICAL, SOCIETAL AND LINGUISTIC METHODOLOGY

4.1 OVERVIEW

4.2 THE PLAYERS

4.3 INITIATION

4.4 INTERVIEWS
  4.4.1 Identify interview topics
  4.4.2 Outline interview content
  4.4.3 Identify interview participants
LIST OF FIGURES

Figure 1: Knowledge transfer from the 'real-world' into a computer ........................................ 4
Figure 2: Building on conventional knowledge acquisition techniques ................................ 11
Figure 3: The growing role of language and society in our knowledge acquisition methodology .................................................. 12
Figure 4: Towards a new methodology for knowledge acquisition? ........................................ 14
Figure 5: Comparison of Think-aloud protocols and focused interviews for brittleness .......... 31
Figure 6: Types of psychological interview techniques leading to different types of expert system .......................................................... 32
Figure 7: Mediating representations facilitate communication between domain expert and knowledge engineer (from Bradshaw and Boose, 1993) ................................................................. 36
Figure 8: Three-schema architecture for knowledge acquisition tools (from Bradshaw and Boose, 1993) ................................................................. 39
Figure 9: Mediating Representations (from Boose et al., 1993) .................................................... 40
Figure 10: Hierarchy of Domain Experts used in the DENDRAL Experiments .................................................................................. 43
Figure 11: Tentative ranking of experts involved in the MYCIN experiments .............................................. 45
Figure 12: University of Surrey AI Group systems ......................................................................... 49
Figure 13: Mapping of concepts within sociology to the artefacts of knowledge engineering ................................................................. 59
Figure 14: Concordance extract from an interview transcript ......................................................... 72
Figure 15: Compound terms from the interview transcript ................................................................ 73
Figure 16: Knowledge acquisition methodology in DIMES .................................................................. 76
Figure 17: Knowledge acquisition methodology in W-RAISA .................................................................. 77
Figure 18: Knowledge acquisition methodology in ELSIE ..................................................................... 79
Figure 19: Increased performance in knowledge elicitation from DIMES to ELSIE .......................... 81
Figure 20: Outline of Chapter 4 with respect to stages of our methodology ........................................ 83
Figure 21: The Domain Interface Group ............................................................................................... 87
Figure 22: Our approach to brainstorming the DIG ............................................................................. 90
Figure 23: Excerpt from the output of a brainstorming session with the ELSIE DIG to outline an expert interview ............................................................................................................................... 98
Figure 24: Questionnaire responses regarding the domain interface group .................................. 120
Figure 25: How much of the knowledge acquired could each member of the DIG provide alone? (none - all) 120
Figure 26: How long would each member of the DIG take to provide the knowledge? (days - years) 121
Figure 27: Questionnaire responses regarding domain coverage of the knowledge 121
Figure 28: How much of the domain do you think is covered by the implemented system? (none - all) 122
Figure 29: Questionnaire responses regarding the importance of different techniques 123
Figure 30: Questionnaire responses regarding expert systems in the workplace 124
Figure 31: Questionnaire responses relating to usefulness of our knowledge acquisition techniques 125
Figure 32: Questionnaire responses regarding the choice of interview participants 126
Figure 33: Questionnaire responses regarding our knowledge acquisition tools 127
Figure 34: How much knowledge do the members of the DIG think has been documented in the project that is undocumented elsewhere? (none - all) 128
Figure 35: DIG's opinion on each expert's charisma 129
Figure 36: DIG's opinion on each expert's authority 130
Figure 37: DIG's opinion on each expert's cleverness 130
Figure 38: DIG's opinion on each expert's knowledge 131
Figure 39: DIG's opinion on each expert's experience 131
Figure 40: Number of rules per score out of 10 in the paper knowledge base audit 132
Figure 41: Percentage of rules in the paper knowledge base edited, deleted, added or unchanged during the audit 133
LIST OF TABLES

Table 1: Processes, methods and knowledge roles for brainstorming and score systems 34

Table 2: Relative frequency of closed class words in special and general language 71

Table 3: Relative frequency of terms in special and general language and their ratios 71

Table 4: Possible object-property-value tuples elicited from the concordance 73

Table 5: Examples of semantic cues for locating 'rules' in text 74

Table 6: Example of candidate rule found with semantic cues 74

Table 7: Evolution of a Knowledge Acquisition Methodology 80

Table 8: Techniques used, adapted and evolved for knowledge acquisition projects since 1988 at the University of Surrey 82

Table 9: The changing focus of task responsibility in DIMES, W-RAISA and ELSIE 82

Table 10: Roles of the Domain Interface Group 91

Table 11: Criteria to be considered / not considered during expert selection 100
1. INTRODUCTION

1.1 Motivation

The primary objective of our work described in this thesis was to find an effective methodology for the acquisition of knowledge in domains where conventional knowledge acquisition techniques had previously been unsuccessful. By conventional techniques we refer to those evolved principally from psychology such as interviewing and think-aloud protocols. The result of our work is a knowledge acquisition methodology which includes these techniques but within a much broader framework, the basis of which owes more to the social sciences than psychology. We emphasise the importance of domain specific language and the peer group in which an expert’s knowledge exists, suggesting that this leads to a more accurate model of the domain.

A second objective was to make such a methodology as efficient as possible, keeping down costs and human resources to a level affordable by the domains in question. This is a pragmatic objective imposed by the circumstances of projects which provide us with the opportunity to research new techniques and innovative uses of existing ones, which must be met if this, or any other knowledge acquisition methodology, is to progress from the research laboratory into the information technology workplace. The resultant methodology is a cohesion of scientific and pragmatic problems and solutions.

Artificial Intelligence, or AI, can be construed as an “experimental science”, in which questions relating to knowledge representation, reasoning and dissemination have been
discussed extensively. However, the issue of knowledge acquisition has not been addressed as enthusiastically, nor as optimistically. This thesis aims to redress the balance. We describe how the application of our techniques, new and old, avoid some of the classic problems associated with knowledge acquisition and provide an effective and efficient methodology for a knowledge engineer to follow. Further, through subsequent analysis of these techniques in practice, we identify what can be done to improve the techniques in order that better results be achieved in the future.

Our results are aimed at knowledge engineers of all levels. For the novice or trainee, we intend to provide an efficient methodology whilst distancing the knowledge engineer from some of the more difficult tasks associated with conventional knowledge acquisition. For the more experienced knowledge engineer, the work carries a message that there are tools and techniques available from disciplines which he/she may not have explored, perhaps not having the opportunity to step back and look at his/her role from any angle other than a pragmatic one.

1.2 BACKGROUND

Our work is based on five years of practical knowledge acquisition in the building of prototype knowledge based systems for "real world" clients. Although only prototypes developed under research and development funding, these systems were always intended to reach the desktop and be used nation-wide. The methodology outlined here has evolved during this time in response to specific difficulties encountered during the practical application of conventional techniques.
The evolution of our methodology has been directed by the advancements in many related but not often associated disciplines, and verification of our contribution to the field has constantly been sought through the publishing of papers in journals and at international conferences, as well as through personal communication with some of the leading experts and institutions within the knowledge engineering community.

The efficiency of a methodology is dependent to a large extent on the domain in which it is to be used. Psychology based techniques, such as expert interviews, were used in the knowledge acquisition phase of classic expert system projects such as DENDRAL and MYCIN (Buchanan and Shortliffe, 1985) with arguably enormous success.

Put "abstractly and somewhat simplistically" by Yost and Newell (1989), the conventional paradigm of expert system development starts with a domain expert articulating the means of performing a task in a natural language. A knowledge engineer then comprehends this task knowledge expressed in the natural language, resulting in a conceptualisation of the knowledge in terms of the task domain. Next, the knowledge engineer maps the task knowledge from the terms of the task domain to the terms of some computational domain or model. Finally, the knowledge engineer composes a set of statements that express the computational conceptualisation of the task in a computer language. Together, the comprehension, domain mapping, and composition are referred to as operationalisation of the task knowledge.

A natural language must be used because both the domain expert and the knowledge engineer must be familiar with it, and it must permit clear and concise description of the task.
knowledge. For the remainder of this thesis we assume that the natural language is English. Yost and Newell stated that "operationalisation remains a task for humans, rather than computers, because natural language comprehension is routine for humans but is much too difficult to perform automatically. Further, operationalisation remains a task for knowledge engineers, rather than the domain experts, because the latter rarely are skilled in the use of computer languages".

The dominant methodology for knowledge elicitation is that of psychological interviewing, where the knowledge engineer and domain expert are involved in a Socratic dialogue, after which the knowledge engineer spends a considerable amount of time animating the knowledge acquired from the expert. Figure 1 illustrates the view of knowledge transfer from the expert to the computer, funnelled by the knowledge engineer to help it through the metaphorical “bottleneck” of knowledge acquisition.

![Figure 1: Knowledge transfer from the 'real-world' into a computer](image)

4
Our aim is to 'widen' the bottleneck by replacing the 'narrow' funnel at the knowledge engineer with a much 'wider' interface between the domain and the system. This wide domain-interface will be provided in two ways. First, by utilising domain practitioners to support the knowledge engineer in as many roles as possible, an approach introduced primarily to remove the reductive bias which accounts for the bottleneck. We call this team of practitioners the Domain Interface Group (DIG).

Second, we aim to support domain experts and the DIG to co-operatively build models of expert knowledge by providing mediating representations in natural language and structures designed to be clear and comprehensible to humans, not necessarily to be easily processed by computers. In this way we concentrate on 'modelling' the knowledge acquired without worrying about the obvious restrictions imposed during implementation. We argue, like Ford et al. (1993), that considerations of human efficiency far outweigh considerations for complex modelling problems.

Students receiving an introduction to Artificial Intelligence often believe that psychological techniques provide the only assistance available to knowledge engineers during the knowledge acquisition phase, a point mentioned here because this reliance on such a difficult subject is off-putting for engineers wishing to embark on expert system projects. We consider this approach as conventional knowledge acquisition because of its popularity, widespread acceptability and long-term usage and also because like most things referred to as "conventional", we believe that it is old-fashioned and needs modernising.
In recent years researchers have started to look for new techniques, progressing from the still popular interview techniques and think-aloud protocols (see for example Wood and Ford, 1993) to the automated use of Kelly’s Personal Construct Theory (Kelly, 1955, 1966, 1969, 1970; Agnew and Brown, 1989a,b; Mischel, 1964) and Ausubel’s Assimilation Theory (Ausubel, 1963; Ausubel et al. 1978) in the form of repertory grids and concept maps.

Bradshaw et al. (1993) reviewed the role of Kelly’s work in knowledge acquisition revealing that several forms of grid have been developed including implication grids, resistance-to-change grids, bipolar implication grids, dependency grids, exchange grids, and mode grids among others (see for example Adams-Webber, 1984; Fransella and Bannister, 1977; Hinkle, 1965; Shaw, 1981). These researchers have also devised a variety of grid formats in which people, objects, events, situations, or other kinds of elements, are either categorised, rated or rank-ordered on a set of constructs over which many newly developed analysis tools can be run.

We do not underestimate the importance of the role which psychology has to play in knowledge acquisition, and indeed devote substantial parts of this thesis to issues within the psychological metaphor which are as useful to knowledge engineers now as they were twenty years ago. However, we argue that psychology-based techniques need to be complemented by issues relating to the interaction of members within the domain community, from expert down to novice, and the understanding of the domain specific language.

We had varying success in the past with interviews and think-aloud protocols, but our circumstances led us to look more and more towards a societal approach. We judge that the
classical interview techniques lack vital input from the society in which the expert, and his/her knowledge, exist; a society without which the expert would be a nobody and his/her knowledge of no significance. We were missing something very important in the knowledge acquisition process, and we believe this was 'consensus': broad agreement within a very wide domain; the ability to please most of the people most of the time. We needed to consider the peer group in which the domain knowledge is extensively used, and the wider community within which the knowledge means everything, outside of which it means nothing.

Consensus is an issue which has caused problems for knowledge engineers for a long time, and a solution has so far eluded the community. Buchanan and Shortliffe, in their conclusions about more than ten years worth of MYCIN experiments, stated that “one important limitation of our model is its failure to address the problem of integrating knowledge from different experts... we have no tools for reaching a consensus” (1985). We believe that the DIG is a step towards lifting this limitation for expert system projects in the future.

1.3 OUTLINE OF THE THESIS

In recent years, several members of the knowledge acquisition community have made similar adaptations to their approach, moving towards a societal metaphor. They consider the importance of what Ford et al. (1993) refer to as the ‘expert-in-context’ and ‘socially situated expertise’, a central theme in our own work.

Gaines and Shaw (1994) reported how sociologists of science have characterised scientific communities as forming invisible colleges which monitor and manage the changing structure
of knowledge in their domain (see Crane, 1972) and that “individuals playing major roles in these communities are experts not only in overt knowledge of the domain but also through skills in its management and development”.

As academics we have over the past six years strived for answers to scientific questions relating primarily to what was, when we started our work, called the knowledge acquisition bottleneck. It is now, however, considered by the majority of knowledge acquisition researchers, ourselves included, that knowledge acquisition is a constructive modelling process, not simply a matter of ‘expertise transfer’.

Ford et al. (1993) point out that from a constructivist perspective, “we would expect that experts in the same domain are likely to agree about much of their knowledge (i.e. widely shared consensual beliefs) and yet each of them might rely also to a considerable extent on a unique fund of experience. Thus a critical task in knowledge acquisition research is the development of adequate tools and techniques for the purpose of assisting the knowledge engineer and expert in their task of collaboratively building a domain model. This modelling activity can make explicit the valuable personally constructed experience that experts frequently use, but are often unable to articulate.” This leads to a “troublesome knowledge engineering paradox” first described by Waterman (1986), that the more competent domain experts become, the less able they are to describe the knowledge they use to solve problems. As practising knowledge engineers, we believe that our consensus approach is a possible solution to this very practical problem.
In Chapter 2, we review knowledge acquisition’s evolution over recent years, discussing these issues of experts, expertise and its modelling in detail, hopefully situating our work within the context of current thinking and practice.

The acquisition of knowledge usually assumes that all experts in a given domain have similar experiences during their apprenticeship and in applying their knowledge. In some disciplines this may be true, however we have found that the geomorphological, agricultural and economic diversity within our research domain leads to substantial differences of opinion amongst experts, particularly at a geographically well-defined local level. For acceptability of a system across such boundaries, consensus must be achieved, something not straightforwardly acquired through conventional techniques. As also pointed out by Ford et al. (1993) “for broad domains political and nonrational criteria will weigh more heavily, and require different knowledge engineering strategies to locate the personally constructed and socially situated expertise.”

In keeping with the recommendations of Shaw and Woodward (1990), following the example set by Wood and Ford (1993), we discuss the working assumptions and principles that have guided our work before describing specific aspects of our techniques. Our experiments, represented by the development of an expert system for the Jarvis Breast Screening Clinic (DIMES) and two expert systems for the National Rivers Authority within the domain of water resource management (W-RAISA and ELSIE), focus on the problems of knowledge acquisition in emergent domains where knowledge is still being consolidated, where there are multiple knowledge sources, including legal, medical, environmental and engineering, and where consensus plays an important part. Named after the final system to be produced by the
author, background to these systems which make up the "ELSIE experiments" is provided in Chapter 3 along with a history of AI research at the University of Surrey.

This chapter also describes how our techniques evolved during the development of these systems. In particular, we discuss the importance of language and its components during knowledge acquisition, drawing from many linguistic disciplines including terminology, corpus management and text analysis. We show how modern text analysis tools can be customised to produce impressive results from archives of text such as manuals and codes of practice and from transcripts of expert interviews. We consider that the societal approach is the most important new metaphor in knowledge acquisition and we also include in Chapter 3 our attempt to draw links between aspects of knowledge engineering and current thinking in sociology.

Chapter 4 describes our methodology as both a theoretical framework for the elicitation of knowledge and a practical guide for a knowledge engineer to follow step-by-step if presented with a similar modelling problem in a similar domain. We show how we have built on a conventional knowledge acquisition technique with our experience in natural language processing and our experimentation in the societal approach, as illustrated in Figure 2.

The conventional approach to knowledge acquisition is deficient in many ways. The knowledge engineer attempts to elicit knowledge direct from the domain expert through a selection of psychology-based interview techniques and then analyses the results with little reference to the language content of the output (see Figure 1 earlier in this chapter).
Figure 2: Building on conventional knowledge acquisition techniques

We attempt to broaden the interface between domain and model through the introduction of domain practitioners to support the knowledge engineer in his/her many roles during knowledge acquisition, forming what we call a Domain Interface Group (DIG). The inclusion of members of the domain community, other than the experts, not only lowers the terminological barrier between the knowledge engineer and the specialist language of the experts, but may also help in ensuring a more accurate model of the domain.

We demonstrate how this societal metaphor can be used in conjunction with the conventional psychological knowledge acquisition technique of domain expert interviewing. Within this metaphor we describe the roles of the DIG: we suggest that many tasks, previously the sole responsibility of the knowledge engineer, can be executed more effectively and efficiently through the use of brainstorming and consensus decision making. We report, perhaps for the first time, the systematic use of group dynamics in knowledge acquisition.

The linguistic metaphor enables us to present corpus-based knowledge acquisition, in which sophisticated text analysis tools elicit knowledge automatically from domain texts and the
transcripts of expert interviews. We believe we are the first to exploit the ‘weirdness’ of scientific texts (Ahmad et al., 1993; Ahmad and Davies, 1994) to semi-automate the acquisition of knowledge from text. Figure 3 illustrates how language and society have played an increasing role in our knowledge acquisition methodology over recent years.

![Diagram](image)

**Figure 3: The growing role of language and society in our knowledge acquisition methodology**

The use of our interdisciplinary approach appears to suggest a changing role for the knowledge engineer, which we discuss in detail. First, it appears that he/she can act more gainfully as a group facilitator, learning how to best use the group dynamic present when domain practitioners work together within the DIG. Second, one can argue that the knowledge engineer’s ability to communicate with the domain experts can be improved if he/she learns about the structure and usage of language constructs used by the experts, the difference between general and specialist language and the role of terminology and lexical semantics. This alleviates the personal bias in a knowledge engineer’s interpretation: the fact that if the same knowledge acquisition task is given to a different knowledge engineer the results of the interpretation can be, in some cases, substantially at variance from each other.
This is the problem known as reductive bias, a primary form of which is caused by the expert "translating" his domain terminology in a misguided attempt to help the knowledge engineer.

We introduce new techniques and adapt existing techniques to support the knowledge engineer's transition and we attempt to show how his/her new role and the use of these new techniques improve the development of 'real-world' expert systems. We believe that the most important result of the research is the attempt to incorporate all of these techniques into one methodology, illustrated in Figure 4.

The methodology involves the use of brainstorming and consensus decision-making from a very early stage. We do not intend to give the impression, however, that the interviewing part of the process is a trivial task. Although much has been written on the advantages and disadvantages of many different interviewing techniques, our work here has focused on the roles of knowledge acquisition 'in context' and consensus decision making, describing an approach for structuring interviews to avoid reductive bias through the understanding of domain-specific language.

"Human beings acquire knowledge of and competence in a specific language via a complex process of socialisation" (Collins Dictionary of Sociology). The knowledge engineer does not have the time or the effort available, even if he/she wanted to, to undertake this complex process of socialisation, which would imply that there is no way he/she could quickly learn to communicate with the expert members of the domain. Like Wood and Ford (1993), our goal is to provide guidelines for making interviews with an expert as productive as possible, and like them our conception of what constitutes interviewing is reasonably broad.
Figure 4: Towards a new methodology for knowledge acquisition?
After the interview has taken place and the transcript has been reviewed and annotated by the experts and DIG, the knowledge engineer can once again turn to sophisticated text analysis tools to help him/her extract domain objects, properties, tasks and rules to produce the mediating representation of the paper knowledge base. Note here that the knowledge is elicited independently of concerns for how it will be implemented in a final, working system. As Wood and Ford (1993) pointed out, this leaves the knowledge engineer free to make a decision about representation and implementation issues without being prematurely constrained in his/her knowledge acquisition efforts.

After more brainstorming sessions with the DIG to refine this paper knowledge base, the knowledge engineer implements the rules in a series of prototypes. The DIG walk through these implementations and are given the chance, through further brainstorming sessions, to once again refine the knowledge before the system goes into user appraisal tests.

We believe that such a methodology, which can be grounded by systematic reference in the societal and linguistic metaphors for knowledge acquisition, may be a more accurate, more efficient, more affordable and therefore more generally acceptable approach to knowledge acquisition than the conventional approach. However, such improvements are difficult to assess: almost any view would be interpretative and possibly biased. Ford et al. (1993) point out that the “crucial question for knowledge engineers is not ‘How do we know the model is correct?’, since every model is, to some extent, an oversimplification, but alternatively, ‘How useful is the model (and the modelling process) as a means of facilitating our understanding of the domain?’” In Chapter 5, we have attempted to show that our model (the paper knowledge base) is accurate and complete, and therefore our modelling process (methodology) is
effective and efficient, by presenting results of questionnaires completed by the DIG, turning part of our methodology on to our own work: the need for consensus and domain wide approval.

Finally, in Chapter 6, we conclude whether or not we have succeeded in providing the effective and efficient methodology which was missing from the knowledge acquisition repertoire when we started our work several years ago. We include a discussion on how, based on our evaluation, our techniques could be further adapted and enhanced for the next generation of knowledge acquisition systems.
2. RECENT TRENDS IN KNOWLEDGE ACQUISITION

2.1 A CHANGE IN APPROACH

Ford and Bradshaw (1993) state that one of the most active and controversial areas of knowledge acquisition research concerns multiple expert analysis and Bradshaw et al. (1993) report “future knowledge acquisition systems can neither assume a single source of expertise nor a closed world.” We open this review in section 2.2 by considering various descriptions of experts, expertise, and the importance of multiple knowledge sources.

Over recent years many researchers including ourselves, have leaned towards the view that knowledge acquisition is a constructive modelling process, not simply a matter of ‘expertise transfer’ (see, for example, Ford et al., 1993; Bradshaw et al., 1993; Boose et al., 1993). The knowledge engineer and the expert are involved in the development of a model based on the expertise of the latter, the model representing “a structured understanding of the entities and the processes that contribute to the solution of a real-world task” (Torra and Cortés, 1995).

Ford et al. (1993) suggest that “to be consistent with this perspective we should advocate practices and tools that facilitate active collaboration between expert and knowledge engineer, that support a serviceable theory in their application, and that support knowledge based system development from a life-cycle perspective.” The agenda for the knowledge acquisition research community includes developing tools and methods to aid experts in their efforts to express, elaborate and improve their models of the domain, as described in Section 2.3.
We believe that interviewing domain experts is still a very important part of the knowledge elicitation process and machine-based automatic knowledge acquisition is still very much in its infancy. To this end, knowledge engineers have developed some interviewing techniques well used by psychologists and sociologists. Wood and Ford produced an extensive report on their research into interviewing techniques for Ford and Bradshaw (1993). We summarise their findings in Section 2.4 below along with our own work on interview techniques reported in Ahmad and Griffin (1991), Griffin and Ahmad (1993) and Griffin (1995).

There is very little discussion on the role of brainstorming in the knowledge acquisition literature, particularly on how a knowledge engineer should facilitate such a session. The most recent and comprehensive report is by Boose and Bradshaw (1987) whose findings we summarise in section 2.5.

We will hear from several researchers in section 2.6 how a good mediating representation fosters the constructive modelling process by empowering domain experts and knowledge engineers to co-operatively build models of expert knowledge. Furthermore, we discuss how mediating representations may facilitate explanation by enabling the system’s eventual users to explore the conceptual domain model without resorting to the low-level representations of the shell or programming language (see for example Bradshaw et al., 1993).

As pointed out by Ahmad (1993), there is a growing awareness of language related issues in the knowledge acquisition literature, however these show little understanding of terminology and lexical semantics which provide the powerful text analysis engine used in
our own methodology. In Section 2.7 we look briefly at this growing trend and more specifically at current text analysis modules in tools reported to be capable of knowledge acquisition from text.

In Section 2.8 we review the use of expert system ‘gangs’ in classic expert system projects of the past such as MYCIN and DENDRAL (Buchanan and Shortliffe, 1985) in an attempt to find early existence of a societal metaphor in knowledge acquisition.

Finally in this chapter, in Section 2.9, we briefly summarise the current state of the art in automated knowledge acquisition tools.

2.2 Experts and Expertise

Ford et al. (1993) state that “expertise is more than a mastery of some set of widely shared consensual beliefs of the kind found in textbooks”. They claim that “the most significant aspects of [an] expert’s socially situated knowledge and skills are those of their own making, constructed out of personal experience with their social constituency.”

This implies that “the expertise does not reside in the expert per se but in the expert-in-context.” This view is fortified by Bradshaw et al. (1993) when they state that much knowledge required to build an expert system “is social and/or political in nature and all of it is context dependent.” Indeed Ford et al. (1993) conclude that we have lost an individual reference for expertise and that “the unit of analysis is an interaction between a constituency and the selected expert.”
Wood and Ford (1993) attempt to distinguish expertise from knowledge in that “experts have developed a socially situated expertise-in-context that in some important respects does not coincide with publicly available domain knowledge.” Gaines and Shaw (1994) provide a description from work by sociologists of science that have “characterised scientific communities as forming invisible colleges which monitor and manage the changing structure of knowledge in their domain (Crane, 1972).” They point out that individuals playing major roles in these communities are considered experts not only in overt knowledge of the domain, but also through skills in its management and development.

Very little has been written recently on the selection of experts for knowledge acquisition, but Burton et al. (1987) presented data relating experts’ individual characteristics (personality and cognitive style) to the efficiency of different knowledge acquisition techniques. They concluded that “introverts take longer on the interview than do extroverts.” This is perhaps not very surprising, but it is surprising that these introverts eventually “produce more rules and clauses than the extroverts in order to convey the same amount of information.” This leads to a difficult decision for the real world expert system developer: quicker interviews save money but at what cost to the knowledge content? Although not well documented, such issues must be considered when choosing experts as Burton et al. found that “personality accounts for over 50% of the variance in elicitation time for the interview technique.”

The problems associated with the selection of experts is compounded when it is felt necessary to use multiple experts. In our experience this is often the case. Bradshaw et al. (1993) point out that one of the most active areas of knowledge acquisition research concerns multiple
expert analysis. They add that “theoretical as well as practical concerns also make this one of the more controversial areas.”

They observe that in some cases, co-mingling the domain models of multiple experts tends to cause a “regression to the mean”, and that the resulting system is “less expert” than either individual (Ford and Adams-Webber, 1992). They have also found that it is difficult to add another expert to an “emerging and typically idiosyncratic domain model” acquired from a particular expert, and indeed they recommend that it is “usually preferable to build a separate knowledge base for each expert, rather than attempting to mingle their expertise in a single unified knowledge base.” However, so as not to completely negate our work, they do qualify this statement by saying that “this is not to say that there are no circumstances which warrant the use of multiple domain experts”.

Schuler et al. (1990) go much further, pointing out that “If there is agreement between the consensus opinions and the expectation of the users, users gain more confidence that these are valid recommendations.” Unfortunately they do not say how to achieve this, however we believe we have found a way, as reported in Chapter 4.

Schuler et al. also point out that an interesting consideration arises when using computers to support collaborative work. They note that participants may have differences in vocabulary or definitions, differences in semantics, differences in concepts or differences in information, disagreeing perhaps on models of process and yet still agreeing on final conclusions. “A consensus view is useful, but it might obscure one or more viable, special-case dissenting views.” The question arises as to how much dissent there should be before a special case can
be flagged? The solution proposed is the use of structured discussion sessions which tend to change weak opinions and expose misinformation.

Finally on this subject, we look at tools that have been developed to support the use of multiple knowledge sources. The MINUS tool (Shaw and Gaines, 1986) compares grids from different experts on the same subject and points out differences and similarities. This information has been used to manage structured negotiation between experts (Boose, 1986). Also there are plans to include SOCIOMBRIDS features (Shaw and Gaines, 1986) in the future to display networks of expertise where nodes and relations show the degree of subsumption of one expert's grid over grids from other experts.

Bradshaw et al. (1993) report on a tool called RepGridNet, which supports integration of repertory grids and socio-analysis tools with an electronic mail subsystem to facilitate the formation and management of 'special interest groups' (Shaw and Gaines, 1991; Shaw and Gaines, 1992). Boose et al. (1993) have developed a comprehensive decision model for group decision support systems as part of their DDUCKS project. This decision model combines current brainstorming-oriented methods, structured text augmentation and repertory grids, all of which we discuss later in this chapter. Boose (1989) also mentions Delphi, AQUINAS, ETS, MEDKAT, KITTEN and KSS0 as tools involved with the acquisition of knowledge from multiple sources. However, he concedes that although these tools can elicit differences, once they are uncovered, group facilitation is necessary to resolve them (Boose et al., 1992b), corroborating our approach as described in Chapter 4.
2.3 A Constructivist Metaphor

Ford et al. (1993) offer a constructivist theory of knowledge as a plausible theoretical foundation for knowledge acquisition and as an effective practical approach to the dynamics of modelling. In this view, human experts construct knowledge from their own personal experiences while interacting with their social constituencies (e.g. supervisors, colleagues, clients, patients) in their niche of expertise. In short, knowledge acquisition is presented as a co-operative exercise emphasising the use of mediating and intermediate representations which we will discuss further in section 2.7.

Two main theories provide the basis for the constructivist metaphor. The most popular approach is based on Kelly’s Personal Construct Theory (Adams-Webber, 1989; Adams-Webber, 1990; Agnew and Brown, 1989a; Agnew and Brown, 1989b; Kelly, 1955; Kelly, 1966; Kelly, 1969; Kelly, 1970; Mancuso and Eimer, 1982; Mischel, 1964) which according to Bradshaw et al. (1993) has provided “both a plausible theoretical foundation and an effective practical approach to knowledge acquisition in a variety of settings.” They report that the application of repertory grid techniques to knowledge acquisition has been enormously successful and that both construct theory and repertory grids are now “so widely known and used that people often equate them.”

As reported by Ford et al. (1993), Kelly’s principle of “constructive alternativism” asserts that “reality” does not reveal itself to us directly, but rather is subject to as many different constructions as we are able to invent, thus any given event is open to a variety of alternative
interpretations. It is important to note, however, that this does not mean that one interpretation is as good as any other.

Bradshaw et al. (1993) report that researchers have developed several forms of grid including implication grids, resistance-to-change grids, bipolar implication grids, dependency grids, exchange grids, and mode grids among others (Adams-Webber, 1984; Fransella and Bannister, 1977; Hinkle, 1965; Shaw, 1981). A variety of grid formats have also been derived in which people, objects, events, situations, or other kinds of elements, are either categorised, rated, or rank-ordered on a set of constructs.

The repertory grid is normally used in knowledge elicitation when an expert finds it easier to provide exemplary cases rather than develop a knowledge structure directly. Gaines and Shaw (1993) describe the technique as eliciting the significant distinctions between cases, all the time feeding back matches between cases to elicit new distinctions, and matches between distinctions to elicit new cases. They suggest that the “resultant grid is clustered to feed back to the expert the overall conceptual structure for validation, and rules may be induced from the comparatively small dataset which are usually meaningful because the feedback has eliminated spurious correlations.” It is this grid and these rules that are then exported as a knowledge base covering the specific domain characterized by the cases.

There are a large number of tools incorporating repertory grids for knowledge acquisition including: Aquinas (Boose and Bradshaw, 1987; Boose et al., 1989), DART (Boose et al., 1992a), ETS (Boose, 1984, 1986), FMS Aid (Garg-Janardan and Salvendy, 1987), KITTEN (Shaw and Gaines, 1987), KRITON (Diederich et al., 1987a; Diederich et al., 1987b), KSS0
(Gaines, 1988), Nicod (Ford et al., 1991), PCS (Chang, 1985; Shaw and Chang, 1986), and PLANET (Gaines and Shaw, 1981; Gaines and Shaw, 1986; Shaw, 1979).

It is worth noting here, however, that amidst all of the recommendations, Bradshaw et al. (1993) also report that Personal Construct Psychology methods provide no guarantee that a sufficient set of knowledge will be found to solve a given problem. Indeed Aquinas, for example, attempts to expand the initial subset of solutions and traits with problem-solving knowledge for specific cases.

The second constructivist approach is concept mapping which, like repertory grids, is used to structure the domain. The concept map is the major methodological tool for ascertaining “what is already known” in Ausubel’s Assimilation Theory (Ausubel, 1963; Ausubel et al., 1978). Like Kelly’s Personal Construct Theory, it is based on a constructivist model of human cognitive processes, describing how concepts are acquired and organised within a learner’s cognitive structure.

Concept maps are of increasing interest to those engaged in the process of knowledge acquisition for the construction of knowledge-based systems. Bradshaw et al. (1993) describe them as a set of concepts in a hierarchical framework where inclusive concepts are found at the highest levels, with progressively more specific and less inclusive concepts arranged below them. In this way, concept maps display Ausubel’s notion of subsumption, namely that new information is often relative to and subsumable under “more inclusive concepts” and concepts at any given level in the hierarchy tend to have a similar degree of generality.
Propositions, represented by the relationship between concepts, form semantic units; thus a
concept acquires additional meaning as more propositions include it. Bradshaw et al. (1993)
report that "much of the expressive power of concept maps comes from the fact that the user
is free to employ an unlimited set of linking words to show how meanings have been
developed. When concepts and linking words are carefully chosen, these maps are powerful
tools for representing and communicating nuances of meaning." We can draw many
similarities between their use as a mediating representation and our use of natural language in
the same role. We describe mediating representations further in section 2.7.

Concept maps offer a flexible framework for eliciting, representing and communicating the
emerging domain model. In this way, "they are well suited to the view of knowledge
acquisition as a constructive modelling process in which the knowledge engineer and domain
expert collaboratively build a domain model" (Bradshaw et al., 1993).

2.4 Interviewing

We believe that interviewing domain experts is still a very important part of the knowledge
elicitation process became machine-based automatic knowledge acquisition is still very much
in its infancy. To this end, knowledge engineers have developed some interviewing
techniques well used by psychologists and sociologists. The style and application of each
method helps to obtain as many facets of the knowledge of a domain as possible.
Wood and Ford (1993) produced an extensive report on their research into interviewing
techniques for Ford and Bradshaw (1993). We summarise their findings below along with our
own reported in Ahmad and Griffin (1991), Griffin and Ahmad (1993) and Griffin (1995).
Both typologies are based on relevant research and techniques from the social sciences, the nature of expertise, a desire to assist a knowledge engineer to avoid reductive bias (one of the major pitfalls associated with the acquisition of highly complex concepts), and the desirability of de-coupling elicitation from implementation. Wood and Ford suggest a four phase approach: descriptive elicitation, structured expansion, scripting and validation.

2.4.1 Descriptive elicitation

At this stage the knowledge engineer attends to the categories, objects, models, and other conceptualisations used by the expert in describing their domain. Because much of this information is reflected in the expert's use of domain-specific terminology, Wood and Ford suggest that a significant amount of the knowledge engineer's time should be spent documenting the expert's use of language. The most important considerations during this early phase of knowledge elicitation are to record verbatim what the expert says and to ensure that the terminology being used is "native" to the expert's community. It is important to remember that a word or phrase need not be obviously unfamiliar to the knowledge engineer in order to be important.

It is for this purpose that we propose informal or overview interviews (Kidd and Cooper, 1985) aimed at familiarising the knowledge engineer with the domain and the particular problem which the proposed expert system is intended to solve. It requires much preparation by the knowledge engineer in collating and learning the relevant technical terminology, however we recommend that this be performed through text analysis, as described in Chapter 4.
2.4.2 Structured Expansion

The knowledge engineer uses questioning techniques that explore the rich, integrated organisational structure of the expert’s knowledge at this stage. The most important technique here is to phrase questions to the expert in domain-specific terminology. A knowledge engineer’s consistent use of terminology different from that of the expert will encourage the expert to translate. Although it is difficult to tell exactly when an expert is translating, Wood and Ford suggest that there are some cues. The use of analogies to everyday objects and situations, formal explanations from texts, or other explanations that the expert takes time to construct, are not likely to represent the expert’s customary modes of thought. Using the domain-specific language, however, provides a context of familiarity and encourages the expert to focus on the domain.

We advocate here the role of focussed interviews (Bradshaw, 1988) which are similar to ordinary “chat show” conversations or discussions where the interviewer is interested in a topic of which the interviewee is knowledgeable. They are normally conducted by following a pre-determined agenda. The interviewee will be familiar with the agenda well before the interview to allow preparation of diagrams, graphs, sketches, and the like. The interviewee is initially prompted with the first topic or question, but is given a great deal of freedom of expression thereafter. To ensure that the current terminology is used, we advocate the use of a domain practitioner, not the knowledge engineer, as interviewer (see Chapter 4).
2.4.3 Scripting

Once the content and structure of a large portion of the expert’s declarative knowledge have been elicited, this information can be used to guide the search for procedural expertise. Wood and Ford make a distinction between two different types of procedural knowledge: the reasoning or mental manipulation of information an expert goes through to reach conclusions about the problem at hand, and the strategies and sources used by the expert to obtain the relevant information.

We have reported how think aloud protocols are the most commonly used technique at this stage. This technique has its origins in cognitive psychology where it was used by psychologists to study the strategies with which people solve problems. Basically, it requires that the expert ‘thinks aloud’ while solving a given problem or case study, the latter being advantageous because the end results are already known. The expert is much more relaxed doing this sort of task: he/she will have been asked to prepare a presentation and will be very familiar with one of his/her “pet” problems. This familiarity will often lead to conjecture about possible exceptions or alternatives that could have occurred at points in the solution; these can be recorded and followed up in subsequent interviews.

The utility of using declarative knowledge to interpret verbal protocols is the reason Wood and Ford recommend obtaining extensive declarative knowledge before attempting to elicit procedural knowledge. Once the important concepts and objects in the domain are understood by the knowledge engineer, a think-aloud protocol is much easier to interpret.
2.4.4 Validation

Wood and Ford recommend that aided-recall questions are used to follow up results obtained in think-aloud protocols. This involves showing video- or audio-tape of the interview to the expert at a later date, freeing the expert from what they call the ‘interference’ resulting from think-aloud protocols.

We have reported that it is essential to record and transcribe all interviews and that transcripts should be clearly cross-referenced to recorder times and interspersed with sketches, photocopies or reproductions of diagrams, tables and the like, that were referred to during the session. Once completed, a copy should be sent to the interviewee and preferably another member of the domain for comments, corrections and criticism (see Chapter 4). Involving the expert in validating his/her own transcript reduces the chance of erroneous information appearing in the prototype’s knowledge base.

We finally recommend the use of structured interviews (Bradshaw, 1988) well into the knowledge acquisition process. They are used when information is required in much greater depth and detail than the other techniques can offer and are more interrogative than conversational. The principle outcome of the structured interview is detail of the domain entities (tasks, rules and objects) to such a level that a decision could be made about the representation schema or data structures required to implement them in an expert system.
2.4.5 Performance

We believe it is important here to record some findings on the performance of these techniques. Burton et al.'s (1987) study on knowledge acquisition highlighted a poor showing for protocol analysis. They found that not only does protocol analysis take longer to perform and analyse than comparable techniques, it also seems to retrieve a substantially smaller amount of the necessary information than other techniques.

We also reported problems with protocol analysis in Griffin (1995). Its use elicited many specific issues but revealed very few generic tasks. During prototyping, it was clear that this approach led to the expert system being very brittle: if a similar case was input to one of the cases used during the interview then the system performed extremely well, but if there was any significant difference in the circumstances of the case, the system could not perform at all (see Figure 5).

![Figure 5: Comparison of Think-aloud protocols and focused interviews for brittleness](image-url)
However, also in Griffin (1995), we noted that the opposite problem can occur through extensive reliance on focused interviews. We achieved our aim to remove the brittleness from the expert system, but unfortunately at a cost: the knowledge elicited from the interviews led to the system being able to support almost every case, however it was not specialised enough to fully determine a single one (see Figure 5).

Ideally, an expert system needs to possess the lack of brittleness and the expertise in specialist areas, but this problem of brittleness and performance appears to be dependent on the interviewing techniques employed. We hope that the following sections, and particular Chapter 4, may help to provide the in-between state of broad coverage and high performance missing in Figure 6.

![Diagram](image)

**Figure 6:** Types of psychological interview techniques leading to different types of expert system
2.4.6 Tools

A final note on interviewing techniques is to mention the vast number of tools now available to support this process. Boose (1989) reports that AQUINAS, ASK, CAP, ETS, KITTEN, KNACK, KRIMB, KRITON, KSS0, MDIS, MOLE, MORE, ODYSSEUS, PLANE, PROTOS, ROGET, SALT, TEIRESIAS and TKAW all interview experts directly and KRITON, LAPS and MEDKAT record and analyze transcripts from experts thinking aloud during protocol analysis.

2.5 Brainstorming

Results of research presented by Dickson Lukose at the ICCS’94 Workshop on Knowledge Acquisition using Conceptual Graph Theory, showed that out of 31 research groups studied, only two groups had reported the use of brainstorming as a knowledge acquisition / elicitation technique (McGraw K. and Seale M.R., Osborn A.) as compared to 13 groups reporting on the use of interview techniques.

The most recent and comprehensive report on the roles of brainstorming is found in Boose et al. (1993) who discuss the role of brainstorming within the context of their work on a decision support model. They report that the processes of generating and scoring alternatives are at the heart of most decision problems and that a fast and simple way to reach a group decision is to brainstorm alternatives and score them. Typical processes, methods, and knowledge roles noted for brainstorm-and-score systems are summarised in Table 1, many of which we use ourselves. They conclude that making criteria explicit sometimes helps the group reach a
better decision: explicit criteria allow alternatives to be examined in detail, and can also illuminate areas of agreement and disagreement.

Another reference reporting on the use of brainstorming is Torra and Cortés (1995) who use brainstorming to elicit “qualitative attributes”. This entails “voting” on the preferred choice of label for objects from a selection of linguistic terms, again something we too use brainstorming for, as reported in Chapter 4.

<table>
<thead>
<tr>
<th>Process / Method / Role</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generate requirements</td>
<td>Define the needs that help identify the problem.</td>
</tr>
<tr>
<td>Define problem</td>
<td>Define the problem and a process for solving it.</td>
</tr>
<tr>
<td>Organize problems</td>
<td>Use outlining to group problems in categories.</td>
</tr>
<tr>
<td>Define participants</td>
<td>Define the members and their roles (such as participant or facilitator) in the problem-solving process with rosters or group lists.</td>
</tr>
<tr>
<td>Organize participants</td>
<td>Categorize participants by session or problem type.</td>
</tr>
<tr>
<td>Generate alternatives</td>
<td>Identify potential solutions for a problem.</td>
</tr>
<tr>
<td>Organize alternatives</td>
<td>Organize alternatives in categories with an outlining or grouping tool.</td>
</tr>
<tr>
<td>Weight (score) alternatives</td>
<td>Score solutions for a problem individually or by group. Common methods include placing items in order (ranking), assigning rating scores, voting (yes / no / abstain), selecting several alternatives from a list, and allocating fixed resources among alternatives.</td>
</tr>
</tbody>
</table>

Table 1: Processes, methods and knowledge roles for brainstorming and score systems

2.6 MEDIATING REPRESENTATIONS

The term ‘mediating representation’ has various interpretations in the literature, however the most commonly referenced is to “convey the sense of ... coming to understand through the
representation” (Johnson, 1989). As Ford et al. (1993) point out, a model based description of the domain in a form that the user can intuitively understand has many advantages, the chief of which is that it can serve to mediate communications between developers, experts and users of the system, helping all of them to understand and articulate the broader, higher level problem context.

Research on mediating representations has generally attempted either to improve the computational expressiveness of human-efficient representations such as repertory grids, concept maps and hyper-text, or to improve the learnability of computationally powerful ones, like fourth-generation languages for example.

Wood and Ford (1993) believe that knowledge should be elicited, organised, and documented with minimal concern about constraints and details associated with implementing a working system in which it will be used. They advocate the documentation of knowledge using a mediating representation that can be used in knowledge acquisition without making commitments regarding system implementation tools, an ideal illustrated by Bradshaw and Boose (1993) and reproduced in Figure 7.

These efforts were made primarily to support the early stage of conceptual modelling which has always suffered from the problems of premature encoding of knowledge implementation-driven representations (Ford et al., 1993). Gaines and Shaw (1994) report, however, that the application of mediating representations, “whilst relevant to the understanding, management and presentation of the project does not in itself result in artificial intelligence in the sense of an inferential computational system. The knowledge acquired is being fed back to facilitate
the intelligence of people rather than being used by a computer to provide direct support of knowledge processes through computational intelligence.”

![Diagram]

**Figure 7: Mediating representations facilitate communication between domain expert and knowledge engineer (from Bradshaw and Boose, 1993)**

On the other hand, Ford *et al.* (1993) argue that “considerations of human efficiency far outweigh considerations for complex modelling problems” when it comes to the use of mediating representations. They state that a “good mediating representation can facilitate modelling processes by providing a medium for experts to model their valuable, but difficult to articulate, knowledge in terms of an explicit external form.” There is common agreement that the mutual development of an external cognitive artefact supplementing the exchange of information between participants promotes and enriches communication, leads gradually to a shared understanding of the emerging conceptual model of the domain (Norman, 1988, 1991) and therefore enables domain experts and knowledge engineers to co-operatively build problem-solving models (Ford *et al.*, 1993; Bradshaw *et al.*, 1993).

These researchers also noted from Johnson (1989) that “where many are gathered together to build an expert system, the team will need access to a statement of the problem and, as the
project progresses, to the emergent knowledge not yet expressed (or expressible) in the
concepts of the final implementation language”. This they see as another ideal use of the
mediating representation, and indeed we make our mediating representations available
throughout the project and include them in the implemented system.

This agrees with yet another advantage of such representations reported in the literature, in
that they may also facilitate maintenance and explanation by enabling both knowledge
engineers and the system’s eventual users to explore the conceptual domain model without
resorting to low-level representations in the later stages of system development (Bradshaw et
al., 1993; Ford et al., 1993).

It is widely considered that one of the most important questions in the use of mediating
representations is how they should be designed and evaluated. Criteria, derived from Johnson
(1989) and Winston (1984), and reproduced by Bradshaw et al. (1993), suggest the following
rules-of-thumb for evaluating the effectiveness of a representation:

- Is the formalism sufficiently expressive?
- Does the formalism aid communication between the members of the development
team?
- Does the formalism actually guide knowledge analysis in a significant way?
- Does it make the important things explicit, suppressing detail and keeping rarely
  used information out of sight, but still available when necessary?
- Does it expose natural constraints?
- Is it complete and concise, efficiently saying all that needs to be said?
Boose et al. (1993) believe that effective mediating representations are critical to the success of knowledge acquisition tools, being “the users’ window on the decision model.” Many knowledge acquisition tools achieve success in part through the development and adaptation of good mediating representations. These tools tend to adopt one of two approaches. Either they contain interfaces that bear a close resemblance in appearance and procedure to the original manual task, for example, cancer-therapy protocol forms in OPAL (Musen, 1988) and our own standard NRA application forms in ELSIE (Ahmad and Griffin, 1994), or they rely on some easily-learned, generic knowledge representation form, for example, object hierarchies and repertory grids in DART (Boose and Bradshaw, 1987; Boose et al., 1989).

ICONKAT’s principal mediating representations are the concept map and the repertory grid (Bradshaw et al., 1993). It uses these complementary mediating representations synergistically. Concept maps depict the conceptual relationships of the domain as constructed during the knowledge acquisition process. In addition to concept maps and repertory grids, ICONKAT supports the use of a variety of other mediating representations such as images, audio, Quicktime movies and documents.

The issue of mapping between representations is a troublesome one. It is obvious that much of what can be modelled in mediating representations can not be directly incorporated in to the current commercial performance systems, therefore there is the need for integrative, intermediate representations that are relatively independent of the constraints of the delivery environment (Ford et al., 1993). Bradshaw et al. (1993) agree that although mediating representations have enhanced the richness and subtlety with which the human participants in the knowledge acquisition process can model the domain, the need for intermediate
representations has become increasingly apparent. For example, the kind of representations that ICONKAT provides as the basis of its modelling environment are designed for the benefit of humans, while implementation formalisms are focused on computational issues, causing a substantial semantic gap. ICONKAT's intermediate representations are hence sometimes referred to as the 'glue' (Bradshaw et al., 1993), an approach illustrated in Bradshaw and Boose (1993) shown in Figure 8.

![Figure 8: Three-schema architecture for knowledge acquisition tools (from Bradshaw and Boose, 1993)](image_url)

Finally in this section we reproduce in Figure 9 a typology of mediating representations from Boose et al. (1993). Examples of mediating representations which we use in our research are italicised and will be discussed in later chapters.
As pointed out by Ahmad (1993), there is a growing awareness of language related issues in the knowledge acquisition literature, many researchers stating the need for glossaries or lexica and the importance of terminology. As we see in Figure 9 above, Boose et al. (1993) sites 'free-text' at the highest level of mediating representation and glossaries and lexica further down his classification, supposedly being “simple structure collections".
This literature, however, shows a poor understanding of terminology and lexical semantics which we consider vitally important in the attempt to acquire knowledge from text. Lenat and Guha (1990), describing their 'common-sense' knowledge base CYC, note that a rule comprises a 'set of terms', but the meaning of these terms remains 'in the eye of the beholder', totally dismissing the importance of terminology in regard to the role of terms, and lexical semantics in regard to the relationships between the terms.

Boose (1989) lists several tools such as KRITON that can generate knowledge directly by analysing text, whereas Diederich et al. (1987b) only imply the role of terminology in KRITON. They do, however, report on the use of statistically-oriented content analysis techniques to search for domain terms within text, as discussed further in Section 3.4.

Lin et al. (1991) developed a free-text processing system called CAPIS to extract Physical Examination (PE) findings from dictated admission summaries, using a concept based matching algorithm on canonical phrases, or sets of words that comprise a medical concept. However, they assume that researchers studying a particular disease already know the terminology of their domain, and therefore CAPIS only tries to identify the target findings that are pre-specified by the user. Furthermore, the PE sections have special properties that make them easier to identify canonical concepts in free-text, for example, the subject of the description is always implicitly the patient, descriptions typically refer only to the present, and adjectives or nouns used in descriptions for one section are not the implied subjects or objects of sentences in other sections. This structure in the PE facilitates the search for canonical phrases which would not be possible in natural language.
Gaines and Shaw (1994) used text analysis tools to analyse the conceptual structures of the Intelligent Manufacturing Systems research programme objectives in GNOSIS. They produced a concept map, generated automatically through analysis of the co-occurrence of words in sentences, a technique commonly used in information retrieval systems (Callon et al. 1986), and again discussed further in Section 3.4.

Yost and Newell (1989) plan to build a tool that helps a knowledge engineer carry out the operationalisation processes described in our introduction. The knowledge engineer will bring natural English descriptions of task knowledge to the tool which will help select and apply appropriate instances of the identification, representation and communication processes required, ultimately producing an implementation of the task. Interestingly, they do not believe that it will be possible in the near future to fully automate operationalisation in a general-purpose expert system development tool, and therefore will not attempt this. They believe that the language skills required are well beyond the state of the art and their tool will leave the language skills with a human knowledge engineer, who can perform them routinely.

2.8 **EXPERT SYSTEM ‘GANGS’: SOCIETAL INPUT FROM THE DOMAIN COMMUNITY?**

Probably the most famous expert system ‘gang’ referred to in the literature is the MYCIN gang (Buchanan and Shortliffe, 1985), however, we begin our search for domain interface groups of the past with the DENDRAL project, simply because it came first. We have produced a hierarchy of domain experts who took part in the DENDRAL experiments (see Figure 10). This hierarchy is based on the authorship of papers in the journal “Applications of
Artificial Intelligence for Chemical Inference” between 1969 and 1976. The higher rank of an expert shows that he/she published more papers on the subject.

Figure 10: Hierarchy of Domain Experts used in the DENDRAL Experiments

The figure shows that at least 14 domain experts were involved with the experiments. Although the DENDRAL literature does not explicitly discuss the importance of such a broad input from the domain community, we consider this to be the first signs of the need for societal input in an expert systems project. Of course at this period in time, expert systems were in their infancy and such a broad input from the application domain may have been due to such system’s novelty value. Were members of the application domain queuing up to take part in such ground-breaking research, which at the time promised so much? Or did the knowledge engineers, led by Buchanan, purposely search for such a broad domain coverage?
Whatever the cause, it would appear that the effect was the presence of societal input as far back as the early 1970’s.

Much has been written about the MYCIN gang, but again there has been no argument as to the benefits of such a broad representation of the domain community being involved: if anything, the literature concentrates more on the broad input from the domain of the enabling technology than the application. Indeed over 15 computer scientists can be identified as working on MYCIN or related expert systems (such as PUFF, GUIDON and ONCOCIN) to a greater or lesser extent.

However, even more members of the MYCIN gang came from professions within the application domain including physicians and research staff specialising in infectious diseases, pharmacology, oncology, haematology, pulmonary diseases and medicine in general. Figure 11 shows our tentative ranking of domain experts who took part in the MYCIN experiments. This ranking is based on a short discussion with Bruce Buchanan and attempts to show the extent of involvement of the individuals within this project.

What we consider to be important, and part of the inspiration for our work here, is that DENDRAL and MYCIN were both very successful systems, particularly the latter which is famous for its out-performance of humans in the diagnosis of several cases of infectious disease. Systems developed since 1980 have not been so successful, but then they did not have such large ‘gangs’ or therefore such enormous domain input. Although one can not prove that this caused their lack of success, since there are many other factors which should be
considered such as the scope of the problem for instance, we believe that the societal aspect is at least worthy of investigation.

Figure 11: Tentative ranking of experts involved in the MYCIN experiments

2.9 AUTOMATED KNOWLEDGE ACQUISITION TOOLS

Although we have not automated our methodology described in this thesis through the coupling of each component, it is something which will perhaps be done in the near future. For this reason, we summarise some of the important issues relating to automated knowledge acquisition tools. Clancey (1993) reported that ethnographic studies suggest computer tools might be based on two considerations. First, enable forms of sharing: make resources more
available, allow people to see the knowledge and incorporate it into their own systems. Second, make tools accessible to everyone in a group: support overlapping responsibilities, do not isolate jobs, accommodate novices and experts alike.

Ford et al. (1993) consider that automated knowledge acquisition tools should be designed to support the active participation by both experts and knowledge engineers in the creation of knowledge bases. More precisely, they should support at least the following four facets: elicitation and model construction, analysis and refinement of the model, maintenance of the knowledge base in the resultant performance system and model elaboration as part of an explanation capability.

This suggests a requirement that automated knowledge acquisition tools accommodate the changes in representations that may accompany successive stages in model construction: from vague mental models to increasingly refined and explicit conceptual models via elicitation and analysis techniques, and eventually, from these highly elaborated models to an operational knowledge base via formalisation and implementation procedures (Ford et al., 1993).

Bradshaw et al. (1993) have produced a typology of automated knowledge acquisition tools based on their evolutionary stage. First they note the era of the single-approach rapid-prototyping tools. These earliest tools (e.g. ETS, Nicod, PLANET) were based on repertory grid interviewing techniques, representations, and analysis tools. They were generally used for rapid-prototyping of classification problems, following which, rules were exported to a file for use by a commercial expert system shell.
The second era is described as *monolithic integration*. Integration became the theme as knowledge acquisition workbenches (e.g. *Aquinas* and KSS0) incorporated several additional tools and representations. They report that export to external shells was de-emphasised in *Aquinas* as internal problem solving capabilities increased and it became difficult to replicate these in traditional shells.

Finally they describe the modern era of *decoupling and interapplication communication*, representing the current “state-of-the-art”. The components of ICONKAT (Ford *et al.*, 1992; Ford *et al.*, 1991b), KSSn/KRS (Gaines, 1991; Gaines and Shaw, 1990) and DDUCKS (Bradshaw *et al.*, 1992a, 1992b, 1992c) are decoupled to allow integration with tools springing from complementary theoretical perspectives (e.g. concept maps, neural networks, influence diagrams, possibility tables, semantic networks) and to exploit emerging operating-system-based interapplication communication protocols between commercial tools such as HyperCard, Excel, Nexpert, databases and internally-developed applications. Also, more general configurability allowing multiple problem-solving approaches is made possible through the use of PROTEGE II and KADS-like architectures (Musen, 1989; Musen *et al.*, 1993; Breuker and Weilinga, 1989).

Knowledge acquisition tools are beginning to target wider applications such as information retrieval, education, personal development and group decision support (see for example Boose *et al.* 1992b; Cañas and Ford, 1992; Gaines, 1989; Shaw and Gaines, 1992).
3. EVOLUTION OF OUR METHODOLOGY

3.1 AI RESEARCH AT THE UNIVERSITY OF SURREY

Research into Artificial Intelligence (AI) at the University of Surrey began in the early 1980's, with a Natural Language Processing (NLP) system called Loquacious. Around this time, work was beginning on rule-based expert system development shells, based on the success stories of the Stanford Heuristic Programming Project's MYCIN experiments. This work led to PROSE and by 1986 PROSE 2, a shell in which two rule-based expert systems were developed in 1986: WIFE, a forward chaining expert system to guide the user of a complex sewer network modelling simulation software package, and SBCCON.

These two systems were born at the start of the UK Alvey Directorate to promote expert systems technology within mainstream UK applications. Under a grant from Alvey and several large water industry partners, the University's Knowledge Based Systems Group, developed a new expert system development environment which supported frames as well as forward and backward reasoning processes: WIESSE, the Water Industry Expert System Support Environment. This environment was used to develop WADNES, the Water And Drainage Network Expert System, the output of the Alvey sponsored project, and later TESSA, the Tolerance Expert System for Stress Assessment. Another output of the Alvey work was SERPES, the Sewer Rehabilitation and Planning Expert System, which was based on the successful WIFE system and attempted to model the whole procedure for rehabilitating sewer networks (known as the Wallingford Procedure).
History of System Development related to the ELSIE experiments

Figure 12: University of Surrey AI Group systems
After the success of WIESSE in the development of WADNES and TESSA, the environment was generalised and became MARVIN (Holmes-Higgin, 1989). MARVIN was what is now called object oriented, all frames and rules being objects, even their prototypes, the entire system being based on a single boot-strapped concept of a generic object. MARVIN was greeted enthusiastically by the development team at the University, who produced three expert systems within the shell: PLAIM, a diagnosis system for problems on oil-rig platforms (Ahmad et al. 1989), GINAS-IFE, another intelligent front-end, and DIMES, the Diagnosis in Mammography Expert System, the first expert system worked on by the author (Griffin, 1990).

At the same time as MARVIN was being developed, another part of the Knowledge Based Systems Group had returned their attention to natural language processing. The KITES project was an early study into the use of terminology within AI. This led to the groups involvement in the very successful ESPRIT sponsored project TWB, Translator’s Work Bench (Ahmad et al., 1990, 1993), which involved the identification, collation and elaboration of domain specific terminology for future use by translators. Since this new direction was not encapsulated within the groups title, the name of the group changed around this time into the Artificial Intelligence Group, or AI Group as it is called today.

Immediately, the knowledge engineers began examining other uses for the text analysis tool which initially identified these terms. They found that the more sophisticated tools, now available through System Quirk (Ahmad and Holmes-Higgin, 1995; Ahmad et al., 1995), a much enhanced version of the terminology management system developed in TWB could identify much more than domain specific terms, but also what we call propositions which
could be adapted into domain objects, attributes and values, as well as rules for reasoning over. A project was needed in which the possibilities being identified could be tried, tested documented for further use: that project was ELSIE, (Ahmad and Griffin, 1994) the knowledge acquisition methodology for which is described in Chapter 4.

3.2 **OUR CASE STUDIES**

As far back as 1990, breast cancer accounted for 15,000 deaths per year among women, with 24,000 new cases introduced annually in the UK alone. The need for an expert system here was due to two main problems. First, there was an enormous lack of expertise outside of the few centres of excellence. Second, mammography reporting had been shown to be highly variable amongst radiologists which affected the confidence in a national screening program. A prototype expert system was developed called DIMES, Diagnosis in Mammography Expert System.

DIMES was written in MARVIN, an in-house expert system shell developed in prolog (Holmes-Higgin, 1989). (It was a hybrid system making use of frames, production rules and tasks, meta level knowledge structures to partition the rule base). The system was entirely forward chaining because it was impossible to describe the ultimate diagnosis without first collecting large amounts of data.

In 1990 we started work on the Water Resources Management Intelligent Assistant (W-RAISA), an expert system to support regional licensing offices in the regulation of water resources: the handling of applications for abstraction and impoundment licences and the
assessing of the impact on the environment (Ahmad and Griffin, 1991, 1993). W-RAISA was implemented in a proprietary expert system shell and uses various knowledge representation schema such as complex objects, including slots, facets and demons, rules, tasks (in a fashion) and procedures. W-RAISA contains 99 rules distributed fairly evenly between nine tasks: and 20 ‘class’ objects which contain numerous instances at run-time. This inference is complemented by the use of over 50 procedures and functions.

Starting in 1992, the ELSIE project was the main experimental environment in which we developed our knowledge engineering methodology: it provided the application and motivation for us to experiment with innovative approaches to knowledge acquisition and representation. The ELSIE system is the realisation of this research: knowledge dissemination through a real world intelligent information system.

ELSIE is the Expert Licensing System and Information Environment, developed under a research and development grant from the UK National Rivers Authority (NRA). The NRA has a statutory duty to regulate the water resources of England and Wales through the issuing of licences enabling the holder to abstract water from underground strata or abstract or impound surface water in rivers and lakes.

ELSIE has several components, most of which are interactive with each other, with external data sources and with the user. WALDES (the Water Abstraction Licence Determination Expert System) is the central component of ELSIE, providing the user with guidance on whether or not a licence should be issued. WALDES covers the whole process of licence application, from initial inquiry to the NRA, through the statutory application procedure and
on to the determination itself. Throughout the three month period in which a licence application must by law be processed, WALDES is continuously requesting data, gathering data, making decisions on that data and producing results with explanations to the user.

ELSIE constitutes many computer science paradigms: artificial intelligence, object-orientated programming, human computer interaction, relational database management, hypertext systems, and so on. The development of the system also drew on many research paradigms outside the domain of computer science, particularly terminology, lexicography, group facilitation and consensus decision making.

WALDES required a hybrid expert system approach encompassing object-oriented programming techniques for the large scale repository of domain knowledge and production rules arranged in a large task base. We used a sophisticated proprietary development environment encompassing most of these features, requiring only the in-house development of meta-level knowledge representation, such as the combination of the object and rule paradigms for the implementation of large scale task hierarchies.

The rules in WALDES follow the classical IF antecedent THEN consequent format of production rules, but they belong to only one ruleset, minimising the inference tree required by the reasoning strategies, and also state whether they are to be used in backward or forward chaining.

Objects are used to model all of the entities used in ELSIE, not just in WALDES, and much of the system's behaviour, through demons and attached procedures. There are over 200 objects
in total, arranged in a parent-child hierarchy up to five layers deep in places. Many of these objects will have multiple instances at run-time, greatly increasing the actual number of objects used during any one consultation.

Tasks in WALDES are also defined as objects, slots containing the name of the ruleset to be applied to perform the task and information such as the chaining strategy to be invoked, the scope of the search and so on. The task objects and associated code were developed separately from the main ELSIE knowledge base and have since been used in several other projects within the development environment (for a full description of the task features implemented see Ahmad and Griffin, 1994).

WALDES contains over 200 tasks, all linked in a hierarchical structure. Not all tasks contain rules: some contain procedures and functions to be executed, some simply contain text descriptions of the task which needs to be performed, but which can only be done by the human user. All rulesets are implemented for backward chaining, but to keep the search space minimised, each task has a set of entry conditions to ensure data required by the rules at the end of the inference tree is available. This technique also enables a user to jump into the task hierarchy at any position without fear of a task completing with insufficient and therefore possibly misleading data.

3.3 SOCIOLOGY: A NEW METAPHOR FOR KNOWLEDGE ACQUISITION?

The discussion in Chapter 2 outlined the evolution of knowledge acquisition techniques from a purely psychological approach to one with much more societal input. Conventional
knowledge acquisition always adopted a psychological metaphor: from interviewing techniques and think-aloud protocols through to repertory grids and personal construct theory. We have adopted, over a period of time, a societal metaphor for our methodology, in particular considering such issues as language use, society and group dynamics.

As with most knowledge engineers since the MYCIN project, we believed that psychological techniques were the most effective. However, we found ourselves under heavy criticism from the domain community after following conventional methods in W-RAISA. The W-RAISA project was undertaken for a regional branch of a national organisation: we studied one expert, renowned in his field, for four hours using a think-aloud protocol to establish the tasks undertaken whilst working on eight case studies from the domain. The harvest of rules from the transcript was not particularly high, a problem not uncommon with think-aloud protocol analysis (see, for example, Burton et al., 1987), however we were able to build a small prototype expert system. This system was very well received by the regional sponsors and we were invited to present the system to a group of practitioners representing the national body.

Very few of the national representatives considered the system satisfactory, not because the results produced were wrong, but because the tasks undertaken, the questions asked, the assumptions made, the defaults used, and so on, were not consistent with their approach to the problem. This is what we call the 'local expert' problem, the use of a single regional expert in a wide, distributed domain.

The solution at first appeared straightforward: use an expert from each region and pool the knowledge of several local experts. However, as Franklin Delano Roosevelt said, "There are
as many opinions as there are experts”. Since this is indeed the case, should the system be exhaustive, containing every expert’s view, or be a common denominator of their views, only covering issues on which all experts agree?

The exhaustive approach was too much, the common denominator too little: we needed something in-between, some common ground on which all local experts could make concessions or introduce exemptions for particular regional issues. We needed some kind of consensus between the local experts.

We began to consider the domain community as a whole, to think about what groups of people made up that community, what sub-groups of people existed within it, how they interacted, what they had in common, and so on. It appeared that we were asking questions more suited to sociology than psychology or computer science: a society of people made up of distinct groups, each with their own culture, their own preferred solutions to fairly generic problems. This section attempts to identify such similarities between sociology and knowledge engineering, particularly knowledge acquisition, by considering some sociological definitions.

3.3.1 Society

The Collins Dictionary of Sociology gives two senses for the concept of society. First, it describes it as “the totality of human relationships”, a sense which is far too broad to help with our problem. Second, however, it defines a society as “any self-perpetuating, human grouping occupying a relatively bounded territory, possessing its own more or less distinctive
culture and institutions”. As with definitions in every dictionary, this concept is defined in terms of others: to understand it, we must first consider the two concepts of culture and institution, then replace their occurrence with their definitions to complete the picture.

The most important find whilst looking for information on culture is that “knowledge of a culture is acquired via a complex process which is fundamentally social in origin”. This statement provides a foundation for our societal metaphor for knowledge acquisition. However, also worth mentioning is that culture is defined as “the human creation and use of symbols and artefacts”, including codes of manners, dress, language, rituals, norms of behaviour and systems of belief.

An institution is considered to be “an established order comprising rule-bound and standardised behaviour patterns”. Due to the wide variety of ways in which the term is used, and therefore the ambiguity which arises, the more specific term social institution refers to “arrangements involving large numbers of people whose behaviour is guided by norms and roles”. As with society above, if we follow the leads shown through the relationships between concepts in the dictionary, we find that a norm is a standard or rule, some form of regulating behaviour, and more importantly, a role is “the dynamic aspect of status, where status refers to the position and role to the performance”.

At this point we attempt to map the artefacts of knowledge engineering with the cornerstones of society as defined within sociology. Working back to the root concept from the branches of our definition tree traversed above, we can note the following similarities.
The guiding roles of our institution could be construed as the actions of the domain experts, defining the best behaviour to be followed by all those members of the society below their status. Previous experts in the domain would have created the norms which are currently followed, the roles of the past which, over time, have been proven without doubt and are now found in textbooks. The social institution therefore becomes the domain practitioners themselves, existing in relatively large numbers and practising within the rules laid out by the experts of past and present.

This institution is indeed distinctive: in no other group of people would the rules in which the domain practitioners work hold any meaning, hence we are half way to our definition of society. But what of culture? Language is of prime relevance here. The importance and distinction of what we term “expert speak”, the language of a specific domain as compared with that in general use, is discussed in detail in the following section. The symbols of communication are indeed distinctive for a specialist domain, as we show very clearly later. But this is not the only evidence of a culture within the institution of practitioners. The norms of behaviour have already been explored and one can find several examples of rituals: all practitioners from the domain of ELSIE, for example, meet every two years to discuss issues relating to their work.

The final part of the definition involves a ‘relatively bounded territory’. It is fundamental to an expert system’s development that the application area is bounded to ensure any chance of high performance, therefore we conclude that the idea of a society, as defined in sociology texts, is relevant to our view of the domain community in knowledge engineering. Figure 13 illustrates our mapping of concepts within sociology to the artefacts of knowledge.
engineering. With this assumption made, we are only left to mimic the "complex process which is fundamentally social in origin" via which knowledge is acquired, to achieve our objective.

![Diagram of Knowledge Engineering and Sociology](image)

**Figure 13: Mapping of concepts within sociology to the artefacts of knowledge engineering**

We would like to show that knowledge may be elicited from "group discussion", either at its inception, during a "brainstorming" session for example, or through a simulation of the inception, during an interview for example. For this we must look deeper into aspects of society, beginning with groups and their behaviours.
3.3.2 Group

Society, although apparently the key to our work, is too large (at least within the domains we have been studying) to communicate with directly. However, we learnt from W-RAISA the risk of selecting a single member of the society, a local expert, and hoping that he/she will provide a fair representation. This led us to consider components of the society, not single members yet not the society as a whole: we looked at groups.

The Collins Dictionary of Sociology defines a group as "any collectivity or plurality of individuals (people or things) bounded by informal or formal criteria of membership. A social group exists when members engage in social interactions involving reciprocal roles and integrative ties... Any social group, therefore, will have a specified basis of social interaction, though the nature and extent of this will vary greatly within groups.”

The collectivity of individuals and the bounding of the group share the arguments of society, but what is important here is the social interactions, in particular the reciprocal roles. A point to note here is that experts of a domain community do not fit this definition well. Experts rarely exist in groups at all but tend to be individuals within a community. Also, an expert’s roles, as described above, are usually performed down a “hierarchy” of expertise: although there is communication in two directions, experts are usually telling someone what to do, whether teaching, providing guidance or simply answering a question; they very rarely expect or receive similar information back. Reciprocal roles, and indeed more interaction in general, is more likely to be found among peers.
3.3.3 Peer group

A peer group is “a group of individuals of equal status”. This alone would not rule out a group of experts, being peers to each other, however the Collins Dictionary of Sociology goes on to say that “the term is most generally applied to children and adolescents, who experience a very different influence on their socialisation by interacting in groups of their own age, as compared to the hierarchical family experience.”

We will again attempt to draw a mapping here between the artefacts of knowledge engineering and the concepts of sociology. If we replace the age attribute with experience, we can see that the hierarchical family becomes the tree often seen in an organisational chart, depicting the senior staff or experts at the top (like parents) and the new employees or novices at the bottom (like babies). This analogy would make adolescents somewhere between a third and half way up the tree: the level at which you would find the domain practitioner who is not, by his/her own admission, an expert, but knows what to do on a day-to-day basis.

This means that we can label the practitioners as a peer group, and expect them to experience the different influences associated with the interaction associated with such a group. It is this interaction that we now consider.

3.3.4 Group Dynamics

Knowing how the group is made up is unfortunately not even half of the problem: remember it is the roles between individuals of the group which must make up the complex process via
which the knowledge will be acquired. George Simmel, the German sociologist and philosopher, presented society as a “web of interactions” (Collins Dictionary of Sociology), so we turn our attention to “the processes involved in interaction within social groups”, the definition of the term ‘group dynamics’.

The interest in group dynamics within sociology has focused especially on shifting patterns of tension, conflict, adjustment and cohesion within groups and on styles of leadership. We believe that conflict, adjustment and cohesion all play their part when a group is taking part in a brainstorming session in an attempt to reach a consensus on domain specific issues. Also, we would point out that a tested knowledge base is itself an artefact that reflects all of these shifting patterns through its creation, development, refinement and validation. We further consider leadership to be one of the primary roles of a good domain expert chosen for knowledge acquisition purposes. All of these issues are discussed in Chapter 4 when we consider the formation of our domain interface group and brainstorming within it.

3.3.5 Consensus

Consensus is “the existence within a society, community or group, of a fundamental agreement on basic values”. Some sociologists, such as Parsons, “emphasise the existence of such shared values as the basis of any persisting social order”. Could this mean that consensus, if achievable, would form a basis for the knowledge which will satisfy all members of our domain society?
No expansion is available within the literature studied, therefore this question must be left open for future discussion. The lack of supporting information on consensus and brainstorming within the sociology dictionaries is disappointing, but the evidence accumulated to date on society and groups is too strong to give up on our societal theories. We consider now the role of language, here as part of our sociology discussion, not as a separate metaphor for knowledge acquisition in itself, as described in later sections.

3.3.6 Language

The Collins Dictionary of Sociology defines language in three senses. First, language is "a system of symbolic communication, i.e. of vocal (and written) signs... [which] have a common significance for all members of a linguistic group". Second, language is the "crucial signifying practice in and through which the human subject is constructed and becomes a social being" (W. Mulford, 1983). Third, language is "the most important, but not the only sign system of human society".

An important quote to note here is that "language is the means whereby subjectivity is stabilised and crystallised (including 'knowledge'...). Language also exists as an 'objective' institution independent of any individual user. In common with all aspects of human culture, language can be seen to be historical and subject to change." Such comments on language are as true for our sub-language, the so-called 'expert-speak' or domain specific language, as they are for the natural languages of the world for which they were initially made.
"Human beings acquire knowledge of and competence in a specific language via a complex process of socialisation" (Collins Dictionary of Sociology). The knowledge engineer does not have the time or the effort available, even if he/she wanted to, to undertake this complex process of socialisation, which would imply that there is no way he/she could quickly learn to communicate with the expert members of the domain. This supports our argument that an interface is required between the knowledge engineer and the experts to ensure no knowledge is lost in interpretation, or what could even be considered translation, from the language of the expert to the language of the knowledge engineer.

3.4 LANGUAGE AS A KNOWLEDGE ACQUISITION METAPHOR

3.4.1 Understanding and using specialist language

It is rare to see references to work in terminology and lexical semantics in knowledge acquisition literature. Terminology is the science of how terms are coined, how terms enter the language of the specialist community, how they are refined and adapted linguistically and epistemologically, how the term and its variant are used, and how terms become obsolete (see Picht and Draskau, 1985). Lexical semantics is the study of the meaning relationships between the lexical inventory of a natural language (see Cruse, 1986, 1992). Lexical semantics emphasises that word meaning can be dealt with exclusively in terms of relations between lexical items. Adequate accounts of word meaning must also take into account the fact that these relations should somehow be related to abstract concepts and the potential interrelationship between the concepts.
We believe that if knowledge engineers were able to exploit the terminological, syntactic and semantic constructs used by experts for disseminating knowledge in their specialist language, through interviews and domain texts, there is a possibility that interpretation will be reduced and better accuracy ensured.

Before the interview, the knowledge engineer, merely by collecting readily available domain text, can build and use a terminology collection of the domain for overcoming the inevitable terminological barrier between him/herself and the expert(s). Such terminology is vital in preparing interview scripts as described in Chapter 2 above. After the interview, textual analysis can be focused, for example, on extracting heuristics from the interview transcript (and other domain texts) or on extracting more terminology for refining the object-base. The transcript is a good example of special language text: full of specialist terms and phrases, it is a narrative text that aims to inform its reader. The knowledge engineer has to understand the terms, sentences and long stretches of text to extract both problem-solving knowledge and meta-knowledge, useful for explanation and justification.

Access to a terminology data bank should, in principle, alleviate problems related to the understanding of specialist terms. There are a number of complications in using a conventional term base. First, term bases are expensive to build and not every specialism comes ready with its own terminology data bank: in the case of emergent sub-disciplines of science and technology, a term base is usually a post-dated artefact, available, if at all, after a gap of five to ten years. Second, assuming the term base is available, the definition of a term, indeed definitions of words in a general language dictionary, are generally expressed in terms of between three to six other terms, therefore in most term bases there are terms with pretty
opaque and at times substantially circular definitions. Third, term bases are designed for the use of translators, technical writers and information scientists, consequently the cognitive bias in the design is more oriented towards language production and learning: it is, therefore, to be expected that data contained in the term base will not enlighten a knowledge engineer about problem-solving tasks. Fourth, the data structures used in the design of term bases stress the atomicity of individual terms: the use of relational tables, records and pointers, do not exploit the interconnectivity and interdependence of the terms (Ahmad, 1993; Griffin and Ahmad, 1994).

Despite the above mentioned reservations, we believe that whilst terminology data banks, with their current structure and cognitive orientation, may not be quite as relevant as the knowledge engineer would like, their use reduces the time spent in expert interviews and improves the accuracy of the output.

3.4.2 Improvements in text analysis and its affect on knowledge acquisition

As far back as DIMES in the late 1980’s, we had started to consider the importance of terminology in knowledge acquisition. In the early phases of the DIMES project a small initiation corpus was put together from medical text books and journal papers about breast cancer. At the time these texts were not available in machine readable form and therefore had to be typed into a computer (or punched as it was known).

At the time computer based text analysis tools were not very sophisticated. However, for DIMES, we used the Oxford Concordance Tool, which provided us with a list of words used
in the texts and a frequency of their occurrence. This "wordlist" was given to the domain 
experts who were asked to mark any word which they considered specific to their field, what 
we call domain specific or Language for Special Purposes (LSP) terms, part of the vocabulary 
of the specialist language discussed above (Ahmad et al., 1991).

These terms were only single words, not compounds, the importance of which we discuss 
later. However, they did provide the knowledge engineer with something to look for in the 
interview transcripts which he/she knew was likely to be important to the system. Also, these 
words were ideal candidates for the objects and properties of the system, and where possible 
such terms were used during object-base building.

By 1990, when the W-RAISA project started, the AI Group at the University of Surrey had 
started to develop a terminology management tool-kit as part of the Translator's Workbench 
(Ahmad et al., 1990, 1993; Ahmad and Rogers, 1992). This included two tools of particular 
interest to knowledge engineers: KonText, a text analysis tool far more sophisticated than the 
concordance program used in DIMES, and a term base editor, which enabled the building of 
terminology databanks as described above.

As with DIMES, we started the W-RAISA project by collecting texts from the domain 
practitioners: codes of practice, manuals and technical reports from particular cases. Again, 
almost all of these were not in machine readable form and, as with DIMES, had to be 
punched: a very time consuming and expensive task.
However, once in machine readable form, these texts could be passed through KonText. KonText enabled the user to request a wordlist, with frequency information of the word’s usage, and an index, which could take the user directly from the output to the occurrence of the word in the corpus. KonText also allowed users to list words they wished to include exclusively in the output or words they wished to exclude from the output (Ahmad et al., 1992a).

By creating a list of known general language words, such as the, a, of, on, etc., and excluding them, KonText’s output was much reduced in size and much richer in LSP content: this made the knowledge engineer and domain expert’s job much easier when asked to find domain specific terminology.

The most important feature of KonText, however, was the concordance facility: an output format which listed any number of words to the left or right of the found term, enabling the knowledge engineer to find compound terms. Terminologists had noted long before that LSP terms were generally compound, that is they contained two or more words. Indeed, if the knowledge engineer looks up single word components of a compound term in the dictionary, he/she would find it extremely difficult to put together a definition for the whole compound: it is not until you see all components together, that is, in context, that a meaning becomes clear.

A termbank was created and the terms, both single word and compound, discovered via KonText and validated by the domain expert, were added. This termbank was used as the wordlist was used in DIMES. First, it provided the knowledge engineer with focal points to search for when browsing, electronically on a computer, the interview transcripts which had
been punched from an audio recording of the video. The transcripts themselves were also used with KonText to reveal further terms which were added to the termbank and subsequently used in the same way. Finding these terms in the text reduced enormously the time it took the knowledge engineer to (manually) produce the paper rule base.

Second, the termbank contained the names of objects and properties for use in the system's object base. When we created the termbank for W-RAISA, we included relationships which we envisaged may be helpful during object base development: this included “part of” to point to properties, and “kind of” to point to is-a hierarchies. These labels and the relationships between them formed what is now known as a mediating representation: encoding knowledge of the domain at a level understandable by knowledge engineers and experts alike.

By 1992, when the ELSIE project started, we had realised that the analysis of text could lead to much more than a list of terminology. By this time, the text analysis and term base development tools in the translator's workbench had evolved into a very sophisticated tool-kit called System Quirk (Ahmad and Holmes-Higgin, 1995; Ahmad et al., 1995). KonText had been updated with a facility which greatly improved the use of text analysis in knowledge engineering. There are a number of well-developed methodologies in information retrieval, communications theory and corpus linguistics that are based on the frequency of occurrence of a linguistic token (Salton 1989 and a selection of papers edited by Jacobs 1992). However, KonText was now able to compare the relative frequency of each (single word) term in the text with its corresponding relative frequency in the Longman's Corpus of Contemporary English, and give a ratio of the two (Ahmad, 1993). This ratio told the user whether or not a term should be considered as domain specific.
The following section describes in more detail how this new facility, as well as other System Quirk components, can be used to elicit objects, their properties, their possible values and even heuristic rules from a domain text or interview transcript.

3.4.3 Automatic elicitation of propositions from text

A curious statistic of written and spoken language is that something like 200 words make up 50% of the words used by speakers and writers of any language. These words, mainly determiners, modal verbs, prepositions etc., are classified as closed class words, because it is only over centuries that new words are either added or subtracted from this class (cf. thee and thou are excluded in English). The rest of the stock, millions of tokens, make up the other half of language in use. These are called open class words, precisely nouns, adjectives and full verbs, in that new words are constantly being added to this category. The distribution of word classes in a specialist text is very different to the distribution of word classes used in everyday English. This variance from general language is often referred to as the ‘weirdness of special language’ (Ahmad et al., 1993; Ahmad and Davies, 1994).

Our primary technique for identifying domain specific terminology is to use this fact, and compare the relative frequency of a word in a specialist language (SL) text with its relative frequency in general language (GL). If the ratio of SL/GL is greater than 1, the this word is used more often in specialist language than in general language, making it a domain specification. The likelihood is greater the higher the ratio, that is, if a word occurs 1000

---

1 This section is based on results presented in Griffin and Ahmad (1994).
times more frequently in a specialist text than in general language, it is much more likely to be a domain specific term than are with a ratio of 2:1.

To illustrate the point, let us consider an interview transcript from the ELSIE project, containing 8202 words, as a specialist text, and the Longman Corpus of Contemporary English, containing in excess of 28 million words (Summers, 1991), as a representation of general language use. We note that the relative frequency of the closed class words is very similar in the interview transcript and the Longman corpus (see Table 2).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>the</td>
<td>521</td>
<td>6.35</td>
<td>6.09</td>
<td>1.04</td>
</tr>
<tr>
<td>and</td>
<td>220</td>
<td>2.68</td>
<td>2.80</td>
<td>0.96</td>
</tr>
<tr>
<td>I</td>
<td>47</td>
<td>0.57</td>
<td>1.08</td>
<td>0.53</td>
</tr>
</tbody>
</table>

Table 2: Relative frequency of closed class words in special and general language

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>aquifer</td>
<td>37</td>
<td>4.51E-03</td>
<td>0</td>
<td>INF</td>
</tr>
<tr>
<td>drawdown</td>
<td>7</td>
<td>8.53E-04</td>
<td>0</td>
<td>INF</td>
</tr>
<tr>
<td>groundwater</td>
<td>29</td>
<td>3.54E-03</td>
<td>3.88E-07</td>
<td>9104.42</td>
</tr>
<tr>
<td>borehole</td>
<td>22</td>
<td>2.68E-03</td>
<td>4.85E-07</td>
<td>5525.44</td>
</tr>
<tr>
<td>catchment</td>
<td>8</td>
<td>9.75E-04</td>
<td>3.88E-07</td>
<td>2511.56</td>
</tr>
<tr>
<td>abstraction</td>
<td>36</td>
<td>4.39E-03</td>
<td>5.92E-06</td>
<td>741.12</td>
</tr>
<tr>
<td>licence</td>
<td>19</td>
<td>2.32E-03</td>
<td>5.53E-06</td>
<td>418.59</td>
</tr>
<tr>
<td>abstractions</td>
<td>7</td>
<td>8.53E-04</td>
<td>3.79E-06</td>
<td>225.40</td>
</tr>
<tr>
<td>river</td>
<td>34</td>
<td>4.15E-03</td>
<td>1.29E-04</td>
<td>32.20</td>
</tr>
<tr>
<td>water</td>
<td>68</td>
<td>8.29E-03</td>
<td>4.77E-04</td>
<td>17.38</td>
</tr>
</tbody>
</table>

Table 3: Relative frequency of terms in special and general language and their ratios
However, if we look at the frequencies of some open class terms we find that the relative frequency of these terms in the transcript is an order of magnitude greater than the relative frequency of these words in the Longman corpus (see Table 3).

Another technique which can present important information about potential terms is text concording. A concordance of a text provides an index of all words in a text corpus showing every contextual occurrence of a word. Figure 14 shows just three lines of an interview transcript when the domain specific term *aquifer* appears.

1_116 There will shortly be starting a national research project aimed at collecting information on the main *aquifer* properties of transmissivity and storativity for a wide range of *aquifers*

1_22 *Aquifers* may either have a granular matrix or fissured matrix and these two characteristics do have quite a marked effect, particularly on the proportion of the *aquifer* that is available to store water - a granular *aquifer* has a much larger proportion of volume which is available for storage of water, and this can be seen in the way that *aquifers* respond to changes in water level

1_18 ... we can distinguish between two basic types - confined *aquifers*, where the water is held under pressure by a layer of impermeable strata above the *aquifer*, and unconfined *aquifers* ...

**Figure 14: Concordance extract from an interview transcript**

Every contextual example of *aquifer* throws some light on the meaning of the word and how it is used. Note that expert use hyponymies (*aquifer* types: confined *aquifer* and unconfined *aquifer*), attributes (*aquifer* properties: transmissivity and storativity), causal relations (*aquifer* responds to change in water level), qualifications (granular *aquifer* has a much larger proportion of volume...) and so on. From the concordance in Figure 14, a knowledge engineer could produce object candidate descriptions, including their properties and possible values, as shown in Table 4.
Table 4: Possible object-property-value tuples elicited from the concordance

Another difference between general and special language is that general language texts seldom use compound terms with the same frequency as special language texts. Compound terms can be identified by the assumption that they must make up any text falling between two closed-class words, e.g. determiners, auxiliary verbs, conjunctions and so on. Figure 15 illustrates some compound terms in the interview transcript.

Figure 15: Compound terms from the interview transcript

Compound terms usually provide the knowledge engineer with more domain model structure than the single word terms. Compound terms are often directly convertible into object-property value triples, such as those elicited from the concordance in Figure 15 above. Further values are often discovered by looking at the compound term in context, through further concording.

These terms may also provide clues to the hierarchy of concepts in the model. Figure 15, for example, hints that rather than aquifer having a type property, several subclasses of aquifer may exist such as confined aquifer and unconfined aquifer in a type-of hierarchy. Our final
text analysis technique involves the searching for linguistic ‘cues’ in a text which may point
to rule or task descriptions. Rules are usually written as IF ... THEN ... structures, and indeed,
if one were to search for the above pattern in the interview transcript, one would find the
following:

IF    casing is required in the borehole
THEN  careful geological control is needed
      AND it may be necessary to carry out geophysical borehole logging.

However, it is rare to find such obvious examples of rules within a text. Ahmad (1993)
described a whole range of words and phrases used by experts to encode heuristics of this type
(see Table 5 below). If we were to look at the interview transcript with cue words because,
when and so that, then we may determine the rule shown in Table 6 below.

<table>
<thead>
<tr>
<th>affect</th>
<th>as a rule</th>
<th>as long as</th>
<th>assuming</th>
<th>because</th>
<th>customarily</th>
</tr>
</thead>
<tbody>
<tr>
<td>due to</td>
<td>effect of</td>
<td>generally</td>
<td>hypothesis</td>
<td>if</td>
<td>if then</td>
</tr>
<tr>
<td>In general</td>
<td>ordinarily</td>
<td>precondition</td>
<td>premise</td>
<td>provided</td>
<td>proviso</td>
</tr>
<tr>
<td>reason</td>
<td>regularly</td>
<td>rule of thumb</td>
<td>seldom</td>
<td>so that</td>
<td>to ensure</td>
</tr>
<tr>
<td>typically</td>
<td>unless</td>
<td>usually</td>
<td>when</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 5: Examples of semantic cues for locating ‘rules’ in text**

| because | Effluent returns, as I mentioned there, that’s a very important thing **because** it can make the difference between an acceptable or non-acceptable abstraction, particularly when one is looking to see the effluent results, particularly, say from water supply abstractions, is coming back to the system above the point of abstraction, or very close to it, **so that** you can be utilising a process of re-circulation of the resource. | If the effluent is coming back to the resource system above or very close to the point of abstraction then a process of re-circulation of the resource can be utilised and a non-acceptable abstraction may become acceptable. |
| when    | | |
| so that | | |

**Table 6: Example of candidate rule found with semantic cues**
3.4.4 Benefits of the Linguistic Approach

Extracting domain specific terms or summarising a transcript, or any other domain text, to sentences containing cues for rule and task descriptions, not only reduces the amount of time required to produce the paper knowledge base, but also the knowledge elicitation expertise. For example, we gave an interview transcript in machine-readable form to a linguist colleague, with no knowledge engineering experience, and asked her to use System Quirk to analyse the text as described above. The results were surprising. The linguist identified 58 propositions, of which only ten were amended and three deleted by an experienced knowledge engineer (the author), who subsequently added only 19 more propositions, mainly concept descriptions. This partial de-skilling of the knowledge engineer’s job could lead to much cheaper knowledge acquisition, which could at last bring knowledge based systems into the mainstream of information technology systems.

3.5 Enhancing a classical approach with societal and linguistic metaphors

In 1988, the DIMES project followed a classical knowledge acquisition methodology, outlined in Figure 16 below, and relied heavily on one expert to provide the knowledge (Griffin, 1990). Ten interviews were recorded on video tape and transcribed by the knowledge engineers. The transcripts were analysed by linguistic tools in an attempt to extract knowledge directly from the text. The linguistic tools available at this time were fairly primitive compared with the sophisticated tool-kits we use today, but we performed a text concordance searching for a set of key words: sentences where one or more of these words occurred were regarded as potential sources of useful knowledge.
These sentences were simply listed for the knowledge engineer, however, if nothing else, this technique accelerated and made easier the interview analysis process. The knowledge engineer was still responsible for eliciting any useful information found in these sentences: he would record his findings in a paper knowledge base. This archive was made up of simple production rules in the typical ‘IF ... THEN...' notation, categorised under several subject headings chosen at will by the knowledge engineer. The first paper knowledge base of DIMES contained 49 such production rules.

The knowledge acquisition methodology employed within the W-RAISA project (1990) is illustrated in Figure 17. The first stage involved the analysis of a Code of Practice document supplied to the knowledge engineer by the project manager. The document was in paper form only, therefore it was punched and then analysed with a more sophisticated term extraction tool, the knowledge engineer making a note of all terminology which appeared to be domain
specific and eliciting any obvious rules. These rules tended to be very high-level and so were considered to be tasks within the system. Objects and slots of the first prototype were based purely on the domain specific terminology extracted.

![Diagram of knowledge acquisition methodology in W-RAISA](image)

**Figure 17: Knowledge acquisition methodology in W-RAISA**

An informal (overview) interview took place between the knowledge engineer and the domain expert to familiarise the knowledge engineer with the domain, to outline the practices and tasks undertaken by a Water Resources Officer, help the domain expert understand the nature of, and problems relating to the development of, an expert system, and verify the terminology elicited from the textual sources.

A think-aloud interview was held later between the same knowledge engineer and domain expert. Eight case studies were selected by the domain expert who explained the
determination process for each. The whole interview, totalling almost four hours, was filmed in a professional studio and produced on video. This was verified in full by the domain expert, however due to the limited time of the project, not all of the case studies were transcribed. Automated text analysis techniques were applied to the transcripts (and to the original reports used as case studies) to elicit problem solving information. The primary technique was a concordance of the domain specific terminology identified earlier in the project. The problem solving information elicited was then transformed into "If...Then..." constructs by the knowledge engineer to form the paper rule-base.

This mediating representation consisted of a list of rules and objects written in an abstract knowledge representation language based heavily on the natural language syntax of MARVIN (Holmes-Higgin, 1987). The high level of abstraction was desirable for two main reasons. First, the abstraction facilitates ready implementation, with little adaptation, within the constrained environment of the development vehicle, in this case an expert system shell (Ahmad and Griffin, 1991). Second, the readability of the rule- and object-sets had to be such that the domain expert was able to read the knowledge. The knowledge was validated and verified through user trials and corrected and enhanced accordingly.

The methodology used for knowledge acquisition within the ELSIE project (1992) is described in detail in the following chapter. An overview of the methodology, to illustrate differences with our previous approach, is shown in Figure 18.
Table 7 shows the growing dependency on the domain practitioners and the greater exploitation of text analysis within knowledge acquisition over the last 6 years at the University of Surrey. Even in DIMES we knew that the texts of the domain were important, but not until ELSIE could we formally identify concepts, properties and values from this text, and then paper rules and task descriptions from interview transcripts.

Figure 19 shows what we consider to be the improved performance in knowledge acquisition between DIMES, W-RAISA and ELSIE. This is based purely on the number of objects, rules and tasks elicited compared to the amount of time spent on interviewing experts, the most
expensive part of the knowledge acquisition process. Although DIMES had ten hours of video taped interviews, (compared with W-RAISA and ELSIE which both had four), the ELSIE project elicited many more tasks, rules and domain objects than the other two.

<table>
<thead>
<tr>
<th>Project</th>
<th>Domain Practitioner Interaction</th>
<th>Text Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIMES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1988-1990</td>
<td>Interviewee</td>
<td>Corpus creation (Initiation)</td>
</tr>
<tr>
<td></td>
<td>Identifying terms from wordlist</td>
<td>Simple concordance analysis of Corpus</td>
</tr>
<tr>
<td></td>
<td>Testing</td>
<td>-&gt; List of words</td>
</tr>
<tr>
<td>W-RAISA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1990-1991</td>
<td>Interview topic identification</td>
<td>Corpus creation (Initiation)</td>
</tr>
<tr>
<td></td>
<td>Interviewee</td>
<td>Treatment of Corpus and Interview transcripts with KonText</td>
</tr>
<tr>
<td></td>
<td>Identifying terms from wordlist</td>
<td>-&gt; List of terms (including compounds)</td>
</tr>
<tr>
<td></td>
<td>Review Paper KB</td>
<td>Termbank creation</td>
</tr>
<tr>
<td></td>
<td>Testing</td>
<td></td>
</tr>
<tr>
<td>ELSIE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1992-1994</td>
<td>Interview topic identification</td>
<td>Corpus creation (Initiation)</td>
</tr>
<tr>
<td></td>
<td>Expert and interviewer selection</td>
<td>Treatment of Corpus with KonText</td>
</tr>
<tr>
<td></td>
<td>Interview content identification</td>
<td>-&gt; object names, properties, values</td>
</tr>
<tr>
<td></td>
<td>Interview scripting</td>
<td>Termbank creation</td>
</tr>
<tr>
<td></td>
<td>Interview validation</td>
<td>-&gt; object base creation support</td>
</tr>
<tr>
<td></td>
<td>Object base validation</td>
<td>Treatment of Interview transcripts with KonText (Mature corpus)</td>
</tr>
<tr>
<td></td>
<td>Task base modelling</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Paper KB validation</td>
<td>-&gt; object names, properties, values</td>
</tr>
<tr>
<td></td>
<td>Prototype validation</td>
<td>-&gt; paper rules</td>
</tr>
<tr>
<td></td>
<td>Testing</td>
<td>-&gt; paper tasks</td>
</tr>
</tbody>
</table>

Table 7: Evolution of a Knowledge Acquisition Methodology

We attribute this performance increase to changes in our knowledge acquisition methodology since the DIMES project. Throughout the last five years we have used some techniques consistently: some we have adapted and improved, and some new ones have evolved from our research (see Table 8).
Table 9 expands on how one of these techniques, brainstorming and consensus decision-making, has increased since the DIMES project. It shows how the responsibility for the majority of tasks has shifted from a single person, whether the knowledge engineer or a project manager, to what we call the Domain Interface Group, leading to results based on consensus of all parties concerned, a central theme in our methodology, described in detail in the following chapter.
<table>
<thead>
<tr>
<th>Shared techniques</th>
<th>Expert interviews</th>
<th>initiation corpus creation</th>
<th>Text analysis (producing wordlists)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improved techniques</td>
<td>Expert interview content identification</td>
<td>Expert selection</td>
<td>Mature corpus creation</td>
</tr>
<tr>
<td>Evolved techniques</td>
<td>Group dynamics</td>
<td>Brainstorming and consensus decision making</td>
<td></td>
</tr>
</tbody>
</table>

Table 8: Techniques used, adapted and evolved for knowledge acquisition projects since 1988 at the University of Surrey

<table>
<thead>
<tr>
<th>Knowledge Acquisition Task</th>
<th>DIMES</th>
<th>W-RAISA</th>
<th>ELSIE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic identification</td>
<td>Predetermined</td>
<td>Discussion + Consensus</td>
<td>Brainstorming + Consensus</td>
</tr>
<tr>
<td>Expert selection</td>
<td>Predetermined</td>
<td>Predetermined</td>
<td>Brainstorming + Consensus</td>
</tr>
<tr>
<td>Interview scripting</td>
<td>None</td>
<td>Discussion</td>
<td>Brainstorming + Consensus</td>
</tr>
<tr>
<td>Object base modelling</td>
<td>Knowledge Engineer</td>
<td>Knowledge Engineer</td>
<td>Brainstorming + Consensus</td>
</tr>
<tr>
<td>Paper knowledge base production</td>
<td>Knowledge Engineer</td>
<td>Knowledge Engineer</td>
<td>Knowledge Engineer</td>
</tr>
<tr>
<td>Task base structuring</td>
<td>Knowledge Engineer</td>
<td>Knowledge Engineer</td>
<td>Brainstorming + Consensus</td>
</tr>
<tr>
<td>Prototyping</td>
<td>Knowledge Engineer</td>
<td>Knowledge Engineer</td>
<td>Knowledge Engineer</td>
</tr>
<tr>
<td>Testing</td>
<td>Consensus</td>
<td>Consensus</td>
<td>Consensus</td>
</tr>
</tbody>
</table>

Table 9: The changing focus of task responsibility in DIMES, W-RAISA and ELSIE
4. A PSYCHOLOGICAL, SOCIETAL AND LINGUISTIC METHODOLOGY

4.1 OVERVIEW

In this chapter we describe our knowledge acquisition methodology, combining conventional psychological techniques with those borrowed and adapted from studies in societal behaviour and linguistics, as described in Chapter 3. We reproduce below the overview of this methodology (Figure 20) as an outline to the contents of this chapter in which each stage is described in detail.

<table>
<thead>
<tr>
<th>4.2 The Players</th>
<th>Knowledge Engineer</th>
<th>Domain Interface Group</th>
<th>Domain Experts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Identify domain texts</td>
<td>Brainstorming and Consensus</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4.3 Initiation</th>
<th>Analyse domain texts</th>
<th>Using sophisticated text analysis tools</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Produce term bank</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4.4 Interviews</th>
<th>Psychological interview techniques</th>
<th>Brainstorming and Consensus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Identify interview topics</td>
<td>Interview experts</td>
</tr>
<tr>
<td></td>
<td>Interviews content experts</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Interviewers</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4.5 Analysis</th>
<th>Transcribe interviews</th>
<th>Validate and annotate transcripts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Analyse transcripts</td>
<td>Using sophisticated text analysis tools</td>
</tr>
<tr>
<td></td>
<td>Produce paper knowledge base</td>
<td>Using organisation chart tool to display task structure</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4.6 Modelling</th>
<th>Review and amend paper knowledge base</th>
<th>Brainstorming and Consensus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Produce prototypes</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4.7 Prototyping</th>
<th>Review and amend knowledge base</th>
<th>Structured walk throughs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Produce prototypes</td>
<td>Brainstorming and Consensus</td>
</tr>
</tbody>
</table>

Figure 20: Outline of Chapter 4 with respect to stages of our methodology
Although we present these stages sequentially, the output of one becoming the input of the next, it is not our suggestion that the methodology can, or should, be followed through from start to finish, to produce an expert system. It is likely that most stages of the methodology will be revisited on several occasions throughout the project lifecycle, whether to acquire 'missing' knowledge discovered during prototyping or to revalidate the paper knowledge base after a domain-wide change in policy, for example.

4.2 The Players

As we discussed in Chapter 2, the knowledge of a specialist domain is the intellectual property of the domain community. The community establishes itself, over a period of time, as a group within a larger society, and like any other distinct group has its own leaders - the experts, those who follow the leaders - novices, those who apply and refine the knowledge - peers, and those within or outside the domain who insist on criticising the leaders and on rejecting the extant domain knowledge - the opposition. We reiterate here that we do not consider the knowledge of the domain to reside purely in the heads of the so-called experts: the knowledge appears to be the result of a group dynamic: the leaders owe their position to the existence of the novices, peers and the opposition; the knowledge is socially situated. This group dynamic has to be taken into account during knowledge engineering, particularly in knowledge elicitation.

As we found in Chapter 2, recent literature in expert systems development discusses the role of 'multiple experts' but does not provide any method to follow in their use; it appears that a single expert is still preferred. Indeed, the knowledge acquisition literature contains many
reasons why not to work with multiple experts on a project (for example Bradshaw et al., 1993; Ford and Adam-Webber, 1992; Greenwell, 1988), however, our experience indicates that in some domains it is the only option.

In many expert system projects, the knowledge engineer has to work without much assistance from the domain community. This means that unless the knowledge engineer is very knowledgeable about the domain, there is an extensive training period where the knowledge engineer has to depend on the domain expert quite significantly. Furthermore, the knowledge engineer may or may not be willing to produce a critique of what the expert has said during a typical interview. Domain practitioners, on the other hand, have an applicative view of the knowledge and may indeed have alternatives to offer: the so-called ‘real world’ input that is so much coveted by the expert systems community.

We believe that a domain expert to be interviewed during knowledge acquisition should ideally be each of the following: charismatic, authoritative, articulate, clever, knowledgeable and experienced. As we described in Chapter 2, more articulate experts have been found to provide more domain knowledge during face to face interview and extroverts deliver this knowledge faster (Burton et al., 1987). To the knowledge engineer, anyone more familiar with the application domain than him/herself would appear knowledgeable and experienced, and quite likely authoritative and clever, yet may not be any of these to a great degree. Such judgement can not be regarded as considered because it lacks domain contact: the knowledge engineer may have only one contact and no time to consolidate.
We believe that this lack of considered judgement in choosing experts for ‘real world’ applications could be disastrous to the knowledge acquisition process. To avoid this, we feel that the knowledge engineer should be guided in the selection process by other members of the domain. The members of the domain that we have in mind are easily characterised but not easily found. They should have a solid understanding of the domain, gained through practical experience not university education, as we are looking here for more real-world input than learned knowledge. They should have good communication skills and the ability to speak about their work in “plain English” so that the knowledge engineer may understand, although domain specific glossaries as described in Chapter 3 are of help here. Most importantly, they must be highly motivated and committed to the success of the project, and have the freedom within their work to make themselves available for the time consuming tasks of the knowledge acquisition process.

The domain experts themselves certainly meet the requirement of having working knowledge of the subject, but it is unlikely that they can meet many of the others. Leading experts in most domains do possess extremely good communication skills, however the ability to relate their knowledge to their peers, at conferences, standards meetings and so on, does not necessarily imply that they have the ability to work with complete novices in their domain, as a knowledge engineer must be considered, to model their concept of the domain knowledge. Another point arising over the past six years is that unless the project was their own idea, an expert may not be committed to its success: on the contrary, he/she has been known to actively oppose the expert system, seeing it as a threat to his/her elitism. Finally, leading experts are a scarce resource, their time extremely valuable, and expensive.
Therefore, we are looking for experienced domain *practitioners* to help the knowledge engineer interface with the domain in question. To effectively manage the input of these practitioners, encourage communication and provide an environment in which each practitioner could feel part of a team, we formed what we call a “Domain Interface Group” (DIG), illustrated in Figure 21.

![Figure 21: The Domain Interface Group](image)

The introduction of the DIG presents a number of opportunities for the knowledge engineer. However, the successful exploitation of such opportunities requires careful planning on the part of the engineer. Our experience suggests that in contrast to typical expert systems projects, where the knowledge engineer is solely responsible to extract knowledge from an area which may be emerging, we have relied almost entirely on the DIG.
The use of the DIG introduces the role of group facilitator for the knowledge engineer, a new role which he/she may not be accustomed to, and has little training in. What the knowledge engineer facilitates is the elicitation and formalisation of knowledge which is undocumented and largely experiential. This role changes during the knowledge formalisation phase, where the engineer focuses on selecting appropriate knowledge representation and reasoning strategies in common with conventional knowledge engineering.

Group facilitation is a subject in its own right and the literature here is fairly recent. The facilitative role requires the knowledge engineer to adopt techniques that can deal with the collective interactions associated with a group dynamic: intra-group processes like leadership, power and power shifts, decision making, group reactions to other members, group cohesion, and so on. The technique we have adopted is that of brainstorming the domain interface group on a range of knowledge elicitation tasks.

A brainstorming session is generally thought of as a meeting of people in order to develop ideas together, or "intensive discussion to solve problems or generate ideas" (New Collins Concise Dictionary). Generating ideas and solving problems is of great importance during knowledge acquisition, however we are more focused at this stage on reaching consensus.

Not all definitions are optimistic about the use of the brainstorming technique: "the attempt, often unsuccessful, to generate useful new ideas..." states the Macmillan Dictionary of Psychology, hinting that there are many potential pitfalls involved. However, this description of the technique involves "encouraging a group to talk in an excited way and to express free associations and novel ideas, however disconnected from one another or from the main theme.
of the discussion”. Although freedom of speech is obviously important, some constraints must be imposed at an early stage: there must be some focus of the discussion, and this direction must be enforced by the knowledge engineer.

The references to brainstorming are not frequent in the knowledge acquisition literature, perhaps indicated by a short and rather vague entry on the topic in Shapiro’s Encyclopaedia of Artificial Intelligence by Boose (1992): “rapidly generate a large number of ideas” using “Crawford Slip Method” (referencing Rusk and Krone, 1984).

Data presented by Dickson Lukose at the ICCS’94 Workshop on Knowledge Acquisition using Conceptual Graph Theory, showed that from 31 research groups in knowledge acquisition, only two reported the use of brainstorming. The most referenced use is that of McGraw and Seale, whose approach to brainstorming multiple experts is described by Greenwell (1988). More recent work by Boose et al., (1993) is discussed in Chapter 2. However, little detail can be found on what exactly the knowledge engineer should do during such sessions.

Our approach to brainstorming the DIG is based on two distinct phases: the discovery phase and the revision phase (illustrated in Figure 22). During the discovery phase, the problem area is described by the knowledge engineer, along with the goal of the session and the rules which they must follow in achieving that goal. Members of the DIG then call out possible solutions, either when they can be heard or in turn if necessary, until all possible solutions are exhausted. The knowledge engineer records all of these solutions on a flip chart or white board, where they can be seen by the participants.
During the revision phase, the solutions are reviewed, amended and constrained where applicable by grouping and scoring (similar to the approach outlined in Boose et al., 1993). All amendments are recorded on the flip chart or white board by the knowledge engineer until he/she is satisfied that the goal has been reached. At the end of the session the knowledge engineer reproduces the results from the flip chart in a distributable format suitable for later use in the knowledge acquisition process.
We believe that the DIG is an essential part of expert system development, particularly during the knowledge acquisition phase. This group has the domain contact to answer key questions with the considered judgement that the knowledge engineer lacks. Details of the topics, goals, rules, the revising and constraining processes and the format of outputs for each brainstorming session are described in the following sections, as are details of the DIG’s roles throughout the development process. An overview of the roles is shown below in Table 10.

<table>
<thead>
<tr>
<th>Initiation</th>
<th>Identify Initiation Corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interview</td>
<td>Identify interview topics</td>
</tr>
<tr>
<td></td>
<td>Identify interview contents</td>
</tr>
<tr>
<td></td>
<td>Identify suitable experts</td>
</tr>
<tr>
<td></td>
<td>Select interviewer</td>
</tr>
<tr>
<td></td>
<td>Conduct interview</td>
</tr>
<tr>
<td>Analysis / Critique</td>
<td>Review transcript</td>
</tr>
<tr>
<td></td>
<td>Review paper rule base</td>
</tr>
<tr>
<td>Modelling</td>
<td>Participate in structured walk-throughs</td>
</tr>
<tr>
<td></td>
<td>Test knowledge base</td>
</tr>
</tbody>
</table>

Table 10: Roles of the Domain Interface Group

4.3 INITIATION

The *initiation* phase of our knowledge acquisition methodology is primarily to elicit the concepts used by members of the domain community in describing their domain. As described in Chapter 3, much of this information is contained in the domain specific terminology used by the community’s members: words, compounds and phrases.
This is identical to the goal of Wood and Ford's (1993) "descriptive elicitation" phase. However, they interview experts at this stage. They have introduced a typology of questions which can be used in an "attempt to create a situation in which the expert will describe the domain in the natural way, using domain-specific labels for important concepts." These questions, including "what would you call..?" and "how would you describe..?", are carefully selected to avoid the reductive bias which so commonly afflicts knowledge engineer's at this stage. A particular problem here is the form of reductive bias which Spradley (1979) calls "translation competence", arising when experts "translate" their "cultural reality" for the convenience of the knowledge engineer, avoiding the terminology which he/she will not be accustomed with: the terminology we need at this stage.

To avoid this risk, we do not use an expert at this stage, but rather rely on domain texts: papers, reports and articles written by members of the community for the community. In such texts there is absolutely no reason for translation of domain terms, therefore the reductive bias is avoided.

The selection of domain texts to be used is the "problem" posed to the Domain Interface Group (DIG) by the knowledge engineer facilitating the first of many brainstorming sessions. The goal of the session is to provide a good coverage of the domain in a variety of textual formats from in-house newspaper / magazine articles to specialist scientific reports. The rules of the session are that only community wide texts can be suggested (localised variants of terms at this stage could be misleading) and where possible texts should be available in machine readable form. No grouping or scoring of the resultant list takes place in the session;
if time, resources and disk space permit, every text listed should be analysed, however, 100,000 words is normally sufficient for what is called the “initiation corpus”.

The texts are analysed by the knowledge engineer with the help of a text analysis tool such as System Quirk (Ahmad and Holmes-Higgin, 1995; Ahmad et al., 1995), as described in Chapter 3. Domain specific single-word terms are extracted by comparing the frequency of occurrence of each word in each text of the initiation corpus with its frequency of occurrence in a large scale, general language corpus, such as the Longman Corpus of Contemporary English (Summers, 1991). These single-word terms are then used in text concording to find compound terms which, as we showed in Chapter 3, are the mainstay of domain specific terminology and from which concepts of the domain are identified.

All terms extracted need to be stored in a suitable mediating representation, which can be referenced by the knowledge engineer, DIG and experts, as and when required throughout the entire knowledge acquisition and modelling process. For this purpose we advocate the use of specially adapted terminology databanks (or termbases). These relational databases provide descriptions, definitions and contextual examples of the terminology stored, as well as relationships between the terms. These relationships are of primary importance as they will eventually form the basis of the domain model. Tools like System Quirk enable these relationships to be viewed in a variety of formats, from hypertext to concept maps (Chapter 2 includes details of the use of these formats as mediating representations).

The analysis and term base creation is undertaken solely by the knowledge engineer, or perhaps by a linguist if one is available on the team. The important point here is that a
member of the domain community is not required to build the term base to avoid reductive bias: this is avoided because the 'human' involvement is removed and therefore so is the interpretation.

Naturally, members of the DIG and experts may not have access to the hardware and software required to view the resulting termbank, therefore, the final task for the knowledge engineer in this stage is to make the results available to all players. Currently, this is achieved through publishing tools in System Quirk which produce paper versions of the terminology in a variety of formats from glossaries to dictionaries and thesaurus based on the underlying relationships (Ahmad et al., 1992b). It is worth noting, however, that current work at the University of Surrey is making termbank access available via the world wide web (Ahmad and Collingham, 1996). This makes the mediating representation available for all players in the knowledge acquisition process, however distributed geographically, and whatever hardware they use, which is particularly useful for the broad domains that we are interested in. Furthermore, making such a resource more available, allowing people in the same, and other related projects, to see the knowledge, incorporate it into their own work, and update it by making the "tools accessible to everyone in the group", brings us one step closer to Clancey's (1993) vision of "designing for communities of practice", as described in Chapter 1.

4.4 Interviews

The next stage of our methodology is the interview phase. Our interview phase involves more input from the societal metaphor than psychology, and involves the Domain Interface Group (DIG) extensively, from choosing the interview topics to choosing the experts to be
interviewed, and eventually interviewing them. The resulting reduced role of the knowledge engineer is once again an attempt to avoid the risks of reductive bias, so commonly the pitfall of conventional techniques because of the interpretation of domain specific language.

4.4.1 Identify interview topics

Interview topic selection is usually conducted in an ad-hoc manner and is the result of an informal meeting between the domain expert and the knowledge engineer. The fact that the knowledge engineer may or may not be familiar with the domain, and has little understanding of the dynamics of the given specialism, makes the knowledge engineer totally dependent on the expert, which we have shown can lead to considerable problems (Chapter 3).

The knowledge engineer facilitates a brainstorming session of the DIG to identify topics for the expert interviews, based on the following scenario. First, the group are told how many topics they must identify within the domain (the number of interviews the project can afford): the problem. They are told that the topics chosen must provide a good coverage of the domain: the goal. They are also told that the topics should cover issues popular with all participants, that there must be consensus in the selection of topics, and that anything available on paper should be identified but not included for interview: the rules.

The knowledge engineer creates a list of all topics mentioned by the participants: initially the importance of each is irrelevant, the aim here is to ensure that all possible topics are listed. Once the DIG is satisfied that the list is exhaustive, the revision phase begins. The revision phase requires the DIG to prioritise the list: there is a fixed number of interviews and the
topics will be taken in order of preference to all members, however, topics may be merged if necessary.

In the ELSIE project there were four interviews and the subject of abstraction licensing was split into the three main types of licence (groundwater, surface water and impoundments) and general legislative issues. This almost covered the entire subject, but it should be noted that the subject of abstraction licensing is already quite focused.

4.4.2 Outline interview content

The DIG is then required to outline the content of each interview identified. Again this is achieved by the knowledge engineer facilitating brainstorming sessions, one for each interview subject. The discovery phase for each session is very similar to the previous session, only the problem now is to identify more specific topics to be discussed within one of the four broad topics previously chosen. The goal is a script for the interview: this is needed to focus the interviewer and the expert on enough topics to cover the subject but within a fairly strict time limit (see Chapter 2 for details on the use of focused interviews). The rules are the same as before: the topics must cover issues important to all, they must all agree on the selected topics, and anything that can be found from a textual source should be excluded.

As with the previous session, the knowledge engineer creates a list of all topics mentioned by all participants. Once the DIG is satisfied that the list is exhaustive, the revision phase can begin. For these sessions there are several forms of revision and several constraints to be added. First, the DIG is asked by the knowledge engineer to group the listed topics together,
making some of them sub-topics as necessary. This grouping ensures that each part of an interview is focused and the expert does not have to switch between topics, which could cause confusion during analysis.

Once the grouping is complete the list is ordered. This scoring is facilitated through further brainstorming, the priority assigned reflecting the consensus view of the importance of each topic. There is no consideration, at the moment, for the length of a topic: the knowledge engineer knows that there is a limited time in which the interview must be conducted and therefore insists that the most important topics are discussed first.

An important point to recall at this stage is that any knowledge which could be elicited from a static source, a text book or video for example should be demoted to the bottom of the list and marked as such. An expert’s time is usually far too expensive to use discussing anything which has already been agreed upon and published as fact, and can therefore be elicited from text as described in Chapter 3. Sub-topics which contain such material should also be marked as such, enabling the expert to reference the knowledge source to be consulted during his discussion.

The final stage of outlining, once the contents have been identified, is to add timings to each topic. Due to budgetary or resource constraints, it is likely that the expert will only be available for a short time for interviewing. Preparation should be made for up to 25% more time than the planned interview duration: it would be an unthinkable waste of resources if a film studio is booked for an hour, a leading expert has given up an hour of his valuable time, and 50 minutes into the interview there are no further questions to answer or topics to discuss.
In our opinion it is far better to not complete an interview script than have any amount of time wasted.

The DIG estimate the time required for each topic, and sub-topic if necessary, to the nearest minute. Despite the above warnings, the DIG should ensure that ample time is allowed in the schedule to fully cover each topic. A knowledge engineer will have great difficulties during analysis with a topic discussion cut off part way through due to lack of time in the interview. If time is running out in an interview, the interviewer can choose a topic to end on which best fits in the time remaining.

The result of this brainstorming session is a flip-chart or whiteboard containing a list of topics and sub-topics, numbered in the order of importance, and annotated with an estimate of time required to discuss each. This should be rewritten by the knowledge engineer as illustrated in Figure 23, an extract from the output of a brainstorming session to outline an expert interview as part of the ELSIE experiments. The output will be used as a script to guide the interview and enable the interviewer to keep control of the time being used.

<table>
<thead>
<tr>
<th>Interview C: Surface Water Licensing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. What is an 'inland water' and what constitutes abstraction from an inland water?</td>
</tr>
<tr>
<td>(mention isolated water bodies)</td>
</tr>
<tr>
<td>2. Multiple channels</td>
</tr>
<tr>
<td>3. Exemptions from licensing - what are they?</td>
</tr>
<tr>
<td>4. Flow regimes - canals, drains/ditches, lowland rivers, upland rivers</td>
</tr>
</tbody>
</table>

Figure 23: Excerpt from the output of a brainstorming session with the ELSIE DIG to outline an expert interview
4.4.3 Identify interview participants

Only after interview topics have been defined to satisfy the requirements of the project do we recommend that the DIG begin to consider suitable experts for the interview: the interview outline provides a “profile of expertise” which must be satisfied by any candidates for the role. The DIG is asked to find an expert who has the knowledge and experience necessary to discuss the issues raised in the interview outline: the problem. The goal is to identify two or three candidate experts: one will eventually become the interviewee, the other one or two will be used for validation. However, knowledge and experience are not the only criteria on which an expert should be selected: many rules are added for the revision phase.

Once the initial list of candidates is agreed upon, the DIG is asked to consider further criteria, including a candidate’s ability to communicate, his articulateness and charisma, and his authority on the subject: whether or not he/she is respected for his/her views and opinions by his/her peers and staff. It has been said on many occasions that there is no substitute for experience. However, in our own experience, we have found that an experienced expert does not always have in-depth knowledge, and what good is knowledge if it cannot be successfully communicated? When we consider some of the expert selections in the ELSIE project, with hindsight we can see the advantages of choosing charismatic and articulate experts over simply knowledgeable and clever ones, an issue which Burton et al. (1987) have reported in the past, and one which we explore further in Chapter 5.

A factor which the DIG should not consider at this stage is the availability of a candidate, whether dependent on how busy he/she is or his/her location: an expert should not be
discarded at this stage just because he/she is based hundreds of miles away; if he/she is suitable then he/she must be listed. Table 11 outlines the criteria which should be considered and should not be considered during expert selection. It is not possible to order the criteria on any basis of importance: an ideal expert will possess all of the criteria to be considered, however we have marked what we feel are the minimum requirements for the successful candidate.

<table>
<thead>
<tr>
<th>Consider</th>
<th>Do not consider</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge*</td>
<td>Availability</td>
</tr>
<tr>
<td>Experience*</td>
<td>Location</td>
</tr>
<tr>
<td>Articulateness*</td>
<td>Personal relationship</td>
</tr>
<tr>
<td>Charisma</td>
<td></td>
</tr>
<tr>
<td>Authority</td>
<td></td>
</tr>
<tr>
<td>Cleverness</td>
<td></td>
</tr>
</tbody>
</table>

Table 11: Criteria to be considered / not considered during expert selection

As with topic identification, it is not important how many candidates are suggested. When the DIG is satisfied that the list is exhaustive, the knowledge engineer must facilitate a difficult brainstorming session which ranks the experts based on the criteria suggested above. This task is difficult because members of the DIG may have personal experiences with some of the candidates, some good and some bad, times of agreement and times of disagreement, and human nature makes it impossible to remove totally the elements of favouritism or grudges based on past encounters. This is another example of the benefit of a group in this role and the importance of consensus.
The final participant in a knowledge acquisition interview is the interviewer. Traditionally, this role has been undertaken by the knowledge engineer, however the DIG consists of people who know what knowledge is required to be elicited during the interview, having specified the topics to be discussed themselves, and are better equipped, with domain specific knowledge and terminology, to ask the right questions in the right context using the right vocabulary. It appears, therefore, that a more suitable candidate for the role of interviewer would be a member of the DIG.

The knowledge engineer once again facilitates a group discussion from which the member most suitable for the interviewing role is chosen. As with the choice of expert, consideration must be given to the member's awareness of the topic to be discussed, his/her practical experience of applying knowledge within the topic, and the member's articulateness: unlike the knowledge engineer, the DIG member is expected to hold a very dynamic discussion with the expert.

The results of this session are a list of expert candidates to either take part in the interview or validate the interview content afterwards, plus an interviewer from the DIG. The candidate experts on the list may then be approached, in order of preference, to take part in the knowledge acquisition process. It is important during the request for an expert's help to describe to him/her what is required: the interview outline with identified topics for discussion should accompany any letter sent to a candidate expert.

The candidate should also be informed that two experts will be required for the project: one to be interviewed and a shadow expert to verify and validate the elicited knowledge. This is
important because a highly ranked candidate may not have the resources available to take part in the interviewing but may be willing to read through the transcripts or watch the videos in his/her own time: rather than being omitted, such an expert should be encouraged to do so. The role of the shadow expert should also be considered by the DIG for a highly experienced, authoritative expert on the candidate list who they believe will be unsuitable for interviewing, for whatever reason, but whose input would be of enormous benefit to the project.

4.4.4 Interview the experts

The interviewing of a domain expert is not an easy task for a knowledge engineer who is not familiar with the domain terminology, or indeed with the domain itself, and has little understanding of the dynamics of the given specialism. He/she is almost entirely dependent on the expert to keep to the topics outlined and can only trust that the information being supplied is suitable for the model. This is not always the case, as the expert may have misinterpreted the objectives of the session, suffering the same problem communicating with the knowledge engineer before the interview as the knowledge engineer will suffer during it: a lack of understanding of the others’ terminology, resulting in reductive bias in both directions.

The DIG members are in a much stronger position to interview the experts for four main reasons. First, they have a thorough understanding of the domain terminology and with their practical experience should understand all of what the expert says. Second, the members are also fully aware of the required results of the interview from the knowledge engineer’s point of view, and can explain this to the expert in his/her own language much better than the knowledge engineer can. These first two points remove entirely the ‘translation competence’
problem of reductive bias described above. Third, practitioners from a specialist domain tend to enjoy discussing problems relating to their domain, and will generally hold a very lively discussion with the expert on issues of interest to themselves. Finally, the practitioner will have no problem correcting an expert or pointing him/her in the right direction if the focus of a topic is lost, something which a knowledge engineer could never do because of his lack of domain contact.

The interview should be professionally recorded, preferably in a studio, where the best possible lighting and sound can be arranged for production of the video. This video is an important by-product of knowledge acquisition: as well as being used during analysis and validation, it may well be useful elsewhere in the domain as a training aid (see Chapter 5). The video should concentrate on any diagrams or photographs used by the experts for clarification of the spoken word during analysis, and where possible, related video clips and close-ups of photographs and soon, should be edited into the interview.

4.5 ANALYSIS

4.5.1 Transcription

It is the role of the knowledge engineer, or more often a junior member of the knowledge engineering team or a secretary, to transcribe the video of an interview. The transcription must be verbatim and it is vitally important to ensure that the correct domain specific terms are used. The initiation term base is extremely helpful here: although these tools do not usually contain phonetic information, the glossary is normally sufficient to enable the correct
word or phrase to be identified. In the case that a term can not be found, it should be marked in the transcript for elaboration from the DIG or the interviewed expert in the next stage.

Also marked in the transcript should be cues referring to the pictorial information used by the expert during the interview. Where possible, such pictures, graphs, diagrams, photographs and so on should be included in the transcript at the appropriate point, otherwise they should be copied (reduced in size if necessary) and attached to the printed transcript. Finally, the knowledge engineer should annotate the transcript at suitable points with a reference to the video recorder’s timer.

The transcript should be written in a word processor to enable sensible fonts and formatting to be used for readability. Validating experts will be more effective reading well sized and spaced text than the ‘old-fashioned’ computer print out in fixed width, small spaced fonts. Also, space between lines of the transcript enables the validating expert to mark changes or add notes in-line with the text, rather than on a separate sheet which needs to be cross-referenced. All word processors have the ability to save the document as plain ASCII text which will normally be required by the text analysis tools in later stages.

The transcript and a copy of the video tape should be sent to the expert interviewed and to the shadow expert identified in the earlier brainstorming session. Interviews are much easier to watch than to read, and experts find it much less time-consuming to sit in front of the video than to read through the transcript. However, since the transcript is cross-referenced with timer readings, the expert can easily refer to the correct page if they wish to make a note or an amendment.
4.5.2 Critique

The interviewed and shadow experts comment on the interview transcript in their own time, avoiding the need to schedule expensive meetings in terms of their precious time and the project's budget. It is interesting, however, to use the DIG as more shadow experts to review the acquired knowledge as a group, in a session similar to the earlier brainstorming meetings.

The problem to be solved in this session is the verification of the knowledge, the goal being an annotated transcript which to the best of their knowledge is correct and complete. There are two main aspects of the knowledge for which the DIG's verification is essential; indeed, these two aspects may not be verified by the interviewed or shadow experts.

First, being practitioners within the domain, members of the DIG may raise practical issues or constraints in the real world which the experts may overlook or perhaps may not even be aware of. Second, since the DIG is made up of representatives from different regions of the domain, they should be able to spot any occurrence of the 'local expert' problem: an expert relying too greatly on his/her local knowledge resulting in a response that fails to reflect the domain-wide practice.

Discovering that knowledge is incomplete or only applicable in special contexts, is not a cause for concern, but does mean that further elicitation efforts are necessary (Wood and Ford, 1993). If the latter case arises, the knowledge engineer must facilitate a discussion of the DIG until a consensus is reached on whether to add knowledge to cover all possible variations in a general case or mark the knowledge as a special 'localised' case and add other localised cases.
as necessary until all regions are covered. If the DIG can not provide a satisfactory solution, it may be necessary to interview another expert on the same topic.

4.5.3 Annotation

The original transcripts must be annotated by the knowledge engineer with all comments, additions, amendments and deletions made by the interviewed experts, shadow experts and the DIG. For this process we recommend the use of a mark-up language such as the Standard Generalised Mark-up Language (SGML), or ISO 8879, a full and detailed description of which is given by Goldfarb (1990). The language defines elements within the text which are enclosed within two tags: a start tag of the form `<elementname attributes>`, and a corresponding end tag of the form `</elementname>`. The attributes describe the element's relation to other elements and external sources, and provide general information about the text enclosed between the tags. We use tags such as `<new>`, `<deleted>`, `<comment>` and `<crossref>` for the possible changes made to a transcript, each tag having a `revision` attribute identifying when the change was made. These provide an audit trail which is extremely useful when the DIG and users refer back to the transcript later in the project.

The annotated transcript can be considered to be another mediating representation of the acquired knowledge (note that the term base provides a mediating representation for knowledge extracted from the initiation corpus). Although the text is primarily in natural language and therefore far too abstract to be encoded in an expert system shell, the annotations and audit trail enable its use in browsing tools, particularly in hypertext, and hence promote and enrich communication between the experts, DIG members and knowledge
engineer co-operatively building the domain model, a description of a mediating representation put forward by Ford et al. (1993) and Bradshaw et al. (1993).

4.6 MODELLING

4.6.1 Automated text analysis

Manual analysis of the interview transcript in search of 'paper' rules and object descriptions is not only time consuming but also an extremely difficult task for the knowledge engineer. The primary problem is the lack of understanding of the domain terminology or specialist language used. It is likely that the knowledge engineer will not understand the meaning of an expert's response to questioning and will therefore either be looking for known domain specific terms, identified in the initiation term base, or trusting to experience to find 'semantic cues' in the text pointing to possible rules or objects.

We have fully automated the search for both of these by adapting the KonText tool of System Quirk, modifying the concordance and collocation facilities to provide a list of all sentences containing the semantic cues identified in Table 5 (Section 3.4.3) and all domain specific terms from the initiation term base (Griffin and Ahmad, 1994). The former produces the most valuable output at this stage: a list of sentences, all of which are potential 'propositions' for the paper knowledge base; candidate rules, tasks or object descriptions.

Rather than searching through the entire transcript looking for the 'cues' him/herself, the knowledge engineer is provided with a list of sentences which need only 'preparing' for the
paper knowledge base. This involves modifying the sentence to fit our next mediating representation: a structured form of natural language enforcing only a format (or syntax) on the sentences to bring some consistency to the paper knowledge base. Three types of propositions can be produced at this stage. First, *facts* can be provided such as ‘the value of A must be B or C’. Second, pseudo-production rules can be produced if the sentence can be structured inside an ‘if ... then ...’ format, with an optional ‘because’ clause. Third, *task descriptions* can be produced in the form of action statements, such as ‘to do X we must first complete Y and Z’.

It is worth reiterating here that, with the help of automated text analysis tools, it is not always necessary to use an experienced knowledge engineer for this stage. We gave an interview transcript to a linguist colleague with no knowledge engineering experience and asked her to use our modified System Quirk to analyse the text and produce propositions as described above. The linguist identified 58 propositions, of which only ten were amended and three deleted by an experienced knowledge engineer (the author), who subsequently added only 19 more propositions, mainly concept descriptions.

4.6.2 Paper knowledge base creation

The three types of proposition that make up the output of the transcript analysis stage need to be structured into the mediating representation that forms the paper knowledge base. When completed, this is the most important artefact produced by any knowledge acquisition methodology. We recommend a well structured yet highly human-orientated mediating representation for the paper knowledge base. We agree with Ford *et al.* (1993) that
"considerations of human efficiency far outweigh considerations of complex modelling problems". As they also point out, a model based description in a form that the user can intuitively understand has many advantages, the primary one being that it can serve to mediate communication between all of the players. If the analysis results are coded directly into an expert system shell there can be no further communication between the players because only the knowledge engineer can understand the model. Furthermore, it is very difficult to re-engineer knowledge encoded in one shell to work inside another; our mediating representation enables the knowledge to be recorded, amended and updated at a level where it can easily be used by any project in the same domain before any engineering takes place, making it a reusable resource.

We recommend that tasks identified from the transcript analysis are represented in a hierarchical tree structure, the most generic tasks at the top of the tree with component sub-tasks, required to complete the super-tasks, placed below; the top-most task will be the overall task for which the system is being developed. This hierarchical structure can be modelled in a variety of ways with a variety of tools. Concept maps, as described by Bradshaw et al. (1993), are a promising representation for this, however, we have in the past used a much simpler format: the organisation chart. In this formalism, the task of the system will be at the top of the chart, in the chairman’s slot; and the work required to complete this task would be represented at the lowest level of the chart, where one would expect to find the names of people who do all of the work!

As well as providing a suitable structure in which to place the task hierarchy, a further advantage of the organisation chart is that numerous cheap and commonly available tools are
available in which to create, and graphically view the structure, most of which have the capability to import and export the information between similar tools, adding to the reusability of the resource.

We do not add any further constraints to the format of the paper rules at this stage: apart from the fact that they must contain an ‘if’ and ‘then’, the contents of the antecedent and consequent can be free-flowing text in the natural language. Each rule is given a unique identifier for reference throughout the project. The knowledge engineer places the identifier for each rule into one of the tasks in the hierarchy. At this stage this can be a best guess position if necessary, as these will be verified later by the DIG. This results in a task hierarchy in which each task contains a list of rule references to rule descriptions stored in a separate text file. Advancements in operating system communication protocols, such as DDE, mean that the reference can easily be linked dynamically to the actual rule text, enabling the rule text to be displayed in an editor, for example, if its reference is selected. We have not, however, pursued such automation ourselves.

The final component of the paper knowledge base is the fact, concept or domain object base. This comprises the full set of domain concepts used in the rule and task descriptors, commonly associated to objects and their properties in frame- or object-based implementations. Again, concept maps could be, and indeed have been (Bradshaw et al., 1993), used for this representation, however we already have such information stored in the initiation term base. This term base is therefore updated with any new terms and relationships extracted from the transcript analysis.
These concepts are linked to the rules and tasks by adding a list of SGML-based tags, containing the entry labels of the concepts in the term base, to each of the rule and task descriptions. Again, links could easily be implemented to retrieve the concept information at the touch of a button from these references, but we have not pursued this automation.

4.6.3 Paper knowledge base review

At this point, the knowledge engineer facilitates a further brainstorming session of the DIG. The problem is to correctly structure and complete any missing elements in the task hierarchy and paper rule base. The goal is the all-important artefact of the paper knowledge base in the chosen mediating representation. The rules for the session are very simple: tasks and rules may be added, deleted or simply moved to new positions in the hierarchy, based, of course, on consensus within the group.

This session differs from the previous sessions we have described in two ways. First, there is no discovery phase, only revision: the knowledge was discovered during the interviews and the subsequent analysis and critique. Second, this is the first brainstorming session in which we introduce a computer, something we have purposely avoided in the past. Although nowadays most domain communities are familiar and comfortable with computers in the workplace, we do not recommend their use in brainstorming sessions for several reasons.

First, the use of a computer could be a distraction for the DIG members. Second, if a computer is used in place of a whiteboard or flipchart, changes made during the session are lost: although we do not at present study this history of change leading to the goal, we think
that somehow it is important and should be recorded. Third, there is generally no need to use a computer: the knowledge engineer is simply making lists, sorting them, scoring them and so on, but in using a whiteboard or flipchart he is obviously participating in a manner which could not be reproduced whilst sitting at a computer screen.

This session, however, is different: we need the computer. The knowledge engineer works at the computer with the task hierarchy displayed to the DIG via an overhead projector. This way, when a task needs to be added, deleted or moved, the knowledge engineer can reflect the changes immediately and graphically on the screen. For ELSIE, where there were over two hundred tasks in the hierarchy, this proved to be invaluable: it would have been impossible to amend a structure of such magnitude on a flip chart or whiteboard. The actual rules need not be on the screen but can be printed out and distributed to the DIG at the start of the session. The important point is to know where the rules fit into the task base, and this again would be difficult to show on paper or board, let alone amend.

The knowledge engineer and DIG should be prepared for this to be a very long session, perhaps several days. For ELSIE, for example, the session took over nine hours to correctly place all of the 200+ tasks and 200+ rules. The result of the session is the paper knowledge base from which the implemented system will be built: a full specification for an expert system.
4.7 Prototyping

4.7.1 Prototype development

Naturally, at some point, the knowledge engineer must start implementation of the knowledge base. Due to representation constraints in the chosen shell or programming language, it is usually impossible to implement the paper knowledge base exactly as modelled in the mediating representation, and the knowledge engineer often looks to make changes to the model to find a better ‘fit’ in the shell. In an attempt to minimise the variance between the paper and implemented model, we advocate an incremental approach to implementation through a series of prototype systems.

Each prototype should be validated by the DIG, as described below, to ensure consistency between models at various stages of implementation. We believe that the first stage of implementation should incorporate the task hierarchy and an algorithm to traverse the resulting tree structure. At this stage we recommend that the respective paper rules be attached to each task as a text description. This enables the DIG to walk through the task hierarchy, presented at each stage with a description of the current task and how it will be performed.

The next stage is to implement the rules in this framework. We recommend that this is undertaken in stages, maybe two or three specific tasks at a time, with walk-throughs by the DIG at each ‘checkpoint’. The rules need underlying objects present to reason over, hence we recommend that the object base is incrementally developed with the rule base. We also recommend that the user interface be developed as and when required to support the growing
A few simple dialog boxes will be required at the very start of implementation to help guide the user through the task hierarchy and display the rule texts. However, sophisticated data input screens will be required when the object- and rule-base begin to grow. We recommend, wherever possible, to create interfaces mimicking existing interfaces or paper forms from the domain: this will make the user feel more familiar with the new system.

A final point on prototyping is that we find it extremely useful to keep the paper rule-base attached to the task hierarchy throughout implementation and beyond. This allows the DIG validating the system and later, the end-users of the system, to see what the system is doing at any stage of a consultation, something which may be completely hidden by the implementation.

### 4.7.2 Structured walk-throughs

At each stage of the incremental prototype development, the DIG should meet for another brainstorming session. The problem that they are solving is the validation of all aspects of the implementation, within reasonable tolerances for shell constraints. By 'all aspects' we mean the completeness and correctness of the task hierarchy, the rules being fired, the descriptions, explanations and results being presented, as well as the comfort and intuitiveness of the user interface. The goal of this session is a (hopefully ever decreasing) list of modifications required before the next review session.

The rules of this session are as follows. The DIG sit at the computer walking through the implemented tasks, checking that the correct questions are being asked, correct input data is
being sought, correct results being returned, and so on. They should check that the model is consistent with the paper knowledge base (as far as possible) and that all modifications from the previous session have been successfully completed.

Eventually, the DIG will actually be ‘testing’ the expert system rather than verifying it, at which point they are asked to provide case studies as input. Once several of these tests have been successful, the system can be released into full user appraisal tests in the field.
5. EVALUATION

5.1 HYPOTHESIS

We present evidence in this chapter to support our belief that our methodology is a good framework for knowledge acquisition. We attempt to show it is an improvement on conventional approaches, primarily due to our utilisation of techniques from the societal and linguistic metaphors described in earlier chapters.

We showed in Chapter 3 how the methodology used in the ELSIE project out-performed those employed for the DIMES and W-RAISA projects by producing more objects, rules and tasks per hour of expert interview. We argued that this improvement was due primarily to the lack of consideration of societal and linguistic issues in the early approaches, and that it could not be accounted for solely by the greater experience of the knowledge engineer (the author) within the domain.

Ford et al. (1993) point out that if as Agnew and Brown (1989a) said ‘reality does not directly reveal itself to us,’ how can we evaluate the adequacy of our knowledge? They concluded that the crucial question for knowledge engineers is not ‘How do we know the model is correct?’, because every model is to some extent an oversimplification, rather, ‘How useful is the model and the modelling process as a means of facilitating our understanding of the domain?’
For the purpose of evaluating our work, our hypothesis will be that 1) our methodology is an efficient and effective modelling process, and that 2) the resultant ELSIE paper knowledge base is a complete and accurate model of the domain.

5.2 Method

We initially wanted to arrange an experiment where several knowledge engineers performed a knowledge acquisition task, with each technique of the methodology and then scored the effectiveness of each. Unfortunately, we had neither the time or resources to arrange such an experiment. However, there are a group of people who have taken part in the employment of all of our techniques, and know better than anyone else how successful or not their application has been. Not only this, but each one of these people is ideally positioned to answer the second question on completeness of the knowledge acquired. Furthermore, this group have proven experience in validation as described in Chapter 4; our evaluators are members of the Domain Interface Group (DIG) from the ELSIE experiments.

The evaluation procedure involved two tasks to be undertaken by each member of the DIG. First, each member was asked to respond to a questionnaire relating to the effectiveness of our tools, methods and techniques, on aspects such as the use of an interface group, brainstorming, interviewing and so on. Appendix A provides a complete list of the questions asked. Second, each member was asked to revise the paper knowledge base (task structure and rules) by amending, deleting or adding tasks and rules, and to score each task and rule for significance on a scale of 1 to 10. We consider the paper knowledge base to be the model
produced by our knowledge acquisition process for evaluation purposes because not all of the knowledge acquired can be encoded during implementation.

We initially planned a brainstorming session between the members of the DIG with the goal of proving or disproving our hypothesis by reaching a consensus score for each question asked and for each rule and task in the paper knowledge base, thus turning our methodology of verification and validation onto itself. Unfortunately, with the project finished, it was impossible to bring the members together, therefore we have used an averaging method as an alternative for consensus. This will, however, provide us with a more useful perception of individual group member’s views: consensus does not take into consideration minority opinions, and perhaps some decisions made by the group as a whole did not reflect a true consensus but a majority vote.

5.3 RESULTS AND DISCUSSION

5.3.1 Questionnaire on the efficiency and effectiveness of our methodology

Completed questionnaires were received from five members of the Domain Interface Group (DIG) over a period of ten months (due to sickness and general availability) after the completion of the ELSIE project. Each participant in the survey answered every question with a numerical value on a scale of 1 to 10 inclusive. Note that the poles of the scale were randomly swapped so that sometimes 1 was the most positive response, not 10, to ensure that the participants had studied the questionnaire carefully and not completed it without due consideration: if this had been the case it would have become obvious during analysis when
responses were contrasted between participants. During analysis, if the poles had been reversed, the scores were converted back (1 = 10, 2 = 9 and so on) so that when statistically and graphically analysed, all scores were on the same scale. All of the actual questionnaire responses are provided in tabular form in Appendix B.

For each set of responses, the mean and standard deviation were calculated and plotted on a bar chart, except for questions relating to expert characteristics where a cumulative total was used. Where no standard deviation is shown its value is zero, which shows that all responses were the same. Also marked on each graph is the value of 5.5: this is the ‘average’ score of 1-10. We argue that any mean score above 5.5 is a positive response to the question, below 5.5 a negative response. Further, if the standard deviation from the mean also falls the same side of this point, then we say that all responses are statistically proven to be positive or negative accordingly.

The first group of questions concerned the use of an interface group to support the knowledge engineer in his/her knowledge acquisition duties. Responses to these questions are illustrated below in Figure 24. The mean response of 8.4 out of 10 to the first question regarding the effectiveness of the DIG was statistically positive, a promising start. The response to the second question was negative, however this is an encouraging result. We asked each member of the DIG how much of the knowledge acquired through the group could they have provided themselves? As the graph shows, the average response was only 4.8, representing ‘less than half’, although the upper point of the standard deviation was slightly above half. On closer investigation, we see that two members of the group felt that they could provide over half of the knowledge, 60% for Member D and 70% for Member E (Figure 25).
The third question asked each participant how long it would take them to provide this knowledge. The mean response of 7.2 out of 10 equates roughly to “many months”. We must note here, however, that this mean is not statistically supported because there is a wide standard deviation, which falls well below the average marked on the scale. Closer investigation shows that this is because one of the group (member E) scored only 2, implying
that he/she could have provided 70% of the acquired knowledge (their response to the previous question) in only a matter of days (see Figure 26).

![Figure 26: How long would each member of the DIG take to provide the knowledge? (days - years)](image)

**Figure 26: How long would each member of the DIG take to provide the knowledge? (days - years)**

The second set of questions related to the coverage and completeness of the knowledge acquired and modelled. Figure 27 illustrates the mean responses to these questions, which we discuss below.

![Figure 27: Questionnaire responses regarding domain coverage of the knowledge](image)
Responses to five of the six questions are encouraging, only question 7 is a negative result. In question 4, an average of 82% coverage of the domain was recorded, statistically supported by the low spread shown. The next two responses show that on average the participants felt that the task structure was 70% complete and the paper rule base 78% complete, again statistically supported, even though the spread is larger.

Question 7 asked the DIG how complete they considered the implemented system. The graph shows an average of only 55% from the respondents with a standard deviation dipping well below that. We note here that only four members of the DIG responded to this question because one member left the group before the system was completed, however we do not believe that this has any bearing on the result. Figure 28 shows the breakdown of responses to this question.

![Figure 28: How much of the domain do you think is covered by the implemented system? (none - all)](image)

The next question asked the members of the group if the knowledge acquired was still current or already out of date. A mean score of 7.4 out of 10 towards still current is encouraging,
particularly as the earliest responses were received over two years after knowledge acquisition for the project ended. Interestingly, the wide standard deviation of the responses is caused by one member scoring only 5 for this question, and further investigation shows that this response arrived 10 months later than the other four.

The final question in this group asked the participants how useful they thought the interviews were at finding a good coverage of knowledge. Another promising score was received here of 7.8 out of 10 on average, with very little standard deviation showing like-minded responses.

The next group of questions related to the importance of the techniques used in our methodology, asking the participants how highly they would recommend the techniques if they were to undertake a similar project in the future. As Figure 29 shows, all techniques would be highly recommended.

![Questionnaire responses regarding the importance of different techniques](image)

**Figure 29: Questionnaire responses regarding the importance of different techniques**

The results were encouraging for all of questions 10 to 13, respectively the recommendation for a DIG (9.6 out of 10), brainstorming (9.6 out of 10), interviewing experts (9.2 out of 10)
and using DIG members to interview the experts (9.4 out of 10), all results statistically supported by a small standard deviation. Interestingly, the lowest score and widest spread is for interviewing experts (Question 12). Structured walkthroughs of early prototypes (Question 14) scored 10 out of 10 from every participant, illustrated by the zero standard deviation, a very supportive response.

Expert systems in the workplace was the topic of the next group of questions, set in an attempt to find out how domain practitioners feel about expert systems helping with their job. It is often considered, rightly or wrongly, that the introduction of systems like ELSIE will inevitably lead to the loss of jobs within a domain. It is therefore important to gauge how well a system will be received if it is to have any chance of acceptance in the workplace. Figure 30 illustrates the responses.

![Figure 30: Questionnaire responses regarding expert systems in the workplace](image)

Question 15 asked how threatened the members of the DIG felt by the possible introduction of an expert system. As the graph shows, there was little fear that ELSIE would cost them their
job, however, it is worth noting that these people had worked on the development of the system and were well aware of its expected role within their domain.

Question 16 is therefore more interesting, asking the participants how many people outside of the DIG did they think felt threatened by the system’s introduction. An average of almost 70% was the response, suggesting that almost three quarters of the work force feel threatened by expert systems. In expectation of such a result, we asked the DIG if they felt that this fear of expert systems led to negative comments about the system during user trials (Question 17), perhaps in an attempt to slow or even stop its introduction. There is a little evidence of this in the responses: although an average of 4.8 out of 10 is negative, there is a noticeable difference between respondents, and one member did reply with a 7 out of 10.

The next group of 8 questions referred to how useful, in retrospect, the DIG members considered our techniques. Figure 31 shows encouraging responses to all questions.

![Figure 31: Questionnaire responses relating to usefulness of our knowledge acquisition techniques](image)
Every score for this set of questions was very high with almost no spread. The first six questions related to the use of brainstorming at different stages in the methodology, all of which had encouraging responses: interview topic selection (10), planning interview contents (9.8), expert selection (10), task base structuring (9.6), paper rule validation (9.6) and paper rule positioning in the task hierarchy (9.8). The remaining two techniques of interviewing and structured walkthroughs also scored very high at 9.6 each.

The next two questions asked how each member of the DIG, in retrospect, considered their consensus choice of interview participants. The responses are illustrated in Figure 32 below.

![Figure 32: Questionnaire responses regarding the choice of interview participants](image)

It is encouraging that all members of the group, with hindsight, are still happy with their choice of experts (scoring 9.2 out of 10) and interviewer (scoring 8.2 out of 10). There is also little standard deviation showing like-mindedness once again.

The final set of questions on tools and related techniques used within our methodology produced four encouraging and one very interesting response, shown in Figure 33 below.
First, each participant was asked how important they considered the professional recording of the interview and how useful they considered the resulting video artefact as documentation of the expert’s knowledge. Both responses were encouraging scoring an average 7 out of 10 and 7.2 out of 10 respectively. It must be noted, however, that the lower standard deviation in the latter falls below the average, caused by Member B scoring only 4 out of 10.

The next two questions asked how useful the DIG found the breaking down of the domain into a task hierarchy and the visualisation of this hierarchy in organisation charting software. Both responses were promising, scoring 9.6 out of 10 each.

The result of the final question is also positive, and therefore encouraging, but is one of the most interesting in the survey. Each member of the DIG was asked how much of the knowledge acquired in the project they considered to be undocumented elsewhere in the domain. The mean response was 64% (6.4 out of 10), however, there was a very large
standard deviation, which on investigation was due to one respondent as illustrated in Figure 34.

![Bar chart showing knowledge documentation](image)

**Figure 34: How much knowledge do the members of the DIG think has been documented in the project that is undocumented elsewhere? (none - all)**

The majority of participants considered the project to be the only source of more than half of the knowledge acquired, however, one member of the DIG considered only 20% of the knowledge to be uniquely documented. Further investigation shows that member D’s response was received 10 months after the last of the other four was received.

Taking a different viewpoint, we also asked each member of the DIG (A - E) to comment on certain characteristics of the experts (1 - 4) used for interviewing during the knowledge acquisition process. The characteristics were charisma, authority, cleverness, knowledge and experience. Unlike the previous results, we show in Figure 35 to Figure 39 the cumulative scores for each expert for each of the five characteristics. This includes a breakdown of each score by DIG member, enabling us to compare personal opinions much easier than in the
previous graphs: wide variances in opinion can be expected here because there is a large element of human nature involved.

Figure 35: DIG's opinion on each expert's charisma

The question of charisma provides some interesting results, the most striking of which is a particularly low score received for Expert 4. The accumulative score for this expert is only 19 out of 50, in comparison to the most charismatic expert (Expert 3) who scored 38 out of 50. Another interesting result is the minimum score achieved by Expert 4 from DIG Member C of only 1 out of 10.

There is very little of interest in the results of the question on authority. All experts scored extremely highly, between 43 and 46 (Expert 1) out of 50.

On the question of cleverness, Expert 1 stands out from the rest as the most clever by scoring 41 out of 50, 7 more than each of the others. A point of interest here is the like-mindedness of each DIG member.
The question of the experts' knowledge was also scored highest for Expert 1, with a total of 42 out of 50. The other experts were again closely scored receiving 38, 37 and 36 out of 50 respectively for Experts 2, 3 and 4. Interestingly, however, the graph does show a wide difference of opinion between members of the group, even though these differences balance out in the cumulative score.
Finally, the question of experience results in Expert 1 yet again scoring higher than the rest, with 42 out of 50 again. Also, there is little once again between the remaining three, scoring 37, 35 and 37 out of 50 for Experts 2, 3 and 4 respectively. There are noticeably more like-minded responses here than in the previous question.
5.3.2 Paper knowledge base audit.

Only one member of the Domain Interface Group returned the audit on the paper knowledge base, (note that the audit takes several hours), and this response had only the rules scored. There are 342 rules in the returned paper knowledge base, the percentage of which for each score out of 10 is shown in the pie-chart the Figure 40 below. The chart shows that the majority of rules (62%) scored 9 or 10 in the audit (36% and 26% respectively). This is an encouraging result, however, it is worth noting that the minimum score of 1 out of 10 is the fourth most popular, accounting for 8% of the rules.

![Pie chart showing distribution of rule scores](image)

**Figure 40: Number of rules per score out of 10 in the paper knowledge base audit**

Figure 41 shows the number of rules added, deleted, edited and unchanged during the paper knowledge base audit. As the graph shows, almost 10% of the paper knowledge base rules were edited by the member of the DIG, a low but not insignificant score. Almost 4% of the returned rules were new, added during the audit, but encouragingly only 1% were deleted, even though, as we saw above, 8% were considered totally insignificant. The most encouraging result from the graph, however, is that the member of the DIG was satisfied with
over 85% of the rules as they were, even though not all of these were considered highly significant.

![Pie chart showing percentage of rules edited, deleted, added, or unchanged.]

**Figure 41: Percentage of rules in the paper knowledge base edited, deleted, added or unchanged during the audit**

It is worth noting here for completeness that after further investigation there is no correlation between the edited rules and any one particular score: as many rules scoring 10 out of 10 for significance were edited as those scoring only 1 out of 10.

### 5.4 Conclusion

Overall response to the questionnaire suggests that our methodology appears to be efficient and effective, as the first part of our hypothesis suggests. Results of over 70% for questions relating to the coverage of the domain by the knowledge acquired (82%), by the task hierarchy (70%) and by the paper rule base (78%), indicate how successful the methodology is at acquiring knowledge.
The extremely high mean scores, all over 9.6 out of 10, achieved for the usefulness of the brainstorming technique in topic selection (10), planning interview content (9.8), expert selection (10), task base structuring (9.6), paper rule validation (9.6) and paper rule positioning (9.6), along with its score of 9.6 out of 10 for recommendation in future projects, suggest that this is a technique which played a vital role in the methodology’s success.

This technique could not have been used, of course, without the introduction of the Domain Interface Group (DIG), the use of which is supported by the score of 8.4 out of 10 for its effectiveness and 9.6 out of 10 for recommendation in future projects. The importance of the DIG is further supported by the results of the questions regarding how much of the knowledge each member of the DIG could have provided themselves and how long this would have taken them. The result that, on average, each member of the group could have provided only 48% of the knowledge acquired, and over a long period of time, is very encouraging. The fact, however, that one member alone (Member E) felt that he/she could provide 70% of the knowledge in a matter of days contradicts these results and suggests that a DIG is not necessary at all, and neither are the experts to any great extent!

The interviewing of experts, the only part of our methodology remaining from the conventional approach, is probably the least well received of the techniques. This said, it is still rated very highly, its importance to future projects scored at 9.2 out of 10 and the DIG stored its usefulness at 9.6 out of 10. The success of the interviews could be attributed to the choice of participants, the selection of experts and interviewers scoring on average 9.2 and 8.2 out of 10 respectively, supporting once again the use of consensus decision making. It is also encouraging to see a positive response to our insistence on the use of a professional recording
studio for the interviews. Although this is a relatively expensive aspect of the methodology, its use is vindicated by the average score of 7.2 out of 10 for the usefulness of the resulting video artefact, even though one member of the DIG (Member B) disagreed.

The role of structured walk-throughs in validating the incremental prototypes is another technique shown to have made an enormous contribution, the only technique, in fact, to score a full 10 out of 10 from every DIG member regarding its recommendation for future projects.

Another technique which has proved a valuable contributor to the methodology is the structuring of the task hierarchy, scoring 9.6 out of 10 on average. It is also encouraging to see a score of 9.6 out of 10 for the use of organisation charting software to view this hierarchy, vindicating our selection of an unconventional tool over more sophisticated concept map based alternatives.

There are few poor results to contradict the hypothesis. The coverage of the domain by the implemented system was poor (only 55% on average), but this is not particularly surprising: as we have stated in previous chapters, the constraints imposed by the expert system shells often mean that all of the knowledge acquired cannot be implemented, a point proven by the much higher coverage by the domain reported for the acquired knowledge in the paper knowledge base (82%), the model we are using for evaluation purposes.

Two interesting results which are highlighted because of less positive responses relate directly to the passing of time between knowledge acquisition and questionnaire completion. The first of these, relating to how current the knowledge acquired is today (that is, when the
questionnaire was completed) shows that one member of the DIG scored only 5 out of 10, compared to the average of 7.4 out of 10 for the group as a whole, meaning that he/she considers the knowledge far more outdated than the other four members. The second question asked the DIG what percentage of the knowledge acquired did they consider to be undocumented elsewhere in the domain. The same member of the DIG felt that only 20% of the knowledge acquired was uniquely documented, an enormous difference compared to an average of over 75% for the other four members.

The only reason for these two differences could be the fact, as noted in the results, that this member's questionnaire was received 10 months after the last of the others. The first response illustrates very clearly how quickly the domain knowledge changes. To explain the second, one can only deduce that a vast amount of new documentation, or perhaps another computer-based system similar to ELSIE, has been introduced into the domain during these 10 months.

The remaining low scores resulted from our questions regarding expert systems in the workplace. These negative responses, however, are good news, because they imply there is little fear among practitioners that the introduction of expert systems would automatically mean the loss of jobs, particularly when the role of such systems has been explained. Such systems should therefore meet little resistance, however the one result of 7 out of 10 (Member D) is worrying, because if this member's colleagues were responsible for user appraisal testing, he/she implies that they may well make negative comments to stop the system being introduced!
An indirect result from all the questions and responses is the extent of like-mindedness between members of the DIG. We can only propose that this is in some way due to the consensus reached amongst the individuals during their time as members of the DIG. There is certainly no evidence to suggest that any minority views were simply overlooked during the DIG sessions, and if not in agreement at the start of the project, the members appear to share common views by the end.

Although the questions relating to expert’s characteristics have no bearing on the proof of our hypothesis, the results are nevertheless interesting. The responses show very clearly that, in retrospect, the DIG made a poor selection of expert in Expert 4 for interviewing purposes. There is no doubt, by considering his scores for the other four characteristics, that this person is indeed worthy of his expert status within the domain. However, the score of 19 out of 50 for charisma, comprising like-minded low scores from every member of the DIG, proves that experience and knowledge alone are not adequate criteria for the selection of experts in knowledge acquisition projects.

The majority of rules (64%) being scored with either a 9 or 10 out of 10, along with 85% of the rules being untouched and very few being added, would suggest that despite the changing domain knowledge, the paper knowledge base is still a fairly complete and accurate model of the domain, as suggested by the second part of our hypothesis.

The fact that 8% of the rules are no longer significant is most likely due to changes in domain knowledge. We note that the average score of 7.4 out of 10 for the question relating to how
current the knowledge still is would imply that 26% of the paper knowledge base could by
now be insignificant.

The poor response of only one DIG member to the paper knowledge base audit makes it
impossible to draw any conclusions on the importance of consensus in the knowledge
acquisition process. Also, it is impossible to know if the DIG would have added or edited
more rules if they had believed that the changes would be made to the implemented system:
lack of enthusiasm once the project has ended is proven by the poor response.

It is interesting that although 8% of the paper rules were considered insignificant, only 1.2%
were deleted, and most of the rest were not edited: the results reporting that there was no
correlation between the 10% that were edited and the low scoring rules. One supposes that
although insignificant, the member of the DIG felt that these rules should for some reason
remain, perhaps in case the domain knowledge changed again, maybe even back to its
previous state.
6. CONCLUSIONS AND FUTURE WORK

6.1 AN EFFECTIVE AND EFFICIENT METHODOLOGY?

We believe that there are four important achievements to be reported from this work, the combination of these going a long way to meeting our initial objectives. The first output of the research is the description of our approach, combining psychological, societal and linguistic metaphors: this, we believe, is a powerful framework for knowledge acquisition. We have put into practice this framework which provides a means for modelling broad domains where geographical and political issues make the expertise more socially situated. We have shown how classical interview techniques lack necessary input from the peer group in which the expert’s knowledge exists, and we have provided a means of capturing this input through the introduction of the Domain Interface Group. This team of domain practitioners support the knowledge engineer in many roles which, classically, had been his/her sole responsibility. The methodology is supported throughout by the use and understanding of domain specific terminology and lexical semantics in all communication with the domain and in the analysis of domain texts and interview transcripts.

The second achievement is a result of the changing role of the knowledge engineer, from primarily a computer scientist to primarily a group facilitator. This may be considered either as deskilling the knowledge engineer or facilitating him/her in his/her complex task. A primary objective of expert system development is the deskilling of the workforce, however, this commonly involves great cost in the use of knowledge engineering expertise for the system’s development. If we have truly achieved deskilling or facilitating the knowledge
engineer, these costs will be reduced, and therefore one obstacle in building experts systems will be removed. We do not suggest that the role of the knowledge engineer is in any way less important, however the guidelines in Chapter 4 could enable less skilled or experienced knowledge engineers to perform the required role.

The third result of our work, again emerging from our use of the societal and linguistic metaphors, is the decrease in, if not alleviation of, reductive bias. This problem, commonly resulting from either the inappropriate application of an analogy that is too simple (analogy bias) or from the use of everyday terms in technical ways without explicitly stating their technical meaning (common connotation bias), has long been caused by the mis-interpretation of domain knowledge by the ill-informed knowledge engineer. The use of automated text analysis tools to process domain specific text during the initiation stage, where the majority of domain terms are elicited, removes any interpretation by the knowledge engineer at conceptualisation. The use of DIG members for the interviewing of domain experts, and the combination of text analysis techniques and consensus amongst the DIG in the subsequent analysis of interview transcripts, allows little opportunity for the knowledge engineer to interpret the knowledge during any stage of the methodology. This should inevitably lead to a more accurate domain model.

The constructivist paradigm, in which we have situated our methodology, advocates the use of mediating representations to support the modelling process. The fourth important output of our research is the combined mediating representation of terminology data banks (term bases) for modelling domain specific concepts, hierarchical tree structures for modelling tasks, and natural language representation of rules. This formalism is expressive, aids communication
between the development team, guides knowledge analysis, makes important things explicit, and is complete and concise, efficiently saying all that needs to be said. This satisfies almost every requirement for an effective mediating representation described in the literature.

We believe that our methodology provides an accurate and working model of the domain whilst keeping down cost and human resources to an affordable level, an accomplishment which is supported by the results of the evaluation, and therefore satisfies our initial objectives.

6.2 Future work

Based on advancements in knowledge acquisition tools, as described in Chapter 2, we believe that our methodology could be automated. This could be initiated through the specification of input and output files for each stage, enabling loose coupling between the various tools and techniques. This would require a knowledge interchange format, sufficiently expressive to represent each of our mediating representation formalisms, and based on well defined semantics which do not assume any particular implementation, such as KIF (Genesereth and Fikes, 1992). Subsequent addition of further tools, perhaps automating the interview and then the brainstorming sessions, and improving the communication protocols between the tools and between mediating representations, could lead towards a “knowledge medium for collaborative development”, as described in Gruber et al. (1992).

We described, in Chapter 3, a number of complications in the use of a conventional term base. Research on the issues addressed is progressing in the field of terminology management,
however, we feel that it would be useful if more collaborative research was undertaken by the terminology and knowledge engineering communities. Of particular importance, and of most benefit in the past, is research regarding the use of term bases as a mediating representation for domain conceptualisation. Similarly, collaboration between knowledge engineers and researchers in text analysis and information retrieval, could also increase even further the automation of knowledge acquisition from text.

Another part of our methodology that we consider worthy of further investigation is the process by which consensus is achieved within the Domain Interface Group (DIG). This study, into the group dynamic, could be achieved, perhaps, through the interpretation of brainstorming results, but more interestingly, through the examination of the process leading to these results, in particular the modifications to the white board or flip chart contents reflecting the shifting ‘state’ of the knowledge model over time.

Finally, requirements for enhancement and improvement of the methodology will only be identified through its extensive use. Initially, this will be within the knowledge acquisition research community, in particular the AI group at the University of Surrey, where components are already being used in other knowledge engineering projects. However, the more pragmatic problems of the ‘real world’ will require more inventive solutions, and therefore more interesting enhancements are likely to be discovered through the application of the methodology within the IT community at large, the role for which it was intended.
REFERENCES


Ahmad, Khurshid and Collingham, Steven. (1996).


Ahmad, Khurshid and Davies, Andrea. (1994).


Ahmad, Khurshid and Griffin, Stephen M. (1994).  

Ahmad, Khurshid and Rogers, Margaret. (1992)  

Ahmad, Khurshid, Davies, Andrea, Fulford, Heather and Rogers, Margaret. (1992a).  
What is a term? The semi-automatic extraction of terms from text. In M. Snell-Hornby, F. Pöchhacker, K. Kaindl (Eds.), *Translation Studies: An Interdiscipline.* Amsterdam: John Benjamins Publishing Company. pp 267-278.

Ahmad, Khurshid, Davies, Andrea and Rogers, Margaret. (1993).  


Ahmad, Khurshid, Holmes-Higgin, Paul, Rogers, Margaret, Le Hong, Khai and Huwig, Christine. (1993).  
An Expert System for Offshore Inspection and Maintenance. In B. Topping (Ed.),
Artificial Intelligence Techniques and Applications for Civil and Structural Engineers.


Educational Psychology: A Cognitive View (Second ed.). New York: Holt, Reinhart and 
Winston.

Personal construct theory and the transfer of human expertise. Proceedings of the 
National Conference on Artificial Intelligence (AAAI-84), Austin, Texas.


Journal, 1.

Boose, John R. (1992) 
Knowledge Acquisition. In S.C. Shapiro (Ed.), Encyclopaedia of Artificial Intelligence 

Boose, John H. and Bradshaw, Jeffrey M. (1987). 
Expertise transfer and complex problems: Using Aquinas as a knowledge-acquisition 
workbench for knowledge-based systems. International Journal of Man-Machine Studies, 

Better group decisions: Using knowledge acquisition techniques to build richer decision 
models. Proceedings of the Knowledge Acquisition for Knowledge-Based Systems 
Workshop (KAW-92), Banff, Alberta: Canada. pp. 2.1-2.22.

Journal, 5, (4).

Knowledge-based design rationale capture: Automating engineering trade studies. In M. 
Boose, John H., Shema, David B. and Bradshaw, Jeffrey M. (1989).

Bradshaw, Jeffrey M. (1988).


KS-3000: An application of DDUCKS to bone-marrow transplant patient support. Sixth Annual Florida AI Research Symposium (FLAIRS '93). Ft. Lauderdale, FLorida, USA.


Bradshaw, Jeffrey M., Holm, Peter D., Kipersztok, Oscar and Nguyen, Thomas. (1992c).


In D.S. Moralee (Ed.), Research and Development in Expert Systems IV. Cambridge:

Mapping the Dynamics of Science and Technology. UK: MacMillan.

Intelligent support for collaborative modeling. Proceedings of the Fifth Annual Florida

Participant systems. Calgary, Alberta: Canada: Alberta Research Council Advanced
Technologies.

The Knowledge Level Reinterpreted: Modelling Socio - Technical Systems. In K. M.

Crane, Diana (1972).
Invisible Colleges: Diffusion of Knowledge in Scientific Communities. University of
Chigago Press.

Cruse, Diana A. (1986)
Lexical Semantics. Cambridge: Cambridge University Press.

Cruse, Diana A. (1992)
Antonymy Revisited: Some thoughts on the relationship between words and concepts. In
p.289-306.

A methodology for integrating knowledge acquisition techniques. Proceedings of the
First European Workshop on Knowledge Acquisition for Knowledge-Based Systems.
Reading, England.

Diederich, Joachim, Ruhmann, Ingo and May, Mark. (1987).
KRITON: A knowledge acquisition tool for expert systems. International Journal of

Knowledge acquisition and constructivist epistemology. In R. R. Hoffman (Ed.), The
Cognitive Psychology of Expertise: Empirical Approaches to Knowledge Acquisition.
New York: Springer-Verlag.
Special Issue on knowledge acquisition as modelling: Parts I and II. *Int. J. of Intelligent.
Syst*, 8, (1, 2).

Knowledge Acquisition as a Constructive Modelling Activity. In K. M. Ford and J. M.

Explanation as a knowledge acquisition issue. *Proceedings of the Fifth Annual Florida AI
Research Symposium FLAIRS-92*, Ft. Lauderdale, Florida, USA.

An approach to knowledge acquisition based on the structure of personal construct
systems. *IEEE Transactions on Knowledge and Data Engineering, 3*, pp. 78-87.


Knowledge acquisition systems for rapid prototyping of expert systems. *Infor, 26*, (4),
pp. 256-285.

Design requirements for knowledge support systems. *Proceedings of the Fourth
Knowledge Acquisition for Knowledge-Based Systems Workshop*. Banff, Alberta, Canada.

Empirical investigation of knowledge representation servers: Design issues and

New directions in the analysis and interactive elicitation of personal construct systems. In
M. L. G. Shaw (Ed.), *Recent Advances in Personal Construct Technolog*. New York:

151-190.

Cognitive and logical foundations of knowledge acquisition. *Proceedings of the Fifth
Banff Knowledge Acquisition for Knowledge-Based Systems Workshop*. Banff, Canada.
Eliciting knowledge and transferring it effectively to a knowledge-based systems. *IEEE Transactions on knowledge and Data Engineering.* 5, (IV), pp. 4-14.


Knowledge Engineering for Expert Systems, Chichester, UK: Ellis Horwood Ltd.


Archiving Knowledge: Before and After the Interview. In D. Lukose (Ed.), *Proceedings of the Workshop on Knowledge Acquisition Using Conceptual Graph Theory, International Conference on Conceptual Structures.* Maryland, USA.


Hinkle, David N. (1965).


Building Large Knowledge-Based Systems: Representation and Inference in the CYC Project. Menlo Park, CA, USA: Addison-Wesley.


Shaw, Mildred L. G. and Chang, Ernest. (1986).


Modelling expert knowledge. Knowledge Acquisition, 2, pp. 179-206.

The Ethnographic Interview. Holt, Rinehart and Winston (Eds.), New York.

Summers, Della. (1991)


Waterman, Donald A. (1986)
A Guide to Expert Systems. Addison-Wesley, Reading: MA, USA.


APPENDIX A: EVALUATION QUESTIONNAIRE

ELSIE’s Knowledge Acquisition Methods

Domain Interface Group

1. How effective do you consider the use of the domain interface group (the project steering committee plus the knowledge engineer)?

1=not effective 10=very effective

2. How much of the knowledge acquired by the domain interface group could you have provided alone?

1=none 10=all

3. How long do you think it would have taken you to provide the knowledge alone?

1=years 10=days

Coverage

4. How much of abstraction licensing do you think is covered by the knowledge acquired?

1=very broad coverage 10=very poor coverage

5. How complete do you consider the task structure to be?

1=far from complete 10=complete

6. And the paper rules?

1=complete 10=far from complete

7. And the implemented system?

1=far from complete 10=complete

8. How out of date do you think the knowledge acquired is now?

1=totally out of date 10=still current

9. How successful do you consider the interview’s were at finding knowledge?

1=complete failure 10=a total success
Importance of different methods

10. If you were to undertake a similar project, how highly would you recommend the use of a domain interface group?

1=would not recommend 10=would insist on

11. And brainstorming domain practitioners (whether in a domain interface group or not)?

1=would insist on 10=would not recommend

12. And interviewing experts?

1=would not recommend 10=would insist on

13. And using domain practitioners to interview the experts?

1=would insist on 10=would not recommend

14. And structured walkthroughs of early prototypes?

1=would not recommend 10=would insist on

Expert systems in the workplace

15. During the user trials, how threatened did you feel by the possible introduction of a system like ELSIE?

1=very threatened 10=not at all threatened

16. Did people outside of the domain interface group feel at all threatened by the possible introduction of a system like ELSIE?

1=no, none at all 10=yes, all of them did

17. Do you feel that anyone who saw the system made negative comments because of such a threat?

1=yes, definitely 10=no, not at all
ELSIE's Knowledge Acquisition Techniques

In retrospect, how useful do you consider...

1. The brainstorming sessions of the domain interface group for interview topic selection?
   1 = waste of time 10 = necessary

2. And planning interview contents?
   1 = necessary 10 = waste of time

3. And expert selection?
   1 = waste of time 10 = necessary

4. And task base structuring?
   1 = necessary 10 = waste of time

5. And paper rule validation?
   1 = waste of time 10 = necessary

6. And paper rule positioning in the task hierarchy?
   1 = necessary 10 = waste of time

7. And interviews with experts?
   1 = waste of time 10 = necessary

8. And structured walk-throughs of the knowledge base?
   1 = necessary 10 = waste of time

In retrospect, how good do you consider our choice

9. Of experts?
   1 = excellent 10 = poor

10. And interviewers?
    1 = poor 10 = excellent
ELSIE’s Knowledge Acquisition Tools

1. How important do you feel the professional recording of the interviews was?
   1 = waste of time 10 = necessary

2. How useful do you consider the videos as a documentation of the expert’s knowledge?
   1 = very useful 10 = waste of time

3. How useful do you consider is the breaking down of the domain into a task hierarchy?
   1 = waste of time 10 = necessary

4. How useful do you consider the visualisation of the hierarchy in a chart?
   1 = necessary 10 = waste of time

5. How much knowledge do you think has been documented in the project that is not documented elsewhere?
   1 = none at all 10 = all of it
ELSIE Experts

Please answer each of the following questions:

1. For each expert, please state on a scale of 1 to 10 how charismatic a person you thought they would be whilst interviewed (1 is not charismatic, 5 is OK and 10 is very charismatic):

<table>
<thead>
<tr>
<th>Expert 1</th>
<th>Expert 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Expert 3</td>
<td>Expert 4</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. For each expert, please state on a scale of 1 to 10 how authoritative a person you thought they would be whilst interviewed (1 is very authoritative, 5 is OK and 10 is not authoritative):

<table>
<thead>
<tr>
<th>Expert 1</th>
<th>Expert 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Expert 3</td>
<td>Expert 4</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. For each expert, please state on a scale of 1 to 10 how 'clever' you think he is (1 is average, 5 is bright and 10 is very bright):

<table>
<thead>
<tr>
<th>Expert 1</th>
<th>Expert 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Expert 3</td>
<td>Expert 4</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. For each expert, please state on a scale of 1 to 10 how the expert's depth of knowledge compares with your own (1 is much more, 5 is about the same and 10 is much less):

<table>
<thead>
<tr>
<th>Expert 1</th>
<th>Expert 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Expert 3</td>
<td>Expert 4</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. For each expert, please state on a scale of 1 to 10 how the expert's experience compares with your own (1 is much less, 5 is about the same and 10 is much more):

<table>
<thead>
<tr>
<th>Expert 1</th>
<th>Expert 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Expert 3</td>
<td>Expert 4</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX B: EVALUATION RESULTS

ELSIE's Knowledge Acquisition Methods

Domain Interface Group

<table>
<thead>
<tr>
<th>Question No.</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>10</td>
<td>8</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>9</td>
</tr>
</tbody>
</table>

Coverage

<table>
<thead>
<tr>
<th>Question No.</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>9</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>9</td>
</tr>
</tbody>
</table>

Importance of different methods

<table>
<thead>
<tr>
<th>Question No.</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>10</td>
<td>10</td>
<td>8</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

Expert systems in the workplace

<table>
<thead>
<tr>
<th>Question No.</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>16</td>
<td>5</td>
<td>6</td>
<td>8</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>17</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>3</td>
</tr>
</tbody>
</table>
### ELSIE’s Knowledge Acquisition Techniques

**In retrospect, how useful do you consider...**

<table>
<thead>
<tr>
<th>Question No.</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>10</td>
<td>8</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

**In retrospect, how good do you consider our choice of...**

<table>
<thead>
<tr>
<th>Question No.</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>8</td>
</tr>
</tbody>
</table>

### ELSIE’s Knowledge Acquisition Tools

<table>
<thead>
<tr>
<th>Question No.</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>6</td>
<td>9</td>
<td>2</td>
<td>7</td>
</tr>
</tbody>
</table>

### ELSIE Experts

#### Charismatic

<table>
<thead>
<tr>
<th>Experts</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

#### Authoritative

<table>
<thead>
<tr>
<th>Experts</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>
### Cleverness

<table>
<thead>
<tr>
<th>Experts</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>

### Knowledgeable

<table>
<thead>
<tr>
<th>Experts</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>5</td>
</tr>
</tbody>
</table>

### Experience

<table>
<thead>
<tr>
<th>Experts</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>6</td>
<td>7</td>
<td>10</td>
<td>5</td>
</tr>
</tbody>
</table>