DOCTOR OF PHILOSOPHY

A knowledge-based systems approach to agroforestry research and extension.

Walker, Daniel Harmen

Award date:
1994

Awarding institution:
Bangor University

A KNOWLEDGE-BASED SYSTEMS APPROACH TO AGROFORESTRY RESEARCH AND EXTENSION

Daniel Harmen WALKER

BSc (Hons) Edinburgh

School of Agricultural and Forest Sciences
University of Wales, Bangor, UK

March 1994

A thesis submitted in candidature for the degree of
Philosophiae Doctor
of the University of Wales

SUMMARY OF THESIS

Surname	**Walker**
Other names	**Daniel Harmen**
For degree of	**PhD**
Title of thesis	**A knowledge-based systems approach to agroforestry research and extension**

SUMMARY

Agroforestry development programmes frequently rely on knowledge from a number of different sources. In particular, there is a growing recognition amongst development professionals of the value of augmenting partial scientific and professional understanding with the detailed knowledge held by local people. Taking advantage of the complementarity of local, scientific and professional knowledge demands the development of effective mechanisms for accessing, recording and evaluating knowledge on specified topics from each of these sources.

The research described in this thesis developed a methodology for the acquisition, synthesis and storage of knowledge. The defining feature of the approach is the explicit representation of knowledge. This is achieved through the application of knowledge-based systems techniques. AKT2 (Agroforestry Knowledge Toolkit), a software toolkit developed in Prolog, an artificial intelligence programming language, provides the user with an environment for the creation, storage and exploration of large knowledge bases containing knowledge on a specified topic from a range of sources. The use of diagramming techniques, familiar to ecologists and resource managers through systems analysis, provides an intuitive and robust interface.

This knowledge-based system drives incremental knowledge acquisition based on an iterative evaluation of the knowledge bases created. The iterative approach to knowledge acquisition provides a coherent, consistent and comprehensive, and therefore more useful, record of knowledge.

Once created, knowledge bases can be maintained and updated as a record of current knowledge. Techniques for the exploration and evaluation of the knowledge base may be useful in :

- giving research and extension staff access to a concise and flexible record of the current state of knowledge;
- providing a resource and mechanisms for use in planning and prioritising research objectives; and
- providing a resource and mechanisms for the generation of extension materials tailored to the needs of particular clients.

ACKNOWLEDGEMENTS

The research presented in this thesis was undertaken as a component of a collaborative research programme entitled "Formal representation and use of indigenous ecological knowledge about agroforestry". Core funding for this programme was provided by the UK Overseas Development Administration Forestry and Agroforestry Research Strategy as projects R4594 (April 1990 to September 1991) and R4731 (October 1991 to September 1994).

The research described in this thesis was supervised by Fergus Sinclair of the School of Agricultural and Forest Sciences of the University of Wales, Bangor, who also led the research programme as a whole.

Software development was undertaken in collaboration with Gill Kendon. Gill Kendon works under the supervision of David Robertson and Mandy Haggith of the Department of Artificial Intelligence of the University of Edinburgh and Robert Muetzelfeldt of the Institute of Ecology and Resource Management of the University of Edinburgh. Robert Muetzelfeldt undertook the implementation of LOCUS and prototype development of software for the diagrammatic representation of knowledgement (see Appendix A).

The research undertaken provided research support for the postgraduate students ('study fellows') participating in the programme (Gamini Hitinayake, Nishantha Jinadasa, Felician Kilahama, Louise Garde, Alison Southern, Pornchai Preechapanya and Balaram Thapa). This both informed and focused the conceptual and software developments described.

This development was given further impetus by collaboration with Fergus Sinclair and Bianca Ambrose (School of Agricultural and Forest Sciences, University of Wales, Bangor)) and Laxman Joshi and Balaram Thapa (Pakhribas Agricultural Centre, Nepal) on an eight month (January to August 1993) pilot phase research programme entitled 'Use of a knowledge base systems approach to the improvement of tree fodder resources on farmland in the eastern hills of Nepal : Pilot Phase' funded by the ODA Forestry and Agroforestry Research Strategy (R5470).

To Lynn and Pia

Contents

LIST OF FIGURES AND TABLES

CHAPTER 1 LOCAL KNOWLEDGE AND AGROFORESTRY

1.1 INTRODUCTION

The objective of the research described in this thesis was to develop an approach to making more effective use of existing knowledge about the ecology of agroforestry practices in supporting target communities in improving the productivity and sustainability of their land use systems. Particular emphasis was placed on the integration of the ecological knowledge held by farmers into development programmes.

The justification for this objective is outlined in this chapter and can be summarised as follows.

- Ecological explanation of environmental phenomena plays a key role in the indigenous management of agroforestry practices.
- On this basis, improving the ecological knowledge used by farmers in managing agroforestry practices is a powerful means of facilitating the development of those practices.
- The results of scientific investigation into the ecology of agroforestry may often be relevant to the farmers. However, scientific knowledge about agroforestry is incomplete.
- While agroforestry research aims to produce precise and predictive tools for the management of agroforestry practices, this development is seriously constrained by the methodological problems encountered in undertaking research in agroforestry.
- As a result, although scientific knowledge provides a powerful means of improving the sustainability, stability and yield of agroforestry practices, it cannot be expected to provide sufficiently robust and detailed information to enable the farmer to make optimum decisions in all aspects of the management of the ecology of an agroforestry practice.
- By contrast, local knowledge may be applied by farmers in relation to many of the detailed aspects of the management of agroforestry practices which have not, as yet, been investigated by science. So, local ecological knowledge may have a useful role to play in furthering scientific understanding about agroforestry as well as providing a basis for the farmer's management decisions. The use of existing local knowledge and participation of

the farmer in research might result in more effective and more efficient development activities.

- Because the fundamental motivations and processes involved in the advancement of indigenous and scientific knowledge are neither significantly different nor incompatible, effective synthesis of scientific and local knowledge should be possible. This may best be achieved through the development and application of a rigorous and explicit approach to the articulation, evaluation and synthesis of knowledge from different sources.

In the following sections, it is proposed that the synthesis of local and scientific knowledge provides a powerful means of enhancing the efficacy and efficiency of development activities. The nature of agroforestry as a development activity is discussed (§1.2), with a particular emphasis on the extension of information to farmers in order to enhance indigenous change. The varying roles of local knowledge in this process are discussed (§1.5), as is the validity and utility of documenting and evaluating local knowledge and combining it with scientific knowledge.

1.2 AGROFORESTRY

Although none are completely adequate, definitions of agroforestry abound in recent literature.

Agroforestry certainly refers to a category of integrated land use practices. These practices involve the deliberate spatial or temporal combination of trees or shrubs with managed herbaceous crops and / or livestock in such a way that there are significant ecological interactions between the woody and non-woody components (Lundgren, 1987). Throughout this thesis these are referred to as agroforestry practices.

Agroforestry has additionally been described as an interdisciplinary approach to the scientific study and management of agroforestry practices, embracing a systems focus and considering the interactions of people, trees, herbaceous plants and livestock within a production system (Sinclair *et al.*, 1994) in terms of productivity, yield stability, resilience, sustainability, fitness for purpose and equitability (Conway, 1987; Marten 1988).

However, this definition does not capture the range of use of the term which, in practice, describes the activities of a loose association of scientists, rural resource managers and development professionals worldwide.

For some, the primary focus is on the study of agroforestry practices, for example Sanchez (1987), Nair (1989) and Monteith *et al.*, (1991). For others, agroforestry refers to the management of agroforestry practices, for example MacDicken (1990). For yet others, agroforestry is primarily a development activity (for example, Scherr, 1990; Getahun, 1990). Amongst this group there are further distinctions. For some, agroforestry as a development activity is technological, the application of scientific understanding to the development of improved land use based on agroforestry practices. For others, agroforestry as a development strategy 'borders on philosophy' (Budd *et al.*, 1990) with strong ideological overtones, for example Mollison(1990) and Agroforestry News.

From any of these perspectives, the ultimate aim of agroforestry as a discipline or approach is to help farmers in maintaining or improving the stability, sustainability and productivity of their land use systems through analysis of the potential impact of the introduction or maintenance of agroforestry practices.

Agroforestry has attracted significant professional interest since the early 1970s and is widely regarded as providing a valuable set of options in providing sustainable, stable and productive systems for rural communities growing in size and aspiration (Pickering, 1988; MacDicken and Vergara, 1990). Interest has been sufficient to generate significant financial and institutional commitment to agroforestry. Between 1978 and 1987 the World Bank lent US $ 750 million for agroforestry projects, indeed, World Bank lending in 1987 represented 37% of total forestry investment (Spears, 1987).

1.3 AGROFORESTRY AS A DEVELOPMENT ACTIVITY

The objectives of a farmer have been summarised as maintaining or increasing productivity, stability (i.e. reducing the variability of yield) and sustainability (Conway, 1987). Agroforestry as a development activity involves helping the farmer in the pursuit of these objectives.

From a farming systems research and extension perspective (Hildebrand, 1990), development programmes can be characterised as an iterative cycle of activities comprising the identification of constraints on the target population in improving their quality of life, the generation of means for overcoming these constraints, the implementation of these means and the evaluation of the implementation. The means identified may include, for example, legislation encouraging and restricting activities, financial incentives and disincentives, direct resource inputs, programmes to mobilise or motivate target populations and, of particular relevance to this discussion, technological development and extension. The latter primarily involves facilitating improvement in land management and, thereby, the stability and/or productivity of land use practices, by disseminating the results of experimental research to farmers. These results may include physical inputs (for example improved crop or livestock strains, chemical inputs or machinery), technology packages (i.e. prescriptive packages which, in principle, require limited understanding on the part of the farmer for successful application, for example alley cropping, Kang and Wilson, 1987) and knowledge. Research generates knowledge about components and interactions in farming systems and their management. This knowledge, if successfully communicated to farmers, may be used by them in decision making and may thereby result in more effective management of an agroforestry practice.

It is suggested that understanding the role and impact of the extension of knowledge is best achieved through consideration of the incremental adaptation of an agroecosystem to a changing environment (in the broadest sense of the term) which is implicit in the application of a systems perspective to land management (Jones and Street, 1990).

Adaptation will occur both in response to a changing environment (including changing aspirations and objectives) and as a result of a changing understanding of the functioning of a land use practice and its interaction with its environment. Development may be achieved by altering the environment in which a practice is managed (for example, by providing irrigation, influencing markets or providing a ready source of improved cultivars) or by improving the farmers' understanding of the functioning of the practice.

It is appropriate for efforts to improve the decision making of a target community to be undertaken in the light of an understanding of the nature of the process of adaptation. Knight (1980) outlines a model of human societies interacting with their environment to produce food that captures this farm-level adaptation. This is based on three premises.

- A person pursuing a production system must be able to monitor changes and act accordingly. So perception and cognition lead to behavioural changes from within (innovation) and from without (diffusion).

- There must be processes of selection against society and the environment to test potential adaptations.

- There must be a memory in the system.

Knight's model suggests that production systems tend to evolve conservatively, incremental changes in the environment leading to incremental adaptation in production practice through adjustment of the knowledge system. There is evidence to support this view (Rindos, 1984). Recognition of the incremental, conservative, evolution of production systems is important in identifying appropriate scales of innovation generated by development projects (for example, the adaptation of existing systems by contrast to the introduction of technologies or practices as a 'package' designed by researchers). Extension activities may seek to completely replace existing practice or to build on and modify existing practice. Experience suggests that building on existing practices ultimately provides the more successful route. In the same way, extension may seek to replace or to build on existing knowledge. If it is accepted that development activities should seek to maintain or enhance the role of the farmer as decision maker, it seems reasonable to suppose that it is more appropriate for extension activities to build on existing knowledge. From this perspective, development activities must start and end with the farmer (Chambers *et al.*, 1989). This demands :

- an accurate evaluation of farmers' current knowledge, particularly its inadequacies;

- an ability either to generate new knowledge through research or to access and collate existing knowledge not available to the farmer, that can be shown, through evaluation of the farmers' existing knowledge, to usefully augment decision making;

- an ability to effectively deliver this knowledge to the farmer.

The identification of knowledge that can usefully augment decision making will demand not only the identification of gaps in farmers' existing knowledge, but also an assessment of their ability to integrate new knowledge into their decision making. This integration depends not only on their ability to understand new knowledge and the implications of that knowledge but also on the opportunity to innovate on the basis of that new knowledge.

Traditional agriculturists are often conservative and cautious in the face of change (Binswanger, 1980). Existing techniques, practices and systems tend to be rational and well adapted to circumstances (Rindos, 1984). Deviation from this accepted norm through experimentation and non-conformity has an associated risk (Barlett, 1980). So, adherence to ritual and established method may be associated with a feeling of security. However, Johnson (1972) argues that farmers will reject conformity and associated security and actively, though cautiously, confront uncertainties in order to comprehend and, ultimately, master them. Johnson (1972) cites evidence for the existence of *"individual differences in agricultural practice and systematic experimentation within the traditional agricultural communities"* and asserts that *"Such individuality and experimentation are probably pervasive in traditional societies, and must be seen as an essential component of their adaptive processes...."* There is, indeed, widespread evidence that agricultural production systems are rarely static in their detailed structure (Soemarwoto, 1987).

Given that agroecosystems constantly undergo adaptation, albeit conservative adaptation, it is reasonable to suppose that they evolve as a result of the combination of prediction, based on experience and understanding, of the consequences of management decisions, experimental evaluation of predictions and experimental exploration of relationships. Knight (1980) talks of a number of 'screens' through which possible production practices must pass to become implemented. Taking this analogy, two fundamental screens are applied in managing change. The first is the prediction of the consequences of particular management strategies and the elimination of inappropriate options accordingly. The second is the experimental or practical evaluation of the remaining set of solutions. In this model, change comes about by incremental adaptation of existing practice as a result of assessment and modification. Assessment might show that the system has become out of tune with an ever-changing world or that it falls short of a goal and can be improved. The resulting modification will be based on a reasoned prediction of the optimum solution or range of solutions and experimental verification of this prediction. Prediction, in turn, may be based on explanatory ecological understanding of mechanisms (mechanistic knowledge) and those based on non-explanatory experience (empirical knowledge). Mechanistic knowledge can be used to model the real world. Equally, empirical knowledge can provide a basis for empirical, predictive, theory. Mechanistic knowledge that facilities accurate prediction is difficult to generate, but the more mechanistic the understanding the more flexibly it can be used.

Improving the ability of farmers to manage agroforestry practices by improving the knowledge on which this decision making is based can result in efficient and effective adaptation of land use practices to a changing world. The extension of new knowledge demands some understanding of 'gaps' in the existing knowledge of the target community and the means by which new knowledge may be integrated with existing knowledge.

Scientific research - although an important means of improving the farmers' information upon which farmers make decisions - cannot be relied upon to address all the pertinent issues over the next decade (§1.5). As a result, complementary sources of knowledge are considered in §1.5, with particular reference to the characterisation and evaluation of farmers' knowledge.

Before discussing different sources and types of knowledge relevant to agroforestry it is necessary to establish a suitable terminology and explore the distinctions implied by this terminology.

1.4 TERMINOLOGY

Knowledge is considered to be distinct from data, information and understanding. For the purposes of clarity these terms are used in this thesis as follows :

- data - a recorded set of observations (which may be quantitative or qualitative);
- information - data that have been processed in such a way as to make them more tractable for interpretation and use;
- knowledge - the outcome, independently of the interpreter, of the interpretation of data or information;
- understanding - the outcome, for the interpreter, of the interpretation of data, information or knowledge.

Three types of knowledge are distinguished : scientific, local and professional knowledge. These are defined as :

- scientific knowledge - documented knowledge generated through the application of scientific method (Davies, 1968);

- local knowledge - the knowledge held by a community as defined by locality;
- professional knowledge - the knowledge that is derived from formal and recognised education or training.

These categories are not mutually exclusive.

'Local knowledge' is used here in preference to the more widely used term 'indigenous knowledge'. Indigenous knowledge has been loosely and variably used in the literature (Brokensha *et al.*, 1980, deliberately avoid definition) and is frequently synonymous with local knowledge. However, where strictly applied, it refers to the knowledge unique to a particular culture or society (Warren, 1991). This strict definition excludes much important knowledge held by farmers. In this thesis the term indigenous knowledge is defined as the culturally specific knowledge held by members of a culturally defined community and only used accordingly.

With the exception of this particular distinction, the term local knowledge is considered, in this thesis, to be synonymous with the range of similar terms encountered in the literature, including 'ethno-science', 'ethno-botany', 'folk-medicine', 'folk-ecology' (Barker *et al.*, 1977), 'indigenous technical knowledge' (IDS, 1989), 'folk science' (Hunn, 1975), 'folk taxonomy' (Berlin *et al.*, 1968), and 'peoples' science' (Richards, 1985).

Further important distinctions are made between explanatory knowledge and empirical knowledge:

- explanatory knowledge - knowledge providing a mechanistic (often predictive) explanation of a phenomenon;
- empirical knowledge - knowledge enabling the prediction, but not the explanation, of phenomena;

and between ecological knowledge and technical knowledge :

- technical knowledge - knowledge about how to achieve particular objectives;
- ecological knowledge - descriptions of organisms or groups of organisms and their interactions with their biotic and abiotic environments, and explanation of those interactions.

These definitions may be superimposed upon one another, so there is an important distinction in the literature between, for example, indigenous technical knowledge (e.g. IDS, 1989) and indigenous ecological knowledge (Woodley, 1991; Chandler, 1990,1991).

1.5 LOCAL KNOWLEDGE AND AGROFORESTRY

Much scientific investigation of the ecology of agroforestry is intended to produce knowledge relevant to farmers in managing agroforestry practices (Gholz, 1987). Indeed, research agendas in agroforestry indicate that future activities, particularly where process based, will be increasingly productive of accurate, predictive understanding of interactions within agroforestry practices relevant to farmers (Anderson and Sinclair, 1993). Nevertheless, research has yet to address many of the issues pertinent to the improvement of agroforestry practices (Nair, 1990). It is proposed that the incorporation of local knowledge into research activities may provide a powerful means of enhancing the efficacy and efficiency of those activities.

There has been a growing interest in the role of indigenous or local knowledge in development over recent years (Brokensha, 1980; Warren *et al.*, 1989) resulting from a diversity of motivations (Tick, 1993).

Three principal roles for local knowledge in the development process can be identified :

- the application of local knowledge through the participation of local experts in development activities;
- the investigation of local knowledge to improve development professionals' understanding of target communities; and
- the integration of local and scientific knowledge to produce a more complete and useful resource base for use in the development of agroforestry solutions to land use problems.

These roles differ in that they demand an increasingly complete, explicit and rigorous synthesis of local and scientific knowledge within development activities.

(i) Local participation and research

The integration of target communities into the research process is being actively espoused and implemented by growing numbers of development professionals (Richards 1985;, Chambers *et al.*, 1989; Amonor 1989). The conservative but incremental nature of indigenous improvement of farming systems and the nature of this process have already been discussed (§1.3). Rocheleau (1987) shows that local experts can often adapt new methods to local conditions better than outsiders, that local control of change often matters more than rapid transformation, and that new technologies are often most successful when built upon existing practice and knowledge.

Furthermore, it has been argued that local participation can complement scientific manpower: local people, as competent observers can extend the range of data channelled to the scientific decision maker (Howes and Chambers, 1980).

(ii) Fuller understanding of target communities

Traditionally, communication in the development process has been one way, from the researcher to the farmer. However, Meehan (1980) argues that although this model of extension is suitable for the environment in which it has evolved, the U.S. and Europe, the one-way flow of information is inappropriate where there is a more serious element of cross-cultural translation. On this basis, traditional extension is recognised as an untenable approach to development. This has led to developments in approaches to farming system research (Byerlee *et al.*, 1982; Hildebrand, 1990; Biggs, 1985; Simmonds, 1985) and to the development of rapid rural appraisal methodologies (Carruthers and Chambers, 1987) such as diagnosis and design (Raintree, 1987b).

These techniques are primarily aimed at better understanding the needs of farming communities and the constraints under which they operate. Resulting experience casts doubt on the prevailing view that the primary requirement in agroforestry is for adaptive research, leading to development of technology packages that are then extended to farmers (Beer *et al.*, 1990; van den Hoek and Bekkering, 1990). The extension of prefabricated technology packages has proved less effective in agroforestry development than incremental improvements that build on local practice (Buck, 1990; Bunderson *et al.*, 1990). The need to synthesise research outputs into technological packages (e.g. the design phase in Diagnosis and Design - Raintree, 1990), has not been justified by field experience. In fact, there is evidence that farmers who do not adopt an integrated package may nevertheless be interested in elements of that package (Buck, 1990). A need to support farmers in decision making about the incorporation of trees within their farming systems based on the results of more fundamental research has been identified as a result (Anderson *et al.*, 1993).

This recognition of the need to provide support for the farmer appropriate for the farmer's role as the decision maker, demands means of assessing the ecological knowledge that farmers use in their decision-making processes, and the knowledge that farmers hold about how to make decisions on the basis of that ecological knowledge. Simply describing actual practice and the constraints contributing to actual practice is inadequate: an understanding of the knowledge used to generate that practice is also necessary. The investigation of local knowledge is not reliant on that knowledge being 'accurate' (Mathias-Mundy *et al.*, 1991). Understanding local knowledge systems can expose inaccurate or technically outdated components and allow the identification of complementary external knowledge (Rochleau, 1987), allowing extension to be targeted accordingly (Gladwin, 1980).

(iii) Combining local and scientific knowledge

Given that farmers have much more intimate experience of their production practices than external professionals, it is reasonable to assume that the people who have been operating them have developed some understanding of them. The knowledge held by farmers can provide a resource for science irrespective of the participation of the farmers in the use of that knowledge. Investigating local knowledge may be a powerful and efficient means of rapidly filling gaps in scientific understanding about agroforestry. So, amalgamating specific local knowledge and general scientific knowledge may be more powerful in designing appropriate land use practices than the use of either alone. Indeed, it has been claimed that agroforestry is in a unique position to learn from traditional knowledge and practice and combine with indigenous experimental initiatives and potentials (Rochleau, 1987). Certainly, local knowledge can provide a useful basis for preliminary formulation of hypotheses which can then be referred to scientists (Howes, 1980).

Productive synthesis of local and scientific knowledge depends fundamentally on whether the potential complementarity of local and scientific knowledge is matched by compatibility.

In principle, scientific knowledge attempts to explain causality and engender that causality in universally applicable and predictive principles. Opinion is divided as to whether indigenous or local knowledge can also be said to provide systematic explanation of causality.

Niamir (1990) suggests that local knowledge is entirely based on what people think it is necessary to know - they may see correlations but feel it is unnecessary to explain causality. She states that local knowledge does not devise general principles and absolutes but allows an understanding of the heterogeneity of local conditions, while formal science seeks general applicability but, particularly in natural science, has difficulty in attempting to cope with the variability that is found in the real world. By operating at different scales, this view regards the two systems as complementary but of a fundamentally different nature and held in different conceptual frameworks.

By contrast, there is evidence to suggest that some indigenous knowledge is not directly utilitarian (Berlin, 1973; Howes and Chambers, 1980). Howes and Chambers (1980) state that *"ITK [Indigenous Technical Knowledge], like scientific knowledge should be regarded in the first instance as something which became possible as a result of a more general intellectual process of creating order out of disorder, and not simply as a response to practical human needs, such as health and sustenance"*.

Berlin (1973) shows that not only are there close similarities between folk and scientific classifications in terms of formal structure but that *"There is, at present, a growing body of evidence that suggests that the fundamental taxa recognised in folk systematics correspond fairly closely with scientifically known species suggesting a universal ordering of the natural world"*.

Knight (1980) argues that *"ethnoscience and formal science share a common 'quest for explanatory theory' in accounting for the apparent diversity, complexity, disorder and regularity in the environment."* The resulting theories produce a wider causal context than simple common sense although they may be complementary in day-to-day activities.

The research decribed in this thesis was predicated on this view that there is no fundamental distinction between indigenous and scientific knowledge in relation to the motivation behind the generation of that knowledge.

In addition to a common motivation, there appears to be no fundamental distinction between the mechanisms by which indigenous and scientific knowledge are advanced. It is claimed that logical analysis and experimentation underpin advances in both (Howes and Chambers, 1980). Indeed, there is clear evidence for deliberate experimentation amongst local people, and it seems probable that active experimentation is widespread (Richards, 1939; Richards, 1985; Howes and Chambers, 1980).

While motivations and mechanisms may be comparable in a fundamental sense, it is clear that there are important practical differences between scientific and local knowledge. Table 1.1 summarises some of these differences.

Table 1.1 Practical differences between indigenous and scientific knowledge (Brokensha, pers. comm.)

	INDIGENOUS KNOWLEDGE	SCIENTIFIC KNOWLEDGE
SCALE	local	universal
LEADERS	secular	professional
TRANSMISSION	informal	formal
ACCESS	generally open	generally open
EXPERIMENTS	limited	vast
ORTHODOXY	present	present
SUPERNATURAL	present	absent

It is clear, then, that scientific and local knowledge differ in four aspects :

- **methodological structure** - Science has, theoretically, a formal structure of hypothesis, proof or disproof (through experimental evaluation rather than logical deduction) and acceptance or rejection. It seems likely that such rigorous procedures are rare beyond the western scientific tradition.;

- **institutional framework** - The international structure of scientific education, dissemination of findings, peer review and debate facilitate progress in science. While comparable procedures exist within some culturally defined communities (for example the Poro and Sande secret societies of the Kpelle of Liberia, (Murphy, 1980)), the network of individuals involved in support, interaction and innovation and the breadth of experience that can be drawn on will be very small in comparison with the scientific network.;

- **technical facilities and ability** - Science depends heavily on advances in instrumentation and methodology to advance understanding. Such instrumentation expands the range of sensory perception open to the scientist. Means of augmenting the basic senses for local communities are much more limited. Richards (1980) points out that *"the farmer will not know that which he cannot observe fully and completely"* and that *"qualitative judgements and farm decisions based on quantification cannot be better than the level of accuracy inherent in the quantitative procedure used by the farmer"*. The level of accuracy attained by farmers will normally be significantly lower than that attainable by scientific research.; and

- **scale of perspective** - Science seeks, in principle, universal explanation. For the practical purpose of managing a land use system, farmers normally only seek understanding with regard to their immediate surroundings.

In summary, it has been argued in this section that :

- Understanding local knowledge is, in conjunction with understanding practice and environment, an important means of understanding needs and constraints facing farmers.

- Understanding local knowledge is an important activity in improving the communication between target communities and development professionals in the development process.

- Facilitating active partnerships between development professionals and local people in research can help to result in effective and well-targeted research.

- Local knowledge provides a useful source of understanding of the functioning of agroforestry practices that may be complementary to scientific knowledge.

- There is no reason to believe that scientific and indigenous knowledge about agroforestry are fundamentally different in origin and motivation. While this does not necessarily mean that they are held in comparable conceptual frameworks, it does suggest some level of compatibility.

- Science is well equipped to generate robust and widely applicable theoretical frameworks. Indigenous knowledge may provide an effective source of the detailed and context-specific information needed to test and apply predictive theory. This suggests a complementarity.

1.6 SYNTHESISING LOCAL AND SCIENTIFIC KNOWLEDGE

In § 1.5, it was argued that while local knowledge about particular components of systems and their interactions may be robust, local knowledge systems may not contain robust broader conceptual frameworks. By contrast, agroforestry as a science can produce useful and predictive knowledge based on a mechanistic understanding of fundamental processes but lacks widespread experimental validation (Nair, 1990). Given the complementary nature of local and scientific knowledge (§1.5), rapid and effective advancement of agroforestry as a development activity will be facilitated by an effective combination of the two.

Professional knowledge represents an informal synthesis of local and scientific knowledge, taking advantage of the complementary nature of their differing scales. In a practical sense it is this synthesised knowledge that is the primary resource used in the development process. However, Budd *et al.* (1990) report a widespread perception amongst agroforestry professionals of a pressing need for more effective means of accessing and using existing information. They cite poor availability, low quality and inappropriate types of information as critical problems. Systematically collating, evaluating and synthesing scientific and local knowledge may provide a more accessible and usable store of existing knowledge than is achieved by informal synthesis.

The development of rigorous, formal, means of recording, evaluating and synthesing local and scientific knowledge is desirable because it may :

- facilitate a more valid and accurate synthesis of local and scientific knowledge than otherwise possible;

- provide means of more rigorously evaluating the knowledge used in development activities: and

- provide mechanisms for creating a durable store of professional knowledge. Much professional knowledge in agroforestry is ephemeral. Development professionals may develop considerable expertise over the years, but much is lost when they move on.

The development of formal approaches to the synthesis of scientific and local knowledge into a clear body of professional knowledge could have a significant impact on the generation and implementation of agroforestry solutions to land use problems in a development context. This would take advantage of the range of knowledge from the range of different sources relevant to agroforestry and facilitate more effective evaluation and control of activities.

1.7 CONCLUSIONS

There is an important gap between the tasks that need to be addressed by agroforestry as a development activity and the ability of agroforestry as a science to provide the information necessary to support these tasks effectively. This gap results from the particular methodological problems that are encountered in undertaking rigorous research to generate predictive understanding of complex systems such as agroforestry practice. Significant developments are being made in relation to many aspects of the management of agroforestry practices, but, it seems reasonable to suggest that significant gaps in the scientific understanding of the functioning of agroforestry systems (in relation to practical management) will remain for some years.

In addressing these problems, it is proposed that much existing knowledge about agroforestry is under-utilised, particularly the knowledge held and used by farming communities actively involved in the management of agroforestry practices. Making better use of this existing knowledge may provide a means of filling the gap in the knowledge available to agroforestry development programmes until more robust scientific knowledge is generated. Equally, it may help to provide a more informed basis for planning research activities. Both of these processes demand the effective synthesis of local and scientific knowledge.

CHAPTER 2 KNOWLEDGE-BASED SYSTEMS AND AGROFORESTRY

2.1 INTRODUCTION

In Chapter 1 the development of formal means of accessing, evaluating and synthesising local and scientific knowledge about agroforestry was proposed as providing a powerful means of enhancing development orientated research in agroforestry. This chapter describes the exploration of knowledge-based systems approaches to the formal representation and manipulation of knowledge upon which the methodological development described in this thesis was based.

A knowledge based system is the combination of a knowledge base (an articulated and defined set of knowledge stored on a computer) and an inference mechanism (a logic-based algorithm that allows that knowledge base to be reasoned with automatically). The use of knowledge-based systems in the research described was based on the assumption that any mechanism designed to augment existing informal synthesis of local and scientific knowledge must be based on an explicit statement of that knowledge. The explicit statement of, or, more durably, representation of, knowledge facilitates rigorous evaluation against criteria such as accuracy, completeness and utility because it provides a basis for explicit statement of the process of evaluation which can, therefore, be evaluated in itself.

The development of formal mechanisms for evaluating and synthesising local and scientific knowledge not only demands means of generating explicit representations of that knowledge but also demands formal (testable) means of manipulating it. In this context, knowledge-based systems are of particular interest. Artificial intelligence, of which knowledge-based systems are one facet, is fundamentally concerned with the development of computer-based approaches to reasoning, making use of automated reasoning procedures that are precisely defined in the code in which they are implemented. As a result, artificial intelligence approaches to knowledge representation and automated reasoning provide a potential basis for the development of evaluable mechanisms for synthesising knowledge from different sources.

The reasoning procedures used in artificial intelligence applications are based on logic and can, in principle, be applied without the use of a computer. However, taking advantage of the speed and reliability of computers in undertaking tasks that are tedious and repetitive and may involve consideration of large amounts of information greatly increases the efficiency with which tasks can be undertaken, and thereby extends the range of formal reasoning procedures that can be practicably applied.

Hence the application of knowledge-based systems techniques provides means of :

- explicitly representing knowledge;
- developing formal and explicit (and therefore testable) means of evaluating and synthesising that knowledge; and
- automating the manipulation of (including the reasoning with) that knowledge, thereby taking advantage of the efficiencies associated with the computational power of desktop computers.

Knowledge-based systems applications are most familiar in the form of expert systems. Expert systems consist of three parts: a knowledge base, an inference engine and a user interface (Bratko, 1990). The knowledge base contains the knowledge that is used in the expert system. The inference engine consists of inference mechanisms that reason with that knowledge. The interface allows effective communication between the system and the user. The interface and inference engine together are referred to as the shell. This distinction between the knowledge base and shell is made because, in principle, the knowledge used will depend on the particular task in question, while the shell may be generic across a number of similar tasks.

The primary emphasis in the development of expert systems is on the automation of specific reasoning tasks. As a result, the contents of the knowledge base used by the expert system are dictated by the requirements of this reasoning task.

In the application of expert systems techniques, approaches to formal representation are developed to allow automated reasoning with the represented knowledge. Automated reasoning has been found to be applicable in a broad range of tasks including: interpretation (inferring situations from data); prediction (inferring likely consequences from a given situation); diagnosis (inferring system malfunctions from data); design (configuring objects under constraints); planning (designing actions); monitoring (comparing observations so as to plan against vulnerabilities); debugging (prescribing remedies for malfunctions); repair (executing a plan to administer a prescribed remedy); instruction (diagnosis, debugging and suggesting modifications to student response) and control (interpreting, predicting, repairing and monitoring system behaviour) (Hayes-Roth *et al.*, 1983).

Expert system applications can be divided into two broad categories - expert replacement and user support. Expert replacement may be applicable in instances such as monitoring and managing telecommunications systems but has no relevance in the present context. However, user support systems may be appropriate. Four groups of users can be identified for user support systems (Wesley, 1985):

- experts in the domain who use a system to check conclusions;
- experts in a domain who use the system to expand their knowledge and develop expertise;
- individuals with no detailed knowledge of the domain who simply seek advice; and
- individuals who use the system as a learning tool.

To be effective, expert systems require a comprehensive and valid representation of the knowledge in the domain. Because they use a suite of rules and other information to generate particular solutions to particular problems, they have proved to be effective at undertaking complex and repetitive reasoning tasks in relation to solving well-defined problems for which unambiguous procedures for solution can be specified (Doukidis *et al.*, 1985). By contrast, they are inappropriate in instances characterised by unclear statement of needs, lack of problem structure and uncertainty (Doukidis *et al.*, 1985). Efforts at devising and applying systems for use in an agricultural context have shown that even tightly defined tasks in simple, monocultural, production systems (e.g. cotton crop management, McKinion *et al.*, 1989), an enormous number of factors must be monitored and considered if reliable recommendations are to be generated (Barrett and Jones, 1989; Jones, 1989). Nevertheless, expert systems have been explored and exploited in renewable resource management; Barett and Jones (1989), Jones (1989) and Plant and Stone (1991) review the applications of expert systems in agriculture, Muetzelfeldt (1984) and McRoberts *et al.* (1991) undertake similar reviews for forest science. Balachardran *et al.* (1989) explored the application of expert systems in rural development. Expert systems have not been reported in use in an agroforestry context although possibilities have been explored by Warkentin *et al.* (1990).

The research described in this chapter undertook an exploration of the application of expert systems techniques as a means of explicitly synthesising and evaluating knowledge about agroforestry.

2.2 A TRIAL APPLICATION OF EXPERT SYSTEMS TECHNIQUES

The application of expert systems techniques to the evaluation and use of farmers' knowledge was first explored through the implementation of a simple expert system based on research into the decision-making strategies of a community of farmers in Mexico in planning maize planting reported by Gladwin (1980).

Gladwin's work explored the reasons for non-adoption of development programme recommendations on fertiliser application to maize, by elucidating the decision-making criteria and synthesising the decision-making strategies of her subject community and representing the results as decision trees. This research enabled Gladwin to conclude that *"an understanding of the logic behind the production* *decisions that comprise 'the traditional way' rather than just surveys gathering data on socio-economic status and facts about production, is necessary for a successful rural development project".*

In Gladwin's paper, information on decision-making criteria is synthesised into a single decision tree comprising only four questions (Is fertilising twice more profitable than fertilising once?, Is there a risk of loss of the plants?, Is there a risk of loss of costs of fertiliser?, Do you have the credit or capital?), each with a 'yes' or 'no' answer, leading to two possible outcomes (apply fertiliser to the maize twice or do not apply fertiliser to the maize twice). This provided a robust tool for predicting farmer practice (i.e. adoption or non-adoption of the recommendation), correctly predicting the actual decisions made by 97% of a small sample of farmers. However, the development of this decision tree involved the elicitation of detailed information on, for example, rainfall patterns, soil types, the influence of fertiliser on crop growth at different stages and the influence of drought on crops under different management regimes. The implications of this information were taken into consideration in developing two 'profitability criteria' decision trees for different soils that provided the basis for the final decision tree. However, these implications were not explicitly stated, nor were they taken into consideration in the final model (i.e. the decision tree). As a result, this model did not (and was not intended to) allow the exploration of the implication of, for example, a fact being incorrect or an unusual rainfall pattern.

It was decided that knowledge representation and reasoning techniques associated with rule-based approaches to expert systems (Bratko, 1990) might provide a suitable means of expressing all this information explicitly, such that the decision-making criteria and the knowledge upon which those criteria were based could be explicitly recorded and dynamically explored. The resulting implementation is described and discussed in the following sections.

2.2.1 Implementation

The information provided by Gladwin was represented as a set of rules and data in Prolog, one of the main artificial intelligence programming languages (Waterman, 1986) using the LPA MacProlog implementation (Johns, 1991). Prolog is a declarative programming language based on predicate logic (Bratko, 1990) and, as such, provides a more expressive environment for representing human decision making than procedural languages (Guillet, 1989a and b).

For example, the following statement :

```
A given amount of rainfall has different effects on the
different soil types in the village.  In soils with volcanic ash
(arenal soils), a few light sprinklings, together with the
residual moisture from the previous year, can moisten 10
centimetres of soil.  In sodic-like soils (barrial soils), a
regular to heavy rain of 1 hour would be needed to do the same
because sufficient moisture cannot be conserved through the
winter months
```

was used to derive the basic data :

```
Arenal soil has a good water-retaining capacity

(represented in Prolog as:
soil_character([arenal_soil]),good_water_retainer)).
```

and the rules detailing the implications of those data :

```
The soil will be moderately humid if it is a good water retainer
and the rainfall during the period in question is light or if it
is a poor water retainer and rainfall during the period in
question is heavy

(represented in Prolog as:
soil_humidity(Month,'moderately humid'):-
        rain(Month,'light rain'),
        soil_type(Soil),
        soil_character(Soil,'good water retainer').

soil_humidity(Month,'moderately humid'):-
        rain(Month,'heavy rain'),
        soil_type(Soil),
        soil_character(Soil,'poor water retainer').).
```

As soon as one or more rules had been represented the knowledge base could be reasoned with. For example, having entered the information that:

```
There is light rain in April

(rain('April','light rain').
```

queries and answers such as ;

```
"Is the rainfall in April heavy?" - "no",

"Is there light rain in April?" - "yes",

"What is the rainfall in April?" - "light rain",

"What months have light rain?" - "April".
```

could be put to the knowledge base. This reasoning is based on a process of proving whether a piece of information is true by matching it to what is known (i.e. held in the knowledge base).

In themselves the above questions are of little utility. However they become interesting when used as part of a more complex reasoning process. A forward-chaining approach to reasoning (Bratko, 1990) was used. This approach breaks rules into subgoals, and those into further subgoals until a level is reached where the subgoals can be proved to be true or false from the content of the knowledge base or through knowledge elicitation from the user. The power of the knowledge-based system is that it can reason down through a large number of rules and subgoals for each small step in the reasoning process without getting lost, tired or bored and without making mistakes (within the context of the rules used in the programming).

For example, the query :

```
"Is the soil moderately humid in April?"
```

might be represented in Prolog as :

```
? soil_humidity(April, 'moderately humid').
```

This query can be matched to the rule in the knowledge base that :

```
soil_humidity(Month,'moderately humid'):-
        rain(Month,'light rain'),
        soil_type(Soil),
        soil_character(Soil,'good water retainer').
```

Month and Soil are capitalised to indicate that they are variables that can be instantiated to particular values (e.g. April and arenal soil). Therefore, to use this rule to establish that soil is moderately humid in April, the system must first establish that the rainfall in April is light. This is achieved by finding the following statement in the knowledge base :

```
rain('April','light rain').
```

Having satisfied the first condition to the rule, the system must identify the soil type under consideration. This is achieved through knowledge elicitation from the user. Once the user has specified the soil type, the third condition can be addressed by finding the statement detailing the water retention characteristics of that soil. If the user specifies 'arenal soil' this is found to be a good water retainer, meaning that this soil will be humid in April.

2.2.2 Discussion

The development of this trial implementation demonstrated a clear utility in the formal representation of local knowledge. This process resulted in an explicit statement of the environmental information held by farmers and the rules applied by the farmers in using that information to make decisions. The use of forward chaining as an inference mechanism allowed a dynamic representation of the process of decision making. The process of creating this explicit representation was found to demand rigorous consideration of the available information and its implications. This systematic analysis resulted in the iterative development of a coherent record of the knowledge. The resulting knowledge base provides a means of evaluating the implications of the knowledge and decision-making strategies applied by farmers. The information given in Gladwin's (1980) paper was carefully and exactingly stated and represented the result of considerable fieldwork and synthesis. Nevertheless, the process of representation suggested some ambiguity, inconsistency and omissions in this carefully constructed set of knowledge. Representation of the knowledge collected during fieldwork might have facilitated the exploration of the reasons for non-adoption of the recommendation to apply fertiliser explored by Gladwin.

In her paper, Gladwin(1980) argues that :

> ... the researcher should find the 'deep-structure' rules underlying the 'surface rules'. An example of a deep-structure rule in this paper is: I will not adopt the recommendation if I plant en seco in arenal soils and do the first cultivation before the rains come.

These deep-structure rules underpin the development of the decision trees. However, explanations for this deep-structure rule were themselves available from the paper (e.g."fertilizer cannot be taken up by the crop if the soil humidity is too low to dissolve it; soil humidity depends on soil characteristics, rainfall, etc."). This fundamental ecological information as well as the deep-structure rules was explicitly represented in the knowledge base developed. Thus, the rationality behind the deep-structure rules was explicitly recorded and could, therefore, be evaluated.

The knowledge base did not fulfil the same function as the decision trees developed by Gladwin (which predict but do not explain farmers' decisions). However, they provide a basis for considering the rationality of the decisions made. Furthermore, it is suggested that the development of such a knowledge base might have greatly facilitated the synthesis and analysis involved in the creation of the decision trees.

2.3 LOCUS - A TRIAL KNOWLEDGE BASE SHELL

Decisions on the structure of the formal representation and user interfaces used in the trial application described in the previous section were dictated by the particular set of knowledge represented. All the information represented as rules or statements was used in reasoning, while the information used in reasoning was restricted to what is represented as statements or rules or can be elicited from the user. The resulting application represented, therefore, a single model of the decision-making processes and the knowledge used in decision making. This model did not provide an environment for the exploration of, for example, the implications of new knowledge, alternative opinions developed or omitting information. Furthermore, the application did not provide an environment for recording and exploring other sets of knowledge and decision-making strategies: this would demand the development of a completely new model with different rules, structures and interfaces.

By contrast, most applications of expert system techniques have involved the creation of generic shells which may be applied to different sets of knowledge for the same purposes. The utility of creating a generic software environment for the formal representation of knowledge about agroforestry and automated reasoning with that knowledge was explored in the development of a second trial application. This trial was also used to explore the implications of different approaches to knowledge representation and different means of interacting with the user of the system.

2.3.1 Implementation

LOCUS (for local understanding) was developed as a simple decision-support system shell for representing and using the type of information that farmers and development professionals have about agroforestry systems. The knowledge used in this example was based on informal interviews about the role of fodder trees in the hill farming system of western Nepal with staff at Lumle Agricultural Centre, Nepal, undertaken during a visit in February 1991.

LOCUS was developed to explore the implications of :

- representing different types of knowledge or knowledge in different ways;
- reasoning with knowledge in different ways; and
- the need to have a variety of means of interaction with the user, according to the variety of needs that different users have.

(i) Representing different types of knowledge

Two types of knowledge, heuristic knowledge and explanatory knowledge were represented in the LOCUS knowledge base.

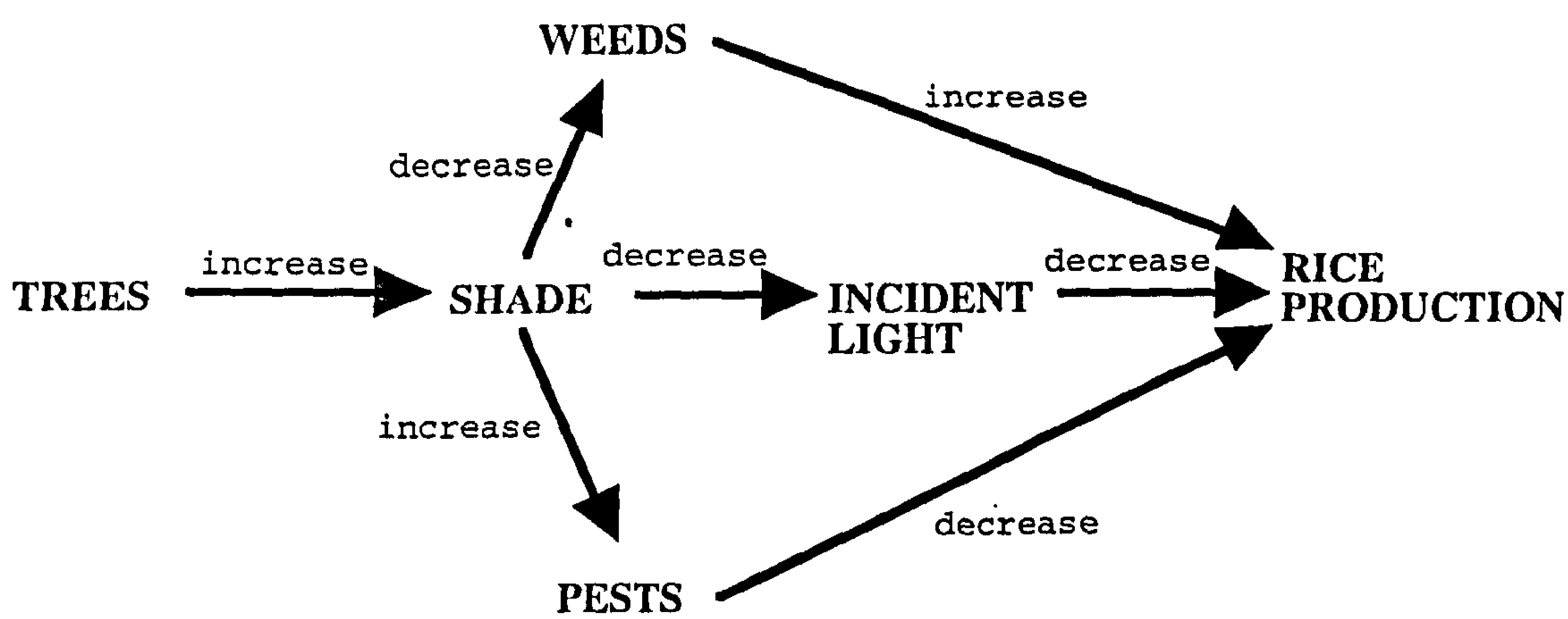
Heuristic knowledge can be described as the empirical understanding acquired through experience. This includes the types of rules that are frequently used in problem solving. Heuristic knowledge need not be based on an understanding of how a system functions, but is employed because it has been found to be useful. Reasoning with such rules of thumb is common in expert systems (Bratko, 1990).

The following is one such rule followed by a representation of this rule in MacProlog (Johns, 1991) for use in LOCUS for problem solving.

```
If your fuel production is inadequate but tree density on your
land is low, the tree canopy is light and you have surplus rice
production, plant some more trees.

superficial_solution(1,
       plant_trees if
       problem(too_little(fuel_prodn)) and
       tree_canopy(light) and
       tree_density(low) and
       rice_prodn(surplus)).
```

By contrast, explanatory knowledge is knowledge that reflects an understanding of how a system works. Explanatory knowledge is illustrated in LOCUS by consideration of causal networks. Causal networks are a familiar tool in helping to structure and communicate an understanding of complex systems (§5.2.2). Figure 2.1 represents a small portion of such a network, and is followed by the formalism in Prolog used to represent this knowledge.




```
causal_link(
        plant_trees,
        increase(shade)).
causal_link(
        increase(shade),
        decrease(incident_light)).
causal_link(
        decrease(incident_light),
        decrease(rice_prodn)).
causal_link(increase(shade),
        increase(pests)).
causal_link(
        increase(pests),
        decrease(rice_prodn)).
causal_link(
        increase(shade),
        decrease(weeds)).
causal_link(
        decrease(weeds),
        increase(rice_prodn)).
```

Figure 2.1 An example of knowledge represented as a causal network and a corresponding representation in Prolog.

(ii) Using knowledge in different ways

LOCUS fulfilled two different functions, providing an accessible store of knowledge allowing investigation of that knowledge itself (for example the importance of particular rules and the amount of knowledge relevant to particular examples) and using that knowledge in solving problems.

The basic interface (Figure 2.3) with the user explicitly shows the knowledge stored in the knowledge base and allows the user to explore that knowledge.

Inference mechanisms implemented in LOCUS allowed reasoning with the two types of knowledge to solve problems. The different reasoning mechanisms required to derive the same solution to the same problem from two different types of knowledge are schematically illustrated in Figure 2.2.

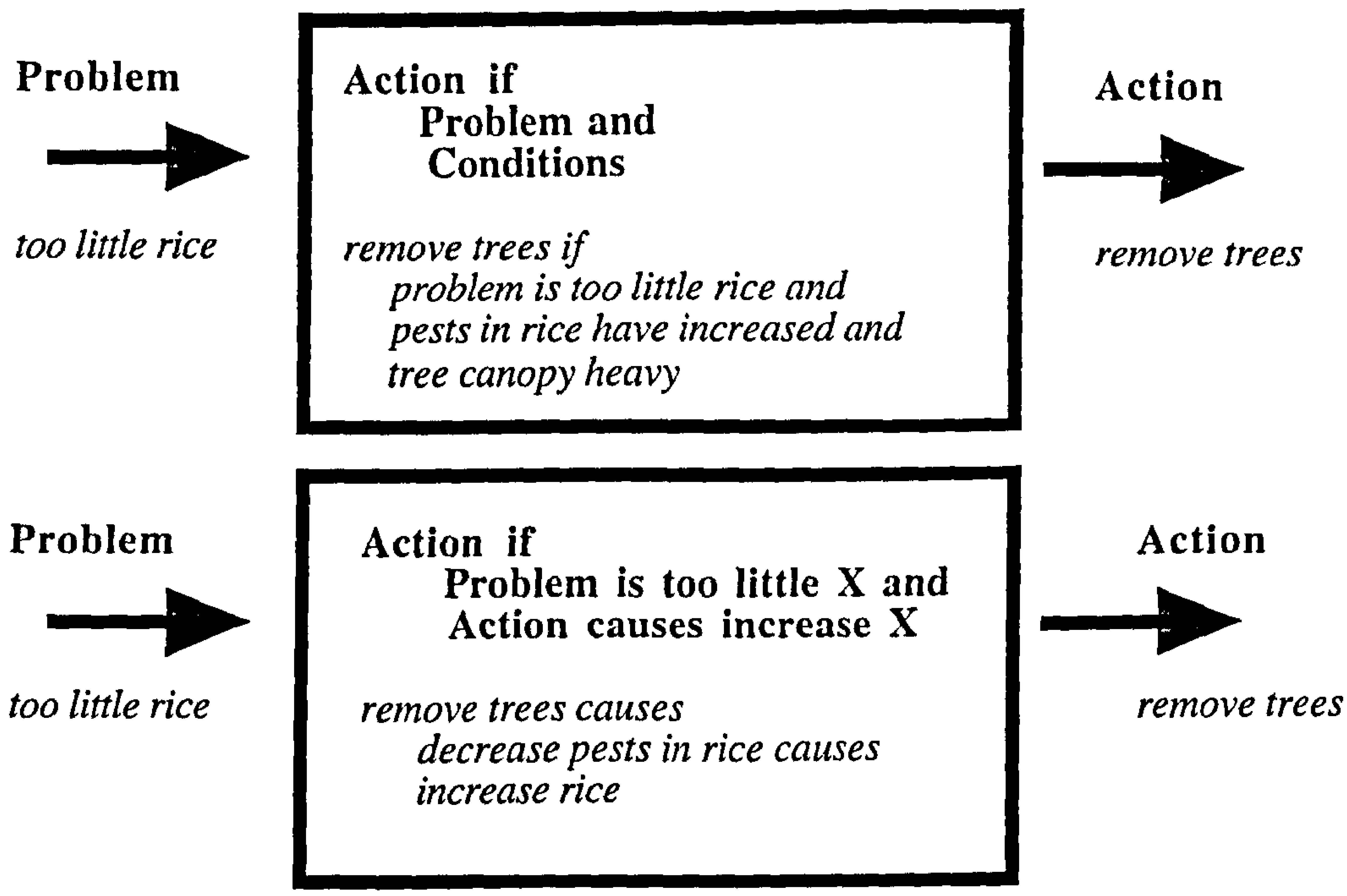


Figure 2.2 Generating the same solution to the same problem using heuristic knowledge (top) and explanatory knowledge (bottom) and algorithms appropriate to each type.

The system exhibited many standard facets of expert systems, such as eliciting further information from the user (Figure 2.4) and providing a full explanation of the reasoning process, including additional information on desirable and undesirable consequences associated with, but not directly pertaining to, the problem being solved (Figure 2.5).

(iii) Different ways of interacting with the user

LOCUS was designed to be used in a variety of different ways, according to the expertise of the user. The control panel and windows of the main interface enabled a number of different means of using the problem solving capacity, for example question and user-led reasoning for the novice user or user-initiative type reasoning for the more experienced user.

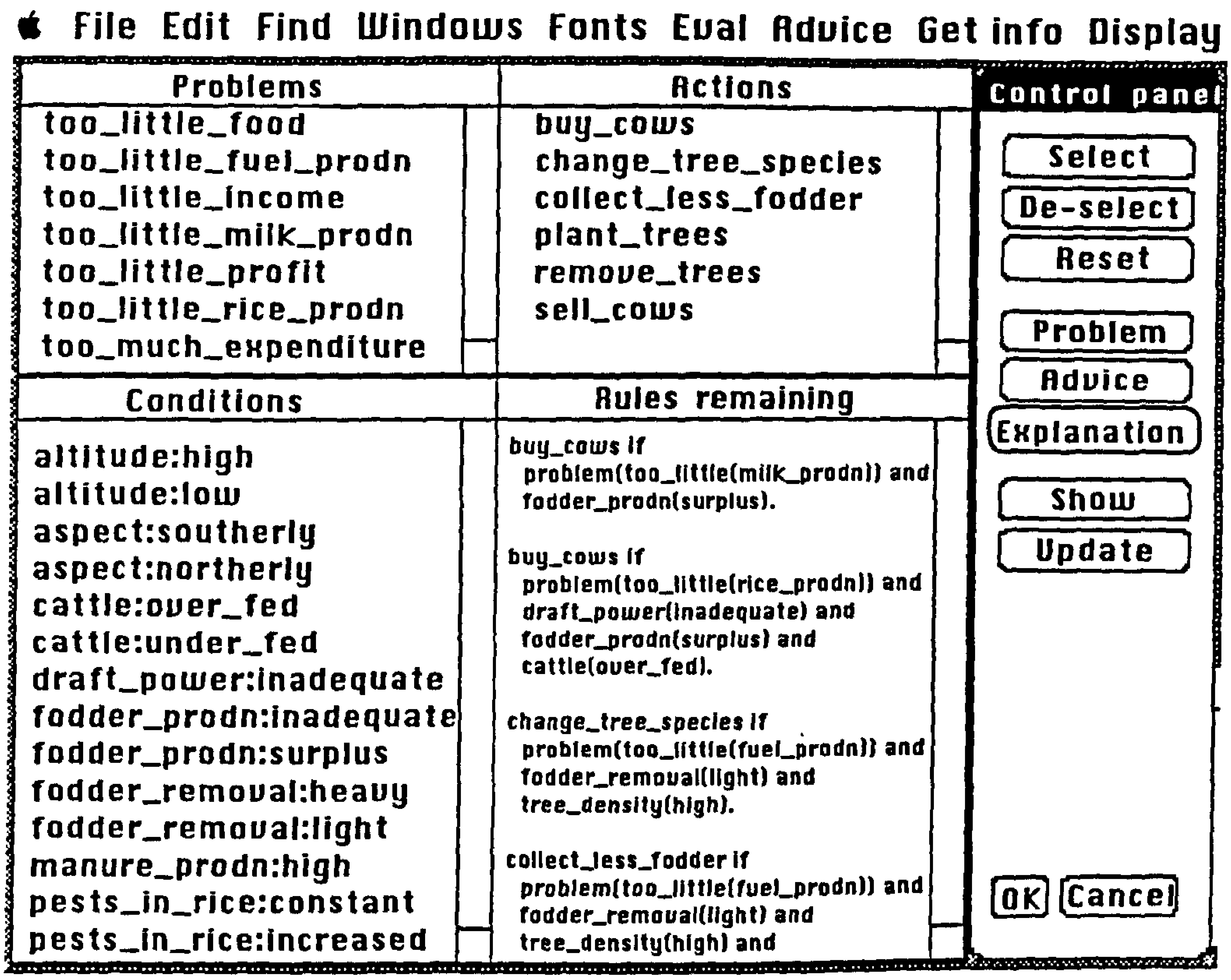


Figure 2.3 The LOCUS primary user interface

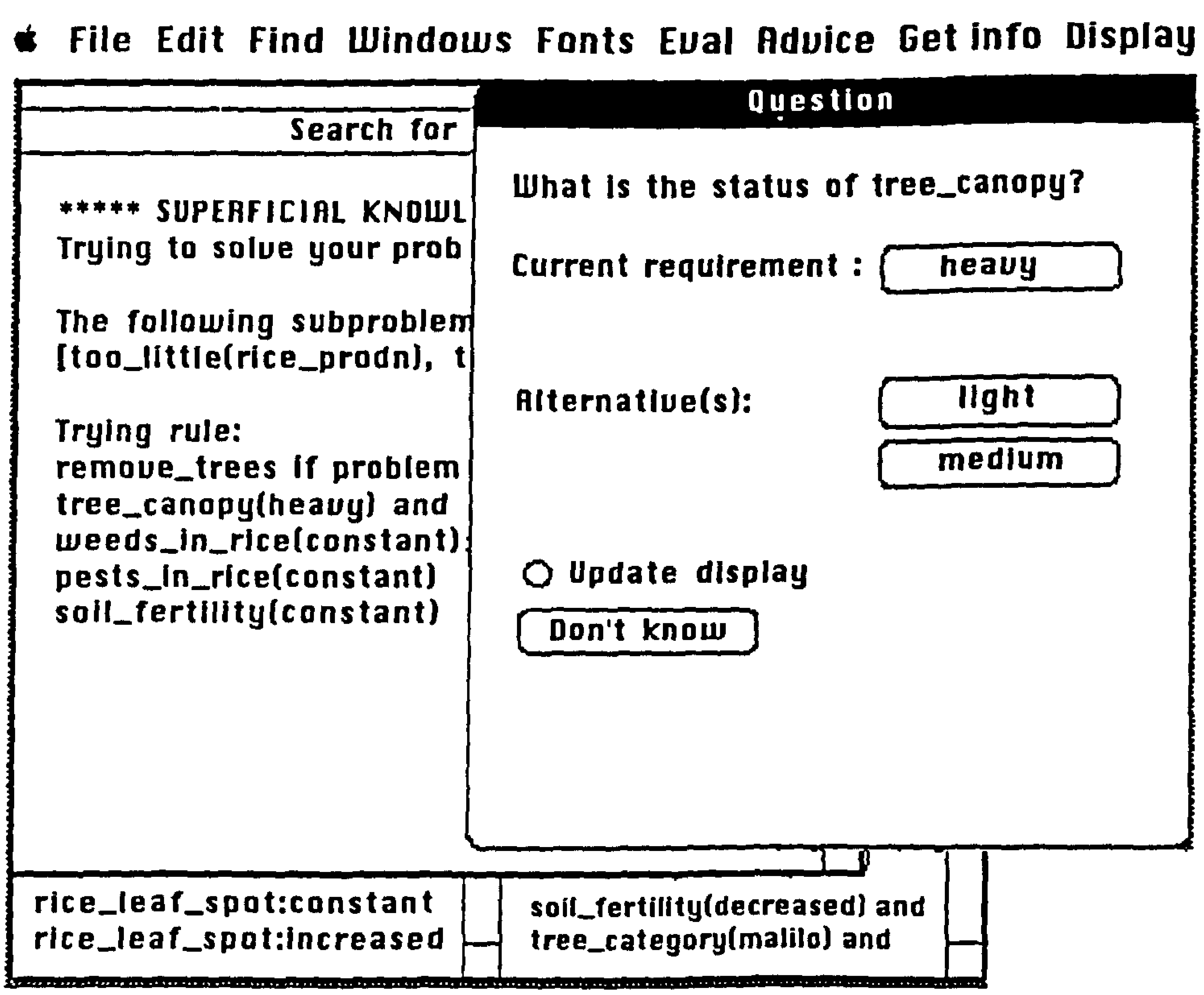


Figure 2.4 Eliciting further knowledge from the user

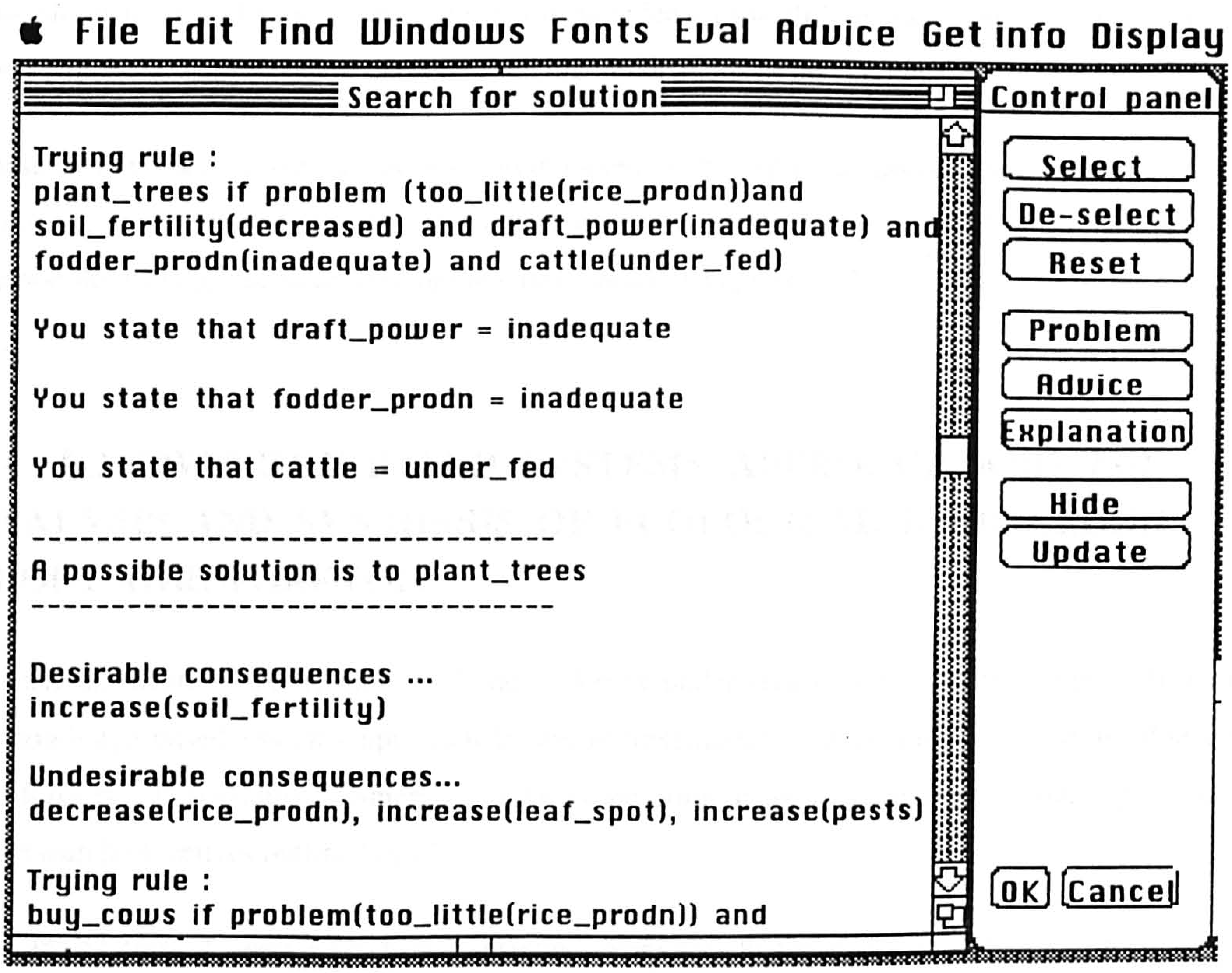


Figure 2.5 Explaining the reasoning

2.3.2 Discussion

LOCUS provides a shell in which heuristic knowledge or explanatory knowledge captured as causal relationships on any domain can be represented and reasoned with, without any modification to the interfaces or inference mechanisms used. The addition of a user interface allowing the content of the knowledge base to be edited would enable a user with only limited understanding of the process of formal representation to explore the implications of changes to the knowledge in the knowledge base or to implement a completely new knowledge base. This is proposed as being a fundamental requirement in developing an approach to representing and evaluating ecological knowledge about agroforestry and in synthesising knowledge from different sources.

As well as providing a generic environment for the representation of knowledge, LOCUS provides an environment for the type of 'modelling' of decision-making processes described in §2.2. As well as the rule-based representation used in §2.2, LOCUS allows knowledge to be represented as networks of causal relationships. Addressing the same query to produce the same answer with the two approaches to representation involved quite different types of reasoning, neither of which could be claimed to accurately represent the reasoning procedures used by farmers in making decisions. As a result, it became clear that automated reasoning could not be claimed to provide a model of the decision-making procedures used by farmers, but rather that it provided a powerful means of exploring the knowledge upon which farmers base their decisions and the implications of that knowledge.

These two utilities, providing a generic environment for the representation of knowledge and means of exploring that knowledge through automated reasoning, have featured prominently in the methodological and software development described in Chapters 4 - 7.

2.4 KNOWLEDGE-BASED SYSTEMS APPROACH AND THE ANALYSIS AND SYNTHESIS OF ECOLOGICAL KNOWLEDGE ABOUT AGROFORESTRY

The trial applications described in §2.2 and §2.3 were undertaken in order to explore the application of a knowledge-based systems approach in the representation and evaluation of knowledge about agroforestry. This section summarises the proposed utility of such an approach, providing a basis for the research objectives outlined in §2.6.

The application of knowledge-based systems techniques to the management of knowledge about agroforestry lies most appropriately somewhere between the creation of expert systems or decision support systems and the creation of a database. The contents of knowledge bases created through the application of the approaches developed in this thesis are usually dictated by a need to capture available knowledge from a defined set of sources on a defined topic for a defined purpose, irrespective of the existence of inference mechanisms that are specifically tailored to the task in hand. However, the ability to apply automatic reasoning techniques to the contents of the knowledge base means that it is a knowledge base rather than a database. Neither the terms database, not expert system adequately describe this objective. The key element of the approach implemented and evaluated in this thesis is the development of a set of knowledge that is formally represented in such a way that it can be automatically reasoned with.

Therefore, formal representation is of interest in the current context for a number of reasons. In particular :

- the process of formal representation demands an analytical approach to knowledge, thereby forcing careful assessment of elicited knowledge;
- formal representation of abstracted sets of knowledge results in knowledge bases that can be evaluated;
- mechanisms for evaluation improve efficacy in the acquisition of coherent and representative knowledge on defined topics;

- formally represented knowledge bases are transportable and can therefore be effectively accessed by the range of individuals in the range of circumstances in which their use might be appropriate;

- a knowledge base can represent an encyclopaedic resource, allowing knowledge from a range of sources to be combined; and

- mechanisms for supporting the use of a formalised knowledge base in decision making in relation to agroforestry as a development activity can be developed.

Successful formal representation requires interpretation of knowledge as articulated in natural language, both in order to capture the implicit meaning of that knowledge and to conform to the restriction imposed by a particular formalism. A formal representation is invariably a stylised simplification, because natural language and reasoning is richer and more ambiguous than any existing computer language.

A knowledge base potentially provides an explicit abstraction of articulated knowledge. Depending on the objectives of a particular piece of research, the development of the knowledge base may be as important as, or more important than, the subsequent use of that knowledge base for automated reasoning. The investigation described in §2.2 and the development of LOCUS (§2.3) has demonstrated that the abstraction of facts and reasoning processes can provide a powerful means of clarifying and illuminating them.

Furthermore, the process of knowledge acquisition (eliciting and representing knowledge) may often involve considerable interaction by the developer and the informant in the construction of a knowledge base. The process of formal representation can rapidly expose ambiguity, inconsistency and omission.

Knowledge representation is more than simply encoding knowledge on a computer: it involves interpretation of that knowledge such that a single, explicit and unambiguous meaning is represented. Techniques for representing knowledge aim to develop as faithful an abstraction of the way real knowledge systems function as is required for the purpose in hand.

This abstraction may be more easily and rigorously evaluated than the source knowledge systems themselves or ethnographic descriptions of those source knowledge systems.

Full and unambiguous statement of the content of a knowledge base, recorded such that information on the structure, source, context and validity of the knowledge is available, may provide a resource that can be flexibly used for a range of different applications.

While a knowledge base will normally be developed for a particular purpose, the full record that can be achieved through formal representation means that the content of the knowledge base may be used in addressing other objectives. Knowledge from a range of sources on a defined domain can be collated in a single knowledge base. This knowledge base can then be flexibly internally partitioned such that any subset may be used for any particular purpose. This flexibility, in conjunction with the collective power of collated knowledge may provide a flexible encyclopaedic resource.

It has been proposed in Chapter 1 that the explicit statement of knowledge in a form in which it can be collated and evaluated is a fundamental requirement in improving the efficacy of agroforestry activities. On this basis, a knowledge-based systems approach is proposed here as providing a framework for facilitating more effective evaluation and use of available knowledge about agroforestry by providing means of improving the acquisition, storage, evaluation and dissemination of knowledge. The explicit representation of knowledge facilitates the creation of comprehensive, coherent, consistent, manageable, and therefore useful, sets of knowledge. This process provides a basis for evaluation of knowledge in terms of its utility, which, in turn, facilitates acquisition and improved efficacy in the acquisition of coherent knowledge on defined topics that is representative of the understanding of farmers and development professionals.

A knowledge base may be used to help a user to:

- synthesise an understanding of the ecology of an agroforestry practice with reference to knowledge from a range of sources;
- evaluate the current state of knowledge on a defined domain;
- evaluate the current statement of knowledge of a target community.

These activities in turn may facilitate :

- the identification of adaptive research priorities; and
- the production of extension materials.

(i) Synthesis

The combination and use of ecological knowledge about agroforestry from a range of sources is premised on the view that ecological knowledge can be articulated and interpreted independently of the cultural context from which it is derived (§1.6). This makes abstraction and formal representation a valid process.

Collating knowledge from a range of sources provides a powerful resource for an individual attempting to synthesise a personal understanding of a particular topic. Inference mechanisms may be applied to the contents of the knowledge base that facilitate this synthesis.

(ii) Evaluation of the current state of knowledge

An understanding of knowledge on a defined domain depends on an understanding of the extent, validity (correctness) and utility of current knowledge. This may be facilitated by abstracting and formally representing that knowledge.

(iii) Evaluating the knowledge held by target communities

It has been suggested in Chapter 1 that developing a detailed understanding of the knowledge of target communities could be a valuable process in many development activities. This process provides a basis for improving targeting and communication in the development process.

Evaluation of the knowledge held by particular communities and, therefore, their development needs, can be achieved through direct investigation. Ethnographic techniques are successfully applied to understanding the knowledge held by communities (Conklin, 1971; Dove, 1985). Similar techniques have been included in methods of rapid rural appraisal for understanding knowledge held by communities (McCracken *et al.*, 1988). However, the formal representation of an abstracted set of knowledge from the subject community creates a more tractable and flexible resource. A representative abstracted knowledge base can be more flexibly and easily evaluated than the knowledge systems from which it has been developed. If sampling during knowledge elicitation is undertaken appropriately, evaluation of the abstracted knowledge base can be extrapolated to the parent knowledge system.

(iv) Facilitating the identification of adaptive research priorities for agroforestry

Effective prioritisation of research objectives for front-line, adaptive research programmes in agroforestry demands the careful management of considerable volumes of diverse knowledge from diverse sources, including current scientific knowledge, professional knowledge and the local knowledge of target communities. The creation and iterative improvement of a knowledge base can provide a powerful means of increasing the objectivity and appropriateness of research priorities identified.

A more rigorous approach to decision making in terms of research priorities is greatly facilitated by the formal representation of current knowledge from the diverse sources relevant.

(v) Producing extension material

Because a knowledge base can provide an environment for collating and synthesising knowledge from a range of sources into a coherent record of current knowledge on a defined domain, it can provide a useful resource from which subsets can be extracted and synthesised in the creation of extension material. Because this may be a much less demanding task than the creation of extension material without reference to a central store of knowledge, and because computer-based support mechanisms can be developed, this approach provides a means of improving the targeting of extension materials to small groups or individual clients.

In summary, the abstraction and formal representation of knowledge may facilitate significantly more rigorous treatment and, therefore, more effective use, of available knowledge than would otherwise be possible. However, the use of formally represented knowledge depends on that knowledge being sufficiently expressive to capture as much of the true meaning of the knowledge as possible, subject to the objectives of any particular piece of research, and formalised in a manner allowing a systematic and meaningful evaluation of the knowledge.

2.5 CONCLUSIONS

The application of artificial intelligence techniques for the formal representation and use of knowledge may provide a useful vehicle for improving the use of existing knowledge about agroforestry because :

- the process of formal representation forces an analytical approach to the consideration of knowledge;
- the formal representation of abstracted sets of knowledge results in knowledge bases that can be evaluated;
- formally represented knowledge bases are transportable such that they can be used by others; and
- a knowledge base can represent an encyclopaedic resource.

In particular the process of formal representation of knowledge provides a basis for :

- evaluation of knowledge in terms of its utility, which, in turn, facilitates acquisition;
- improved efficacy in the acquisition of coherent knowledge on defined topics that is representative of the understanding of farmers and development professionals;
- analysis of knowledge to identify and prioritise research objectives; and
- the evaluation of research outputs against objectives.

On this basis, the development of a knowledge-based systems approach to research and extension is required that facilitates :

- helping users to synthesise their understanding of the ecology and management of agroforestry practices and activities;
- helping users to evaluate the current knowledge held by target communities;
- facilitating the identification of adaptive research priorities for agroforestry; and
- producing extension materials.

Five distinct requirements for the creation and use of knowledge bases can be identified :

- an effective, repeatable means of eliciting ecological knowledge;
- a means of abstracting and storing that knowledge;
- a means of evaluating the abstracted knowledge;
- a means of assessing the representativeness of the abstracted knowledge and, therefore, the validity of the extrapolation from the abstracted knowledge; and
- a means of facilitating the use of the abstracted knowledge in a decision support role.

The creation of a knowledge base must be an iterative process in which elicited knowledge is evaluated in order to drive further knowledge elicitation.

Evaluation of elicited knowledge during the creation of a knowledge base must include evaluation against a specified and clear objective for the use of the knowledge base.

2.6 RESEARCH OBJECTIVES

The conclusions resulting from the research described in this chapter provided a basis for specifying the research objectives and tasks addressed in the remainder of this thesis.

2.6.1 Background

At least four characteristics of knowledge about agroforestry are central in considering the application of knowledge-based systems to evaluating and using current knowledge about agroforestry.

- Knowledge from different sources may not be immediately comparable or compatible in terms of, for example, terminology, accuracy, validity or range of application.
- Knowledge may often be contradictory.
- Knowledge about agroforestry will normally be incomplete and therefore vague and uncertain.
- As a result of the complexity of agroforestry practices, the volume of knowledge available for use in the management of those practices is considerable and its effective application demanding.

2.6.2 Objectives

The following objectives were addressed in exploring the application of knowledge representation techniques to the evaluation, synthesis and use of local and scientific knowledge.

- To develop an approach to knowledge representation that enables the comparison and combination of local and scientific knowledge without an undue distortion of either.
- To develop a knowledge-based systems approach to the representation, evaluation and use of knowledge from different sources that accommodates the existence of alternative opinions.
- To develop an approach to knowledge representation that facilitates the evaluation of the ambiguity and incompleteness of knowledge about agroforestry.
- To develop means of formally representing knowledge resulting in an abstraction of the knowledge held by target communities on defined topics that can be demonstrated to be sufficiently representative to make it useful.
- To explore the viability of a knowledge-based systems approach to agroforestry research and extension in furthering agroforestry as a development activity.

2.6.3 Approach

These objectives were addressed in the research described in this thesis through the specification and implementation of a knowledge-based systems approach to research and extension designed to :

- enable the development of an updatable, explicit and comprehensive record of existing knowledge on an identified topic from a defined set of sources;

- allow the represented knowledge to be accessed and explored; and
- facilitate the evaluation of the represented knowledge in terms of, for example, accuracy, completeness, conflict, utility and distribution.

2.6.4 Tasks

The specification and implementation of a knowledge-based systems approach to agroforestry research and extension has required the development of means of :

- effective, repeatable, knowledge elicitation;
- abstracting and storing elicited knowledge;
- evaluating the abstracted knowledge;
- assessing the representativeness of the abstracted knowledge; and
- facilitating the use of the abstracted knowledge.

The remainder of this thesis details the specification and implementation (Chapters 3 to 6) and the application and evaluation (Chapters 7 and 8) of a knowledge-based systems approach.

CHAPTER 3 INTERMEDIATE REPRESENTATION

3.1 INTRODUCTION

On the basis of the research objectives summarised at the end of Chapter 2, research was undertaken to develop a detailed set of specifications for a software toolkit for the creation of knowledge bases about agroforestry (this chapter and Chapters 5 and 6). These specifications were then implemented as an integrated toolkit called AKT (for Agroforestry Knowledge Toolkit) (Chapter 7).

Knowledge bases created in AKT are intended to be an explicit record of existing knowledge on a defined domain from a specified set of sources. They need not represent a unified statement or explanation of the domain, may contain contradiction and are intended to provide a resource which can be accessed and explored in such a way as to help users to enhance their understanding of the domain. If advantage is be taken of artificial intelligence reasoning techniques, they must also be structured such that they can be reasoned with by suitable inference mechanisms to provide support in accessing and exploring, evaluating and using the contents of the knowledge base.

The creation of a knowledge base is divided into a number of stages (Figure 3.1).

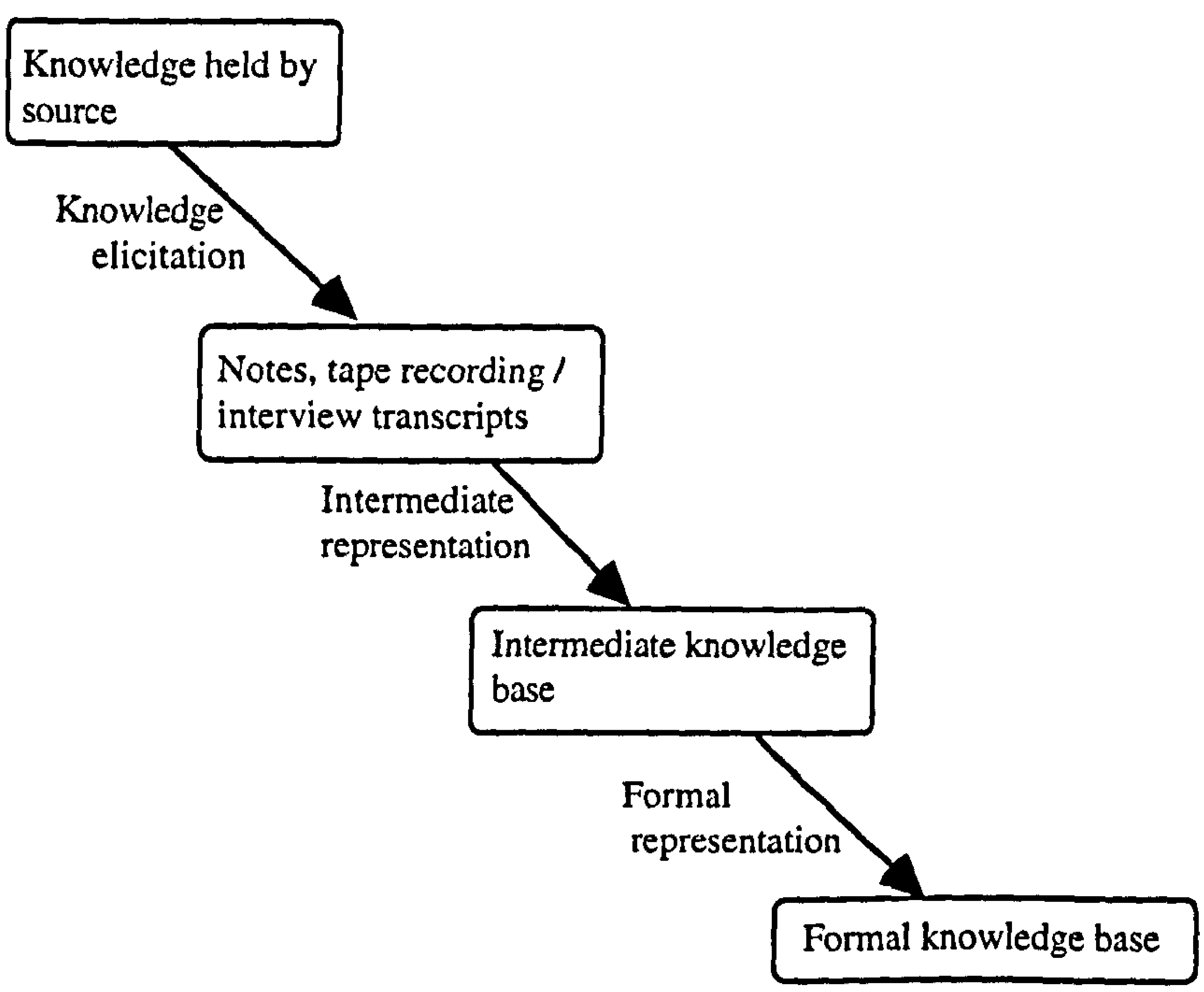


Figure 3.1. Stages and processes in the creation of a knowledge base

Knowledge elicitation involves accessing knowledge from pertinent sources. Typically the knowledge base developer interviews purposively selected informants, although use may be made of written material or other previously articulated knowledge. A record is kept of the knowledge articulated in the interview.

The approaches to knowledge elicitation that have been applied in the creation of AKT knowledge bases are discussed in §7.3.1.

Knowledge that is to be included in the knowledge base is then abstracted from this record. This process involves interpretation by the knowledge base developer and disaggregation into useful units of knowledge that are recorded in a form that is more restricted than natural articulation and provides a resource for subsequent formal representation. This process is termed intermediate representation.

Knowledge elicitation and intermediate representation together are referred to as knowledge acquisition.

Formal representation involves coding knowledge in a restricted format that can be used for automatic reasoning. Formal representation is the final step in the creation of a knowledge base, although in practice it will usually be followed by a further cycle of development (§7.1).

By definition a knowledge-based systems approach involves the development and use of a formal knowledge base. Therefore, the intermediate representation is, in principle, transient. However, formal representation may not completely capture the meaning of the intermediate representation. As a result it may be desirable to save the intermediate statement of a piece of knowledge coupled to the formal statement derived from it.

In this chapter, a specification for the creation of an intermediate knowledge base is developed. This specification resulted from the exploration of objectives and constraints, particularly through the development and exploration of trial approaches. In Chapter 5 a specification for the formal representation of the intermediate knowledge base is developed. In Chapter 6 the use of diagramming as a means of developing knowledge bases is proposed.

This order rationalises the iterative cycle of development involved in the creation of these specifications (Figure 3.2). Unconstrained knowledge elicitation provided a basis for the initial development of specifications for intermediate representation (Walker and Southern, 1992). Subsequent application and iterative development of these specifications facilitated the development of more comprehensive and coherent intermediate representations (§3.2). These in turn provided at basis for the development of a specification for formal representation (§4.4). The development and application of specifications for formal representation in turn resulted in further modification of approaches to intermediate representation, both through illuminating inadequacies in existing intermediate representations and as a result of the limitations imposed by the 'state of the art' in knowledge representation and automatic reasoning techniques (§7.7 and Thapa, in preparation). In particular, the recognised need for the development of intermediate representations that better captured the linkages between knowledge (Chapter 6) was reinforced by experience in formal representation.

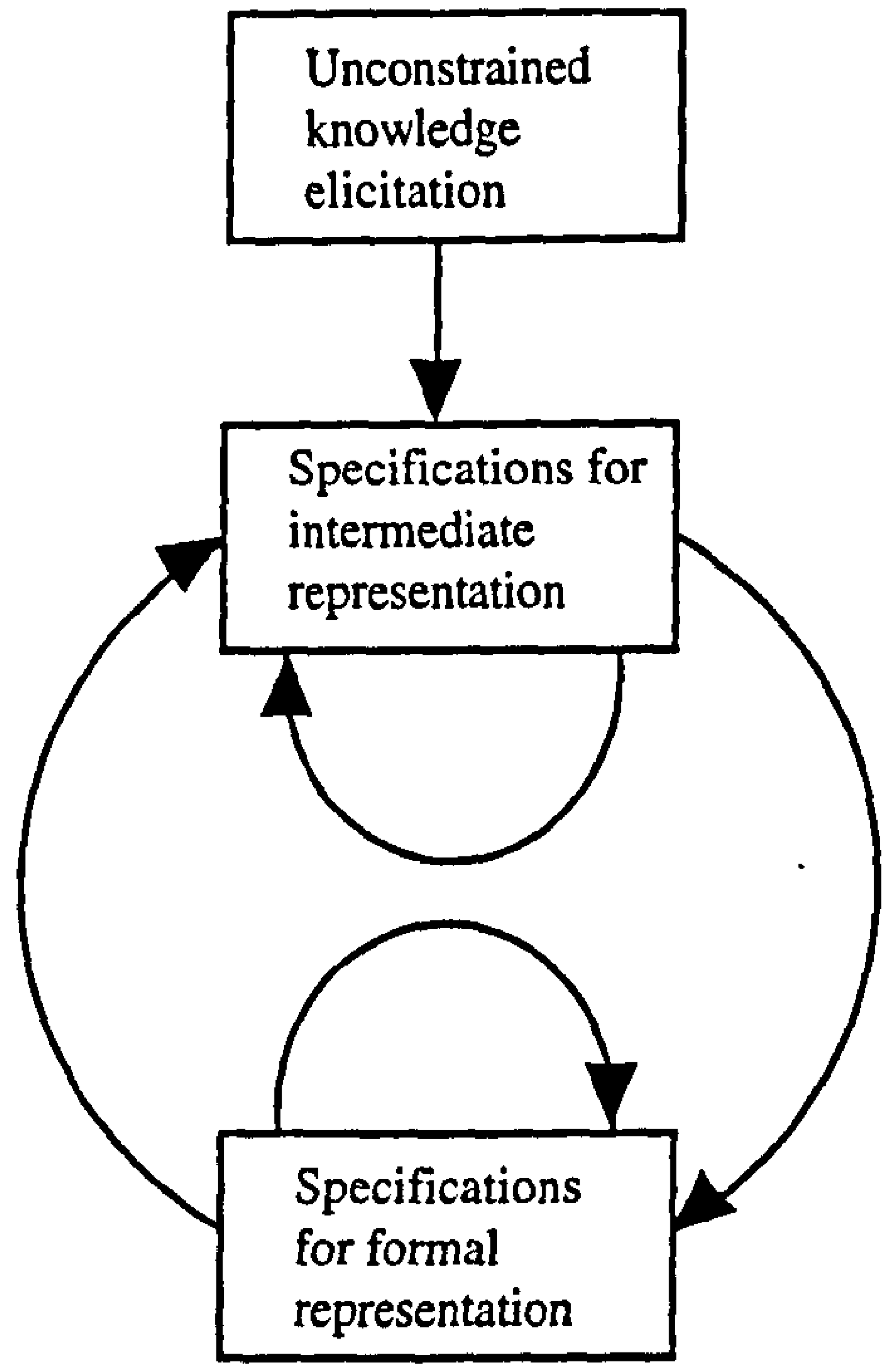


Figure 3.2 Activities in the development of the software specifications. Arrows represent evaluation and consequent modification.

The implementation of the specifications as an integrated software toolkit (Chapter 6) provided a rigorous means of evaluating their practicality. Use of the software, both on a trial basis and within other research (Chapter 7) provided an evaluation of the utility of the specifications as implemented.

3.2 INTERMEDIATE REPRESENTATION OF ECOLOGICAL KNOWLEDGE

The knowledge articulated by an informant during an interview or in written text comprises a particular combination of basic units of knowledge, structured according to the context of articulation. In the creation of a knowledge base, this particular combination is of less interest than the basic units used in constructing that combination and an understanding about the ways in which those units can be linked. Means of disaggregating and recording the units of knowledge that make up the particular argument, along with enough information about them to allow them to be used in combination with others to explain different aspects of the behaviour of a system, are of more interest than recording the original combination. So, the first step in the development of specifications for intermediate representation was to identify appropriate units of knowledge into which to disaggregate interview material.

Many different units of knowledge can be defined. Rauscher (1987), for example, defines three: statements (sentences that express assertions), concepts (units of thought expressed as sets of statements) and chunks (logically consistent sets of concepts). Disaggregation of the same articulated knowledge down to these different levels will result in sets of knowledge of differing characteristics and, therefore, utility. In the current context, it is proposed that it is most appropriate to treat the smallest useful units of knowledge as the basic units in the knowledge base. This maximises the flexibility with which these units can be combined.

3.2.1 Statements of fact as basic units of knowledge

The term 'statements of fact' is used here to refer to the smallest useful units of knowledge. Knowledge is useful if it can be used in combination with other knowledge in reasoning.

Statements of fact express assertions. A statement is not the same as a sentence because different sentences can be used to express the same statement (Giere, 1984). Statements of fact contain knowledge that is useful without reference to other statements of fact but that cannot be broken down into two or more statements of fact. So:

```
The biomass production of a plant is proportional to leaf area
index of that plant.
```

is a statement of fact. By contrast:

```
Rainfall is low.
```

is not a statement of fact, because it does not contain enough information to be used in reasoning.

```
Bananas and apples are types of fruit.
```

is not a statement of fact, by this definition, because it can be broken into two statements:

```
Bananas are a type of fruit
Apples are a type of fruit.
```

Statements of fact are divided into two categories, binary statements and attribute-value statements.

Binary statements of fact capture a binary relationship between two entities. For example :

```
Rainfall causes increased soil erosion.
```

Attribute-value statements describe an attribute of an entity or classes of entities. For example :

```
Bananas are yellow.
```

Statements of fact can also be divided into statements referring to classes or types of entities, referred to here as statements of class :

```
Barley seeds germinate in seven days at 10°C.
```

and statements referring to particular instances, referred to here as statements of instance :

```
Seedling 97 germinated after nine days at 10°C.
```

These two forms of statements are syntactically identical, but they are significantly different in utility. Explanatory ecological knowledge can be derived from statements of instance, however, statements of instance are, in themselves, data and are unlikely to constitute a useful component of a knowledge base.

Factual statements are generally perceived as being, in some sense, verified or true. By contrast, the knowledge base will, in fact, consist of statements of perceived truth, or belief. With limited exceptions, where knowledge has been experimentally validated, it is, therefore, hypothetical. However, most people holding and using knowledge, particularly beyond a strictly scientific context, do not, indeed could not, view it in this way. The term 'statements of fact' is used because it captures the notion that they represent knowledge that is perceived to be true even if not necessarily scientifically verified.

The statement of fact provides a suitable basic unit for the creation of a knowledge base because a set of statements of fact represents a resource that can be flexibly explored and provides basic building blocks for the user in the construction of an integrated understanding of the domain in question. Experience in the disaggregation of knowledge from interview has suggested that restrictions to the syntactic structure of statements of fact and recording information associated with the context of the statements of fact can enhance their utility.

(i) The syntactic structure of statements of fact may be restricted to ameliorate the ambiguity associated with natural language.

Because of the ambiguity and complexity inherent in natural language, accurate interpretation of statements of fact in unrestricted natural language may often be difficult. Natural language is extremely robust in its use and interpretation; it can contain a great deal of ambiguity and imprecision, yet still serve a useful function in communication. Meaning in natural language is often context specific and, therefore, flexible. However, it cannot be assumed that the implicit contextual meaning of a statement of fact will still be understood by users once it has been included in a knowledge base. Furthermore, automated reasoning techniques cannot cope with flexibility of meaning according to context. The formal representations derived from intermediate representations must therefore be syntactically unambiguous.

The creation of an intermediate representation of ecological knowledge that provides an explicit representation of knowledge suitable for subsequent formal representation may be facilitated by limiting the permitted syntactic structure of the intermediate statement of fact. Such statements will still contain the ambiguity in the knowledge itself but will contain less ambiguity in the statement of that knowledge, such that the resulting statements can be interpreted, compared and used with greater certainty and accuracy.

(ii) Recording the context of a statement

Knowledge is inescapably contextual. Disaggregation of knowledge into a set of statements of fact will cause this context to be lost. It has already been suggested that this loss of contextual understanding demands explicit statement of the disaggregated knowledge. However, this does not, in itself, adequately compensate for the reduction in utility of the statement that results from the loss of context. It is proposed that further information of three types be appended to statements of fact to compensate.

The subject(s) of a statement of fact may be identified in order to facilitate the exploration, evaluation and use of the knowledge base

One aspect of the context of a statement of fact is the set or sets of other statements of fact with which it is often used in reasoning. These sets may be defined as a particular domain, a statement being defined as falling into a particular domain, or potentially falling into several domains, according to circumstance. So, the statement :

```
Increasing the vegetation cover on a slope reduces the volume of
soil removed by run-off.
```

might be defined, for example, as being 'about' soil erosion, or watershed management. Capturing this flexible association between statements of fact in a knowledge base may facilitate natural exploration of the knowledge base by a user.

Known limitations of validity of statements of fact may be recorded.

Most statements of fact will have only a limited validity. Validity relates to the circumstances under which the statement of fact is held to be true and the certainty that can be placed on the statement of fact being true. Neither are adequately captured in statements of fact themselves. Appending conditional information and certainty to a statement(§3.2.4), results, therefore, in a more complete record of the knowledge articulated by an informant.

Contextual information about the source of a statement of fact may be recorded

One important piece of contextual information for the interpretation and use of knowledge is the source of that knowledge. Where a user of knowledge has some experience or understanding of the source of knowledge, this understanding will inform the use of that knowledge. For this reason it is proposed that knowledge can be usefully tagged according to source. This has further utilities in facilitating the assessment of the internal consistency of knowledge from a particular source (§7.4.2) and the distribution of knowledge between sources (§7.6).

These extensions to the statement of fact as a structure for intermediate representation are discussed and evaluated in the following sections.

3.2.2 Restricting the structure of statements of fact for intermediate representation

Three levels of syntactic restriction of intermediate statements of fact have been evaluated. These have been based on the hypotheses, respectively, that:

- an adequate proportion of ecological knowledge about agroforestry can be efficiently captured in statements of fact restricted to a set of predefined sentence structures called template sentences;

- statements of fact generated in unrestricted, natural, language can be made significantly more tractable by classifying each statement of fact as falling into one or more categories of structural type; and

- familiarity with the demands of subsequent formal representation means that it is unnecessary to impose syntactic restrictions on statements of fact because the knowledge base developer will tend to restrict intermediate representation to structures appropriate for subsequent formal representation.

(i) Template sentences

The set of 'template sentences' in Table 3.1 was developed to provide a restricted set of forms for recording intermediate statements of fact about the ecology of agroforestry. These do not represent an exhaustive set of the sentence types but were proposed as a significant proportion of a hypothesised, finite set. These templates provided a basis for evaluating the utility of restricting the contents of intermediate knowledge bases to statements of fact of restricted syntactic structure.

Table 3.1 The template sentence structures applied in evaluation of the template sentence approach

____ will achieve ____.
____ will occur as a result of ____.
____ is a reason for ____.
____ is a way of ____.
____ is used for ____.
____ is important for ____.
An increase in ____ causes an increase in ____.
A decrease in ____ causes an increase in ____.
An increase in ____ causes a decrease in ____.
A decrease in ____ causes a decrease in ____.
A change in ____ causes no change in ____.
A change in ____ causes a change in ____.
____ causes an increase in ____.
____ causes a decrease in ____.
____ causes no change in ____.
____ causes a change in ____.
____ is a result of ____.
____ causes ____.
____ influences ____.
____ is defined by characteristics ___ and ___...
____ is a sort of ____.
____ is a part (component) of ____.
____ is next to ____.
____ is close to ____.
____ (activity) is done ____ (place).
____ is ____ (distance) from ____.
____ is at ____(place).
____ happens ____(place).
____ happens at the same time as ____.
____ happens before ____.
____ happens at time ____.
____ happens during ____.
____ takes ____ (duration).
____ happens ____ (frequency).
____ is an attribute of ____.
____ (attribute) can range from ____ to ____.
____ (attribute) has the value ____.
____ (attribute) is measured in ____ (units).
____ (attribute) can take possible values ____.
____ is ____ than ____.
____ is preferred to ____.
The effect of ___ is greater on ___ than on ___.

AND verb type sentences '___ verb ____' such as '____ eat ____' and '____ shade ____'.

(ii) Evaluation of the utility of the template sentences

The utility of a template sentence approach to restricting the structure of statements of fact was evaluated by the development of a knowledge acquisition software tool[1] and its use in fieldwork in Sri Lanka in April and May 1992 (Walker and Southern, 1992). The template sentence approach proved unacceptable as a practical means of creating an intermediate knowledge base for two reasons.

- The template sentence options were overly restrictive. It was usually possible to squeeze a sentence into an available template type but frequently this significantly altered the original statement of fact such that the sense of the informant's original assertion was not adequately captured.

- Generating statements of fact using a template sentence approach was time consuming. After a reasonable period for familiarisation with the task, creating statements of fact from source material using the template sentence approach took up to ten times as long as creating the same statements of fact in an unrestricted form. This difference was not related to software implementation. It resulted from the time required to select an appropriate template structure for capturing each statement of fact. First the appropriate category of template sentence had to be identified, deciding, for example, whether a temporal or causal type statement best captured the sense of the source knowledge. Then within the chosen category the template structure best capturing the meaning had to be identified, taking into account any possibly distortions (as above).

That the set of templates was overly restrictive shows that the set used was not comprehensive. A more comprehensive set of statement types increased the ability of the user to find an appropriate template type, thereby resulting in a more acceptable representation and, to some extent, reducing the time devoted to deciding between marginally adequate structures. However, the resulting set of template types was found to rapidly increase in number. Finding a desired structural type, from the hierarchy of types, became unacceptably time consuming.

While this experience did not cast doubt on the validity or the usefulness of the template sentence approach in improving the tractability of elicited knowledge, the serious difficulties encountered by the users of the approach led to the conclusion that it was practically unacceptable during knowledge acquisition.

---

[1]The Template Tool, implemented in the HyperCard by the Department of Artificial Intelligence of the University of Edinburgh (Haggith *et al.*, 1993, and Appendix A).

However, experience in the development and use of the template sentence approach has had a significant impact on :

- the development of a diagramming approach to knowledge acquisition (Chapter 5); and
- the development of specifications for formal representation (Chapter 4) by illustrating the problems associated with a 'slot-filling' approach to formal representation, resulting in the creation of a more flexible approach, starting with the identification of the basic elements of a statement and working up from these, rather than down from a prescribed structure.

(iii) Identifying structural types

Identifying the structural type of a statement of fact stated in natural language was explored as an alternative approach to restricting the structure of statements of fact. It was proposed that requiring the user to specify the structural type of a statement of fact at the point of intermediate representation would encourage more explicit, clear and consistent recording of statements of fact, which might, therefore, be more tractable for formal representation, while still allowing the statement to be rapidly and naturally articulated in natural language.

For example:

```
Fruiting in pepper occurs after rains.
```

states a temporal sequence and could be tagged accordingly. There is , however, a possible implication of causality in this statement. Tagging the statement as temporal implies that there is not necessarily a causal link between fruiting and rain, tagging a sentence as causal and temporal provides different information.

Temporal and causal statements are two of a set of 'structural types' identified for ecological statements of fact from experience in generating statement templates (Table 3.1). Tagging statements according to structural type (listed in Table 3.2) differs from fitting a sentence into a defined template structure: while the form in which the sentence is entered is not restricted, enabling the user to enter a statement substantially reflecting the way in which it was held by the source, information is available on the basic structural type, thereby facilitating interpretation of the statement. However, the detailed structure of the statement is less explicit, limiting possibilities for automated manipulation of statements in accessing, managing or exploring the intermediate knowledge base.

The tagging facility was implemented within versions of TEAK (Tools for Eliciting Agroforesty Knowledge, a prototype software package, see Haggith *et al.*, 1993, and Appendix A) and early versions of AKT (Appendix A). As well as encouraging a rigorous approach to intermediate representation, this process provided an opportunity for enabling on-line assistance in the process of formal representation (Chapter 4). However, in practice it was of limited utility. It encouraged but did not demand syntactic restriction and could not easily be checked for validity. Identifying a statement's structural type proved, in practice, to fit more effectively into the process of formal representation (§4.6).

Table 3.2 A categorisation of structural types

causality
definition
taxonomic status
spatial relationships
temporal relationships
attribute statements
comparative statements
certainty
rationale (reasons for doing something)
location for an action
function
means of achieving a goal

(iv) Restricting the structure of intermediate statements of fact through familiarity with the procedure for formal representation

Intermediate representation is, by definition, a step in the process of creating a formal knowledge base. Coherent, consistent and therefore tractable intermediate knowledge bases were required prior to the development of specifications for formal representation to provide a sound basis for that development(§4.2). This requirement demanded more exacting intermediate representation than was necessary after the completion of specifications for formal representation.

However, the systematic restriction imposed by the process of formal representation means that a formally represented statement of fact is not as expressive as natural language and therefore does not always capture the entire meaning of an intermediate statement. As a result, it is appropriate to treat intermediate statements as more than simply a transient structures in the process of formal representation, and to permanently record them in the knowledge base. Unambiguous and interpretable intermediate statements therefore remain important, as a basis for future evaluation and modification of the formal representation. This process is facilitated by the use of a syntactic structure that is reasonably consistent with the formal grammar (§4.4). In practice, experience has shown that familiarity with the process of formal representation results in expression of intermediate statements using reasonably consistent and interpretable syntactic structures.

As a result, the proposed practical advantages of syntactic restriction on the form of the intermediate statement were shown to be less valuable than the rapid and natural articulation of knowledge facilitated by the generation of a natural-language statement. This unrestricted approach does however, require that the researcher creating the intermediate statement of fact also be responsible for the formal representation of that statement. While evaluation of the formal grammar developed in this research (Chapter 7) has demonstrated that requiring both intermediate and formal representation to be undertaken by the same person is a practical possibility for postgraduate researchers, any future division of the tasks of intermediate and formal representation would demand reconsideration of imposing syntactic restrictions on the structure of intermediate statements.

In summary, it has been demonstrated that intermediate representation is most effectively undertaken without imposed syntactic restriction (beyond adherence to the basic definition of a statement of fact) but that familiarity with the formal grammar results in the *de facto* imposition of a consistent syntactic structure on intermediate statements of fact.

3.2.3 Capturing the subject(s) of a statement of fact

If a set of intermediate statements of fact, brought together as an intermediate knowledge base, is to be evaluable prior to formal representation (as may be required during a period of intensive knowledge acquisition) then it is necessary to be able to abstract subsets of that knowledge base by content. For example, it may be desirable to collate all current statements about a particular species. This can be achieved by identifying keywords in the statement and developing mechanisms for keyword searches of the intermediate knowledge base (§3.9).

The subject(s) of a statement of fact can be divided into two types: subject as captured by the explicit content of the statement of fact (content keywords) and subject captured by the implied content of a statement of fact (topic keywords).

In the statement:

```
Fruiting in pepper occurs after the rains
```

Fruiting, pepper and rains might be identified as content keywords.

This statement might be considered to be about seasonality, which might be defined as a topic keyword.

In practice, topic keywords have rarely been specified in the creation of knowledge bases (§7.1). Several facets of the identification and use of topic keywords may account for this :

- the identification of topic keywords is significantly more subjective than the identification of content keywords;

- while there are only a small number of words in the intermediate statement that can be defined as being keywords, by contrast a statement may be defined as being about a very large number of topics, depending on circumstance; and
- the function of the topic keyword is, to some extent, fulfilled by the use of properties of inheritance in the creation of hierarchies of keywords (§3.5) and in the creation of means of handling equivalence (§3.3).

While, in principle, the identification of topic keywords provides a flexible means of grouping statements, it is suggested that a more objective basis for identifying and using topic keywords is demanded before the practical utility of this approach can be realised.

By contrast, the identification of content keywords within an intermediate statement of fact has proved, in combination with an appropriate search mechanism (§3.9), to result in the creation of more tractable intermediate knowledge bases.

3.2.4 Capturing the validity of statements of fact

Knowledge about the validity of a statement of fact is important in determining the way in which that statement of fact can be used. Most statements of fact will be known to have a limited validity. Validity can be described both by the circumstances under which a statement of fact is believed to apply (conditionality) and in terms of the confidence with which the statement is considered to be true (certainty).

Conditions can be captured as a set of statements attached to a statement of fact by the operator IF and structured through the use of ANDs and ORs, for example :

```
Crops are prone to lodging IF (there is a strong wind AND crop
roots are exposed) OR (there is a strong wind AND crop stems are
weak).
```

or, equally :

```
Crops are prone to lodging IF (crops stems are weak OR crop
roots are exposed) AND (there is a strong wind).
```

The precedence of AND over OR, or of OR over AND, is arbitrary. Throughout this thesis OR is taken as being dominant over AND (as in the first example above).

Statements of fact may contain an implicit conditionality. For example the statement :

```
Heavy rainfall causes soil erosion
```

contains the implication that soil erosion only occurs where rainfall is heavy. Because this conditionality is not explicitly stated it is less easy to interpret and use, particularly for automatic reasoning. The same statement is better represented as:

```
Rainfall causes soil erosion IF rainfall is heavy.
```

The precise and complete statement of conditionality has proved to be a key determinant of the utility of a knowledge base.

Even where the conditions under which it applies are detailed, statements of ecological fact will almost always have an associated degree of uncertainty. The confidence that can be placed on a piece of knowledge will have an impact on the way that that piece of knowledge is used. Efforts were made, on a trial basis, to elicit a certainty rating from 0 (known never to occur) to 5 (known to occur with absolute certainty) for statements of fact. However, it was found that informants were rarely comfortable or confident in stating a degree of certainty associated with a statement of fact. Furthermore, it is difficult to standardise classification of certainty. So, while in principle a statement of certainty provides important information, in practice it cannot be easily collected or meaningfully used and was excluded from the specification for intermediate representation.

3.2.5 Recording the context of a statement of fact

One means of recording the context of statement of fact is to record the linkages between that statement of fact and others. Most potential linkages between statements of fact can be captured by a consistent wording of those statements, such that linkages can be deduced, for example where two statements are used to deduce a third :

```
Harrowing happens after ploughing
Sowing happens after harrowing
```

therefore:

```
Sowing happens after ploughing
```

or :

```
Root waterlogging causes complete flower loss
Complete flower loss causes crop failure
```

therefore:

```
Waterlogging causes crop failure.
```

These types of chains of linkage may be used in reasoning and explanation. However, such chains may not be immediately apparent during knowledge elicitation. Instead, explanation is often provided by linking one statement to another with the term BECAUSE. For example :

```
Growing nebharo on crop terraces reduces crop yield BECAUSE
Nebharo has large leaves AND
Large leaves result in a high tapkan2 effect on crops
```

To capture this type of linkage, a mechanism for recording linkages of explanation through the use of BECAUSE was implemented in a version of TEAK (Appendix A). This led to sets of statements of fact being recorded as explanations for single statements even where explanation was not otherwise or not necessarily deducible from the statements themselves (Thapa and Walker, 1992). However, in practice this mechanism militated against rigorous consideration of knowledge. It was found that in the majority of instances where a BECAUSE link was used, further work could identify a deducible chain. For example :

```
An increase in tapkan effect causes a decrease in crop yield
An increase in tree leaf size causes an increase in tapkan
effect
Nebharo has large leaves
```

This reinterpretation (which may need to be validated by future reference to the source) is preferable because :

- a general statement of relationship is more useful than a statement of a particular instance of that relationship; and,

- causal representations are more tractable for automatic reasoning (§4.6).

This approach additionally requires a consistent use of values for the attribute leaf size (for example small, medium and large) and information about the relationship between those variables (for example, large leaves are bigger than medium-sized leaves) if automatic reasoning is to be possible. While more demanding, the resulting knowledge base will capture general meaning concisely and allow that general meaning to be applied to specific instances, rather than resulting in a data base of particular cases of a response.

3.2.6 Source

One of the aims of developing a knowledge-based systems approach to agroforestry research and extension was to facilitate the use of knowledge from different sources, including professional, local and indigenous knowledge (§1.6). Where knowledge from a range of sources is used, it is proposed that each statement of fact needs to be tagged according to source in order to :

- facilitate the clarification of information through reference to the source;

---

[2] Erosion effect caused by drops of water falling from the leaves of trees, see Thapa (in preparation)

- allow a comparative evaluation of the nature, extent and utility of knowledge from different sources or types of source; and,

- ensure, for ethical reasons, that due recognition of the source of potentially valuable information is maintained.

The content of any single statement may be separately articulated by more than one source or a statement from one source might be corroborated by another source. For this reason a statement of fact might be attributed to multiple sources.

3.2.7 A specification for statements of facts as an intermediate representation

This section details the specification for the structure of intermediate statements used in constructing knowledge bases that was derived from the research described in the foregoing sections.

The generic structure for an intermediate statements is :

(Intermediate statement of fact, [Contents keywords], [Conditions], [Source]).

Note : [ ] indicates that the contents of the brackets are stored as a list.

(i) Intermediate statement of fact

The intermediate statement of fact is a natural language statement which conforms to the definition of the structure of a statement (§3.2.1).

(ii) Content keywords

Content keywords in the informal statement of fact are identified by the user and stored as a list.

(iii) Conditions

Conditions are stored as lists of statements of condition divided by the infix operators AND and OR. OR has a bracketing precedence over AND (§3.2.4).

(iv) Source

Two types of related information on source are identified: background information about the sources themselves, and specific information about the interview or reference during or in which the source articulated the statement of fact. The resulting basic structure for recording source information is, therefore : (Who, When).

For example, the sentence

```
Fruiting in pepper occurs after rains (Singh, 8/8/88).
```

may be linked to information about Singh (for example, full name, age, gender, ethnic origin, education) and about the interview from which this statement was taken (for example location, interviewer).

Alternatively the source information may refer to written text. In this case the 'Who' information is linked to further details on the authors (full list of authors, institution of each) while the 'When' information is linked to further details of this particular publication (for example date, title, publisher, number of pages).

Multiple sources are stored as a list with the following structure :

```
[(Who,When),(Who,When),(Who,When)....].
```

3.2.8 Trial development of an intermediate knowledge base

An intermediate knowledge base comprising a set of statements of fact conforming to the specifications detailed above was created in collaboration with Alison Southern as a result of fieldwork in Sri Lanka (Walker and Southern, 1992). The resulting knowledge base was an effective and explicit statement of knowledge that adequately captured the majority of relevant items of knowledge articulated by interviewees. However, the coherence, consistency, and therefore tractability of this knowledge base was constrained in a number of ways :

- terms were frequently used inconsistently because there were no mechanisms to encourage consistency, seriously diminishing the utility of the knowledge base; and,
- the list of statements of fact created an entirely flat knowledge base, which therefore did not capture the efficiency in the naturally hierarchical storage and use of knowledge, and therefore contained considerable repetition.

For these reasons mechanisms for intermediate representation beyond the creation of sets of statements of fact were developed. These were :

- mechanisms for creating glossaries of terms used in the knowledge base;
- means of generating and recording definitions of the terms in the glossaries;
- means of defining the hierarchical relationship between terms in glossaries, where appropriate, and thereby providing a hierarchical structure for the knowledge in the knowledge base; and
- means of diagrammatically representing the linkages between statements of fact.

The first three of these are discussed in the next two sections, the fourth in Chapter 5.

3.3 GLOSSARIES OF KEYWORDS

The inconsistent use of terms in the creation of a knowledge base has proved to be a significant source of ambiguity. Different terms may be used to capture the same meaning, or a single term may be used to capture different meanings. Ambiguity in the representation of knowledge has a profound impact on the utility of the resulting knowledge base. Resolving this ambiguity during knowledge acquisition is much more effective than attempting retrospective resolution.

Experience has shown that the consistent use of terms is encouraged by the development and updating of a glossary of keywords. It has proved valuable to compare all the terms in a new statement of fact with terms previously used in the knowledge base (and, therefore, recorded in the glossary) to ensure consistent use of terms and to avoid a proliferation of terms to capture the same meaning.

Glossaries also provide structure in which synonymous terms can be identified, definitions for terms can be developed and the hierarchical relationships between terms can be recorded. Definitions and hierarchical relationships are discussed in §3.4 and §3.5 respectively.

Synonyms are frequently encountered in the creation of knowledge bases. The most widespread example is a set of names for the same species, for example *Piper nigrum*, 'gammiris' and 'pepper', the binomial, Sinhalese and English names for the same species. However, synonymous terms also frequently occur in the same language. For example, for the purposes of a particular knowledge base it may be considered acceptable to view 'solar radiation' and 'sunlight' as being synonymous. The ability to identify synonymous terms allows the development of a more compact knowledge base and results in some tolerance for inconsistent use of terms provided that alternative terms are identified as being equivalent.

Four fundamental types of term are encountered in the creation of a knowledge base : objects, processes, attributes and values (§5.4). These terms have different characteristics and roles. It has proved valuable to store them in separate hierarchies. Object and process glossaries may be hierarchical (§3.5). Attribute and value glossaries cannot normally be hierarchically structured but can be linked such that for each attribute in the attribute glossary, there will be a list of the values that that attribute takes in the knowledge base. This provides an opportunity to ensure consistency of application of attributes and the values of those attributes.

3.4 DEFINITIONS OF TERMS

Consistent and precise use of terms has proved to be an important determinant of the utility of a knowledge base and is critical if a knowledge base is to be evaluated and used effectively. Some terms can reasonably be assumed to be understood (depending on the particular domain, knowledge base creator and set of potential users). These form a common basis in experience and language. However, there will be terms for which precise understanding cannot be assumed. So, mechanisms to improve the precision of the content of the knowledge base by explicitly recording definitions of component terms are important both to the developer in producing a complete and consistent knowledge base and to the user in accurately interpreting the content of the knowledge base.

Definitions may be implicit in a set of intermediate statements of fact in a knowledge base. For example, some or all of the following statements (from fieldwork in Sri Lanka, Walker and Southern, 1992) might be combined to provide a definition for the term 'Jak' :

```
Jak is an edible fruit.
Artocarpus heterophyllus is the binomial name for Jak.
Jak is grown throughout the humid tropics.
Jak is an evergreen tree.
Jak grows up to 20 metres tall.
Jak is a member of the family Moraceae.
```

However, these statements may be dispersed through a large knowledge base. Instead of relying on dispersed intermediate statements of fact it is desirable to have an explicit and accessible definition for each term, demanding a separate mechanism for the definition of keywords in the intermediate knowledge base.

It has not proved necessary to generate a definition for every term used in the knowledge base. Instead, only words occurring in glossaries (i.e. words that have been identified as keywords) are considered to require definition.

The process of creating definitions of keywords in the knowledge bases is diagramatically stylised in Figure 3.3.

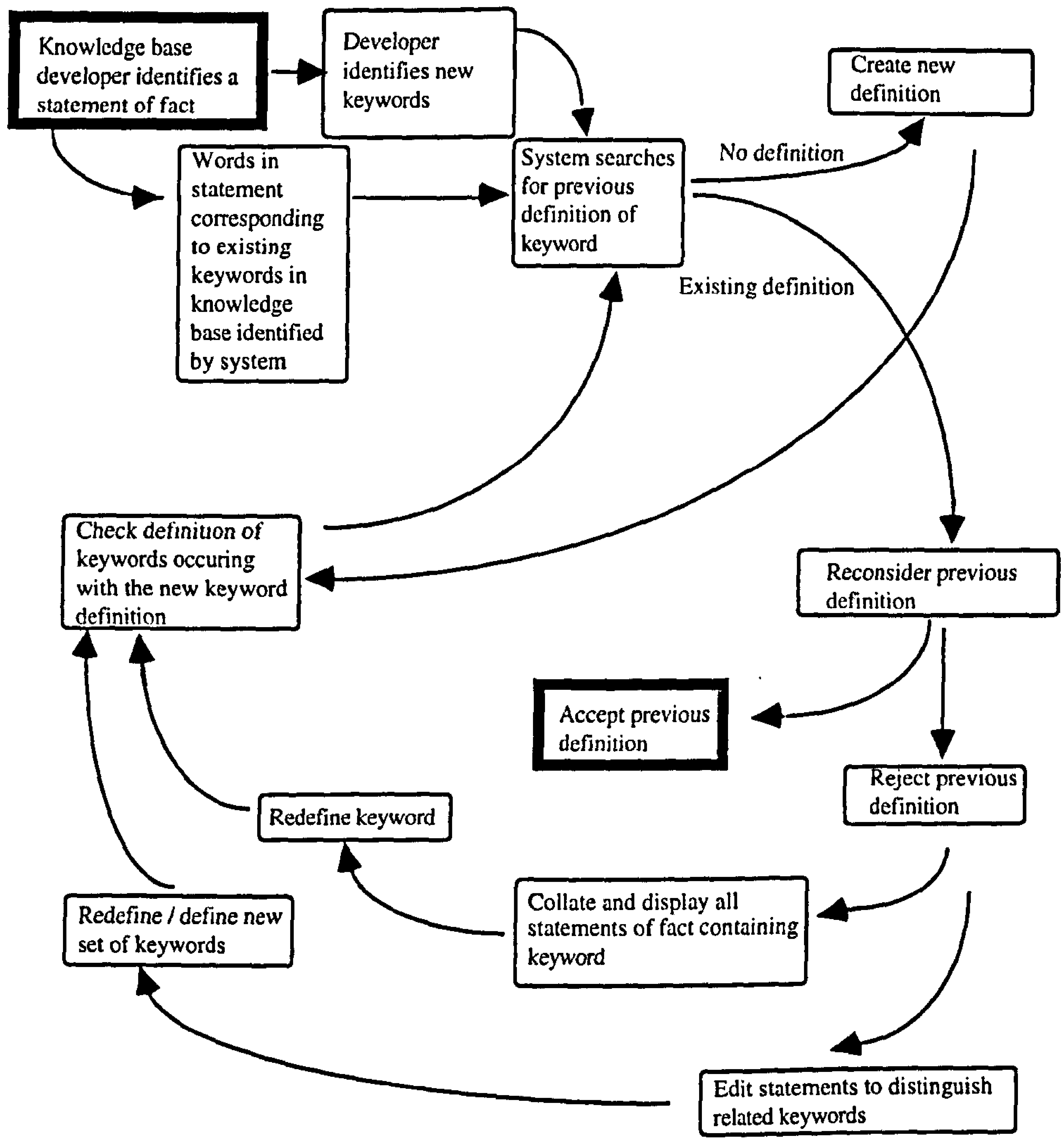


Figure 3.3 A procedure for defining terms (start and end points are identified in bold).

Definitions are rarely achieved by a single statement, usually demanding a more complex characterisation. Experience has suggested that representing definitions as a set of statements of fact results in a more explicit articulation of the meaning of a term than unrestricted natural language description.

The set of types of statement of definition in Table 3.3 were used to generate the set of template statements of definition in Table 3.4.

Using this approach, a definition can be generated by developing appropriate combinations of template statements of definition (which correspond to statements of fact).

Precise definition of all the contents of a knowledge base has proved impossible, not least because ecological knowledge includes terms that are vaguely defined or undefinable and others for which a consensus of understanding cannot be assumed. While some definitions are precise, unambiguously stating what is and is not included in the term, others do no more than characterise the use of a term.

Where the template statements of definition have been applied on a trial basis (§7.2.4) they have been considered to help reduce ambiguity.

Table 3.3 A set of statements of definition (adapted from Werner and Schoepfle, 1987a)

Attributive : Term defined in terms of distinctive characteristic Function : Term defined as a means of achieving something Spatial : Term defined through description of relation to other objects Operational : Term defined with respect to an action which is its characteristic goal or recipient Comparison : Term defined through similarity or contrast with others Class inclusion : Term defined through inclusion in a hierarchical class Synonymy : Term defined as equivalent to another term Antonomy : Term defined as opposite or the negation of another term Quantitative units : Term defined with respect to units of measurement for quantitative variables Qualitative units : Term defined in terms of the set of qualitative values it may take or as a qualitative variable for a particular attribute

Table 3.4 Template statements of definition (X indicates the term being defined)

X is/are _ X has/have _ X happens after _ X happens at the same time as _ X is/are used to _ X is/are found close to _ X is/are found next to _ X is/are found on top of _ X is/are found below _ X is/are similar to _ X is/are different to _ because _ X is/are a type of _ X means the same _ X is/are the opposite to _ X is/are measured in _ X may take the values _ X is/are a value of the attribute _

3.5. HIERARCHICAL STRUCTURES FOR GLOSSARIES

3.5.1 Advantages of a hierarchical structure

The identification of hierarchical relationships between objects, concepts or events provides a means of:

- capturing the hierarchical nature of knowledge;
- enabling the inheritance of properties between keywords up and down the hierarchy, thereby facilitating the development of a compact knowledge base without a loss of information; and
- facilitating hierarchically structured exploration of the knowledge base.

The first two of these are expanded below. The third is discussed in §3.9.

(i) The hierarchical nature of knowledge

Description and explanation of ecosystems tends to be hierarchical. This reflects both the way in which knowledge is held and used, and fundamental attributes of agroecosystems (Conway, 1985).

Ecologists tend to perceive ecosystems as being complex and tend, through a systems perspective, to rationalise some of this complexity through a hierarchical classification of the system, such that statements at any one level are summarised as fewer units with less detail in the level above and explained by more statements at greater detail in the level below (Allen and Star, 1982).

The simplest example of a hierarchical structuring of ecological knowledge is the taxonomic relationship between species. Binomial taxonomy reflects the theoretically predicted evolutionary relationships between species. This means that the binomial taxonomy contains knowledge that describes relations that can, in principle, be verified through real-world observations.

There is evidence to suggest that classification is a universal process in rationalising the human environment and that classification as a means of perceiving the environment will tend to be hierarchical. For example, Berlin (1973) argues that the naming of plants and animals in folk systematics is essentially the same in all languages and can be described by a small number of principles. Most taxa in natural folk taxonomies are members of one of five ethno-biological classes : the unique beginner (e.g. all living things); life forms (e.g. tree, herb, fish), generic (e.g. hickory, maple, cottonwood); specific (e.g. white oak, sugar maple) and varietal (e.g. baby lima bean, butter lima bean). Names referring to these types can be divided into two linguistic categories, primary names (generic and life form taxa) and secondary names (specific and varietal taxa). Life form and generic taxa are usually labels (i.e. one word) whereas specific and varietal names are generally binomial (the generic name being modified by an adjective for the species, and the species descriptor without the generic name but modified by a further adjective for the variety). Exceptions are explainable. For example, where mononomial (i.e. one word) species names are found they are usually 'polysemous with the superordinate genus' (i.e. the same name is used to describe the genus and its most widespread and important species - the type species). This universal approach to identifying the relationship between plants or animals is hierarchical. Furthermore, Berlin (1973) identifies a growing body of evidence to show that the fundamental taxa within folk classifications are frequently close to scientific taxonomy.

Taxonomies consist of simple sets of knowledge that are clarified through hierarchical structuring. These advantages may be extended to more complex sets of knowledge about ecosystems. Indeed, the complexity of systems is often found to be hierarchical (Simon, 1962). Hierarchy theory, an attempt to resolve some aspects of system complexity (Allen and Starr, 1982), has been applied to the theoretical consideration of ecosystems (O'Neill *et al.*, 1986) and the evolution of ecosystems (Burns *et al.*, 1991) and has been proposed as a means of improving predictability (Allen and O'Neill, 1991).

Hierarchy theory remains restricted to theoretical ecology and is either not sufficiently useful to encourage, or not sufficiently defined to allow, validation through experimentation. Irrespective of its utility, the application of hierarchy theory to ecology illustrates an application of a hierarchical perspective to the study of ecosystems.

(ii) Compacting the knowledge base through properties of inheritance

Hierarchies provide a means of increasing the parsimony of the knowledge base because they allow knowledge to be recorded at its most general level of application, but used to consider more specific instances, through consideration of the hierarchical relationship between terms. If, for example, 'wheat', 'barley', 'maize' and 'fava beans' are identified as being annual crops within a hierarchy, the information that annual crops only live for one year is best recorded as generic information about annual crops, rather than for each type of annual crop. This can impact on the size and, therefore, tractability of the knowledge base. Compare Tables 3.5 and 3.6. Each captures the same information; in Table 3.5 this information is explicitly stated, in Table 3.6 it is more implicitly captured because it is assumed that a mechanism exists for recognising the hierarchical nature of the taxonomic statements, therefore all the statements not explicitly stated can be deduced by applying the general rules to lower orders of the hierarchy.

Table 3.5 A complete statement of a set of statements of fact

Crops are economically useful
Legumes are economically useful
Root crops are economically useful
Cereals are economically useful
Chick peas are economically useful
Pigeon peas are economically useful
Cow peas are economically useful
Crops are deliberately cultivated
Legumes are deliberately cultivated
Cereals are deliberately cultivated
Root crops are deliberately cultivated
Chick peas are deliberately cultivated
Pigeon peas are deliberately cultivated
Cowpeas are deliberately cultivated
Crops are plants
Legumes are plants
Cereals are plants
Root crops are plants
Chick peas are plants
Pigeon peas are plants
Cowpeas are plants
Legumes photosynthesise
Chick peas photosynthesise
Pigeon peas photosynthesise
Cowpeas photosynthesise
Legumes have roots
Chick peas have roots
Pigeon peas have roots
Cowpeas have roots
Legumes have leaves
Chick peas have leaves
Pigeon peas have leaves
Cowpeas have leaves
Legumes transpire
Chick peas transpire
Pigeon peas transpire
Cowpeas transpire
Legumes are crops
Root crops are crops
Cereals are crops
Chick peas are legumes,
Pigeon peas are legumes
Cowpeas are legumes

Table 3.6 A compacted statement of knowledge based on a hierarchical structuring

Crops are economically useful Crops are deliberately cultivated Crops are plants Legumes photosynthesise Legumes have roots Legumes have leaves Legumes transpire Legumes are crops Root crops are crops Cereals are crops Chick peas are legumes Pigeon peas are legumes Cowpeas are legumes

In knowledge base creation to date, the knowledge base developer has identified the most generic level at which knowledge can be recorded. However, in some instances it may be possible to implement an inference mechanism that uses deductive reasoning to identify generic rules about a set of entities from specific instances of those entities, thereby moving knowledge up the hierarchy and making it more generically useful.

Suppose, for example, that in a particular knowledge base:

- four crop species are classified as annual crops;
- these four are the only annual crops represented in the knowledge base; and
- all are recorded as only living one year.

An appropriate mechanism might use this information to propose that annual crops generically only survive one year. This knowledge can be recorded and new records of the survival of annual species checked against this hypothesis in an attempt to find instances disproving the suggested statement of fact.

3.5.2 Representing hierarchical relationships

Hierarchical information is captured by identifying a parent-daughter relationship between two terms. In principle it is not necessary to specify the nature of this relationship. So, hierarchical ordering can simply provide a mechanism for indexing the keywords in a set of statements of fact.

In practice it has proved more productive to explicitly state the meaning of a hierarchical relationship. By doing so the hierarchy produced becomes a more meaningful representation of source knowledge and provides a resource that can be more flexibly used in automatic reasoning. The following types of hierarchical linkage are identified:

```
A is a type of B,
A is a part of B,
A is an example of B.
```

'Is an example of' identifies instances of a class and is, therefore, not relevant in representing generic and explanatory ecological knowledge.

The 'is a type of' and ' is a part of' links may identify relationships between objects, for example:

```
Gliricidia is a type of leguminous tree
The stamen is a part of the flower
```

or processes, for example:

```
Splash erosion is a type of soil erosion.
```

The hierarchical relationships between objects or processes can be captured in standard intermediate statements of fact. However, because of the special role of information about the hierarchical relationship between terms, it has proved useful to view the identification of these relationships as a separate task from the development of standard intermediate statements of fact. In the process of creation of the intermediate knowledge base, hierarchies are created which, as a minimum, contain all the keywords identified in the object or process glossaries. Object glossaries will frequently have a complex hierarchical structure. Process hierarchies may often be flat but can, in principle, be hierarchical.

Defining a keyword hierarchy for a set of keywords and topics is a familiar sorting task. Hierarchical classifications are a universally intuitive process widely used in ethnographic research (Werner and Schoepfle, 1987 a and b), for example in deriving species taxonomies (Berlin 1973, 1978) and soil taxonomies (Benfer and Furbee, 1990). This approach has been successfully applied to generating community consensus on soil taxonomies (Alison Southern, in preparation).

If the keyword does not fit into a hierarchy it can be placed in a separate hierarchy. An ability to apply more than one hierarchy to a set of statements of fact is important given that there will be more than one way of classifying that many sets of statements. This ability demands, in turn, an ability to hold a single keyword in two or more separate hierarchies. The resulting set of hierarchies can be visualised as being laid on top of one another, with common junction points resulting in a network rather than a simple tree (in a tree there will be a single route between two nodes, in a network there may be two or more). For example, knowledge elicitation in Nepal during the creation of a knowledge base on tree fodder has revealed at least three separate systems of classification of fodder trees in regular use. A species may be *chiso* or *obaṇo*, *rukho* or *malilo*, *posilo* or *kam posilo* (Thapa, in preparation). These three classifications can be captured in a single network but are better captured in three separate trees that may be used in conjunction (Figure 3.4).

The representation of hierarchical relationships as networks rather than tree has significant implementational implications in using the resultant hierarchy.

3.6 MANAGING KEYWORDS

The preceding three sections have discussed different aspects of the management of keywords in the creation of an intermediate knowledge base. Intermediate knowledge bases developed in accordance with the specifications developed in this chapter have the following features.

- All keywords in an intermediate statement of fact in a knowledge base appear at least once in the glossaries.
- Attributes, objects and processes are stored in separate glossaries.
- Synonymous terms are attached to keywords in glossaries.
- Synonymous terms do not occur elsewhere in the glossary.
- All the keywords are defined.
- Object glossaries are hierarchically structured where appropriate.
- Process and attribute glossaries may, exceptionally, be hierarchically structured.
- Where alternate hierarchical classifications occur, a glossary may be divided into a set of separately developed hierarchical trees which together can be used as a hierarchical net.
- All the links within the hierarchical tree mean either 'is a type of' or 'is a part of'.

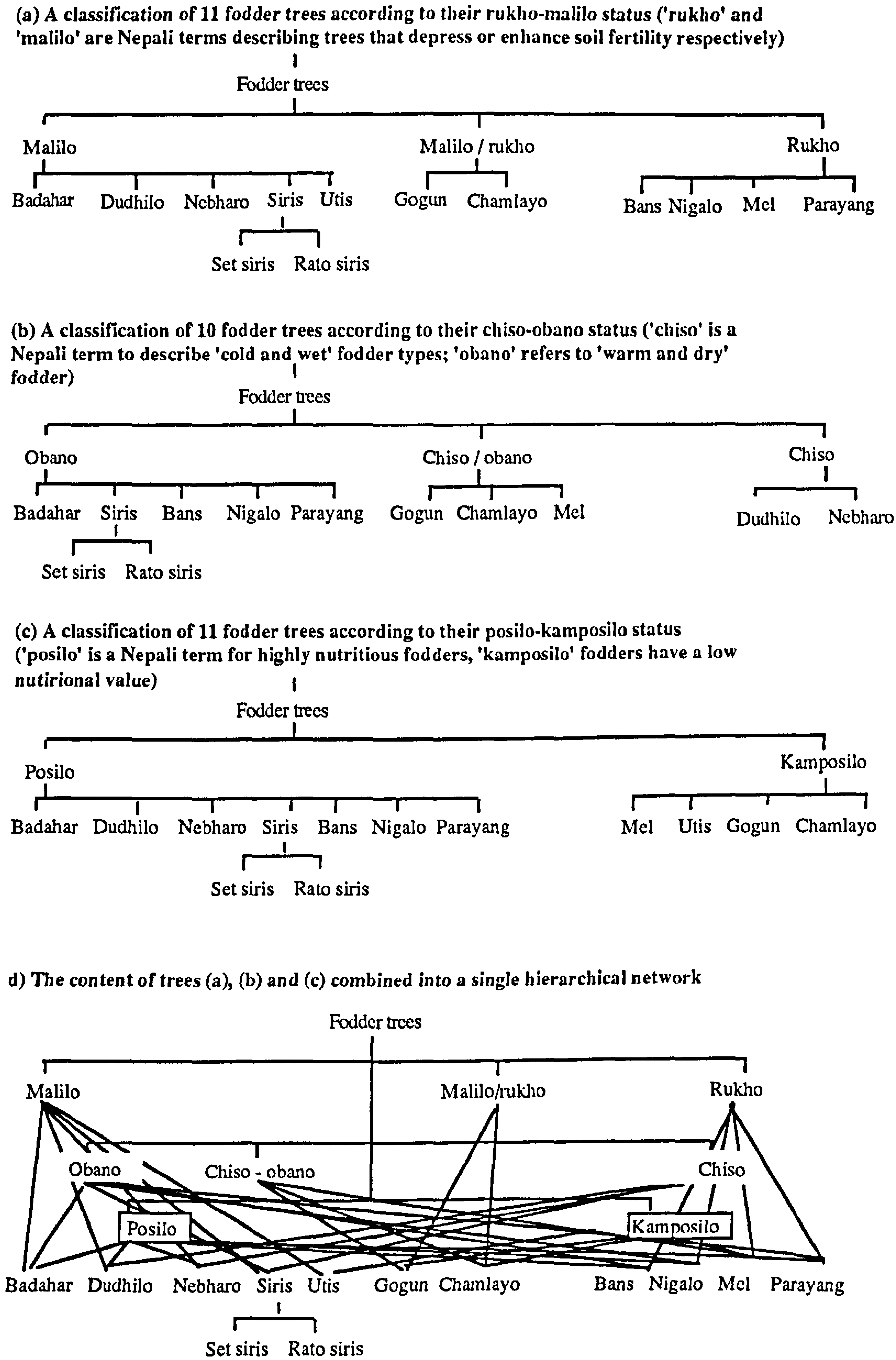


Figure 3.4 Three separate hierarchical trees (a,b,c) (from Thapa, in preparation) for three separate classifications of the same set of species and the three classifications integrated into a single diagram (d).

3.7 MANAGING SOURCE INFORMATION

In §3.2.5, the value of recording information about the source of a statement of fact was identified. The specification for an intermediate statement of fact includes the identification of the source(s) of the knowledge and the date(s) of articulation. This information plays a role in some types of evaluation of the knowledge base. However, a record of the name of the informant and date of the interview is only meaningful if the name and date recorded provide a cross reference with more detailed information about both. This information could be held completely independently of the intermediate knowledge base. However, its inclusion in the knowledge base provides for more flexible knowledge retrieval by source, including the retrieval of knowledge from a group of sources sharing a particular set of characteristics. Two mechanisms for storing and managing this additional information were proposed: a mechanism for recording primary information about each source recorded in a standard fashion, and a mechanism for recording secondary information about the source that can be customised for the individual application.

(i) Recording primary source information

In the specification developed in §3.2.7, each intermediate statement of fact is tagged with the name(s) of the source of the statement and the date of articulation. It was proposed that further primary information be stored in the knowledge base in association with each of these combinations. Knowledge may be articulated through interview or through written material. The primary information associated with an interview was the name of the source, interview date, gender of the source and age of the source, on the basis that these four attributes are widely used in evaluating a knowledge in relation to source. The primary information associated with written material was author(s), date, title and publication details (if any).

Table 3.7 shows the primary source information recorded for four of the informants interviewed by Balaram Thapa during knowledge acquisition in Nepal (Thapa, in preparation).

Table 3.7 Examples of primary source information

Name	Interview date	Age	Gender
Nidhi Nath Khanal	18.12.92	23	male
Maha Nanda Timilsina	18.12.92	39	male
Lok Nath Khanal	20.12.92	27	male
Chandra Pd Khanal	20.12.92	43	male

(ii) Recording secondary source information

Recording primary source information provides a resource for evaluating the content of a knowledge base according to source. However, assessment of the distribution of knowledge within the knowledge base (and, by extrapolation, within source communities) may demand the use of information extending beyond the content of these primary records.

A trial implementation of the information source tool as a part of the prototype knowledge acquisition tool TEAK (Appendix A), provided both a more extensive list of categories of information about a source and the opportunity for user-defined categories. Figures 3.5 and 3.6 illustrate the type of information recorded as a result (from Southern, in preparation). The resulting opportunity for customisation proved valuable in practice (Thapa and Walker, 1992). However, the variability in source information recorded using this flexible approach introduced significant implementational difficulties in developing automated means of evaluation, and in making comparisons both within and between knowledge bases.

As a result, development of a specification for a tool for recording secondary source information had to address the following issues.

- The mechanisms for recording primary source information perform an important set of functions in the management and interpretation of the knowledge base and should not be compromised by customisation.

- In order to allow comparability, any customised set of secondary source information must be consistently recorded across the knowledge base. Any alterations to the secondary information recorded should either be disallowed or retrospectively applied across the existing contents of the knowledge base.

- The utility of secondary information recorded is significantly dependent on the consistency with which it is recorded, which in turn depends on the manner in which the records are structured. Clear rules must be identified.

- In contrast to the above, the construction of record sheets for secondary information should be as intuitive as possible. The envisaged users cannot be expected to digest and understand complex rules of construction.

- Primary records are recorded for every instance of a source. This results in repetition. Where an individual is interviewed on five occasions, details about that individual, for example age and gender, should not have to be repeated for each occasion; only information that is specific to each interview or reference (interviewer or title for example) need be recorded for each instance, and the result should be accessible as secondary information.

- The secondary information recorded should be accessible for use as criteria in searches of the knowledge base.

In order to enable the knowledge base developer to customise the secondary information stored in a knowledge base, it must be possible to customise any of four different structures, 'Reference - author details', 'Reference - publication details', 'Informant details' and 'Interview details'. Up to four customised dialogues, one for each of these categories, are allowable for each knowledge base.

Choice of one of these options then triggers a dialogue allowing the user to customise a dialogue for recording information in addition to that which is already recorded as primary information.

The dialogue created is then be triggered after either the existing 'Informant' or 'Reference' dialogue, as appropriate, to elicit all secondary information on a source.

Use, particular automated manipulation, of the secondary source information depends on explicit statement of the information. So, for example, the attributes in the secondary source information may have qualitative or quantitative sets of values. Specifying the units for quantitative values, or the complete set of permissible qualitative values (where applicable) and ordering of qualitative values (where applicable) will result in a much more tractable set of information.

Figure 3.6 summarises the procedure for creating a secondary information dialogue.

A further set of dialogues are required for circumstances under which the user wishes to alter a secondary information dialogue for which information has already been recorded. This would display the existing attributes and their units or sets of values as appropriate such that the user can edit or delete them. Additionally the user would be able to specify further attributes. All changes should be retrospective such that any records already entered are automatically amended where possible (e.g. deleting any attribute and associated values from previous records) and eliciting further information for all previously recorded sources where necessary (e.g. eliciting values for each source for a new attribute).

Whose

Name	Abeysinghe A M G
Age	39
Gender	male
Education	GCE AL
Landholding	1/2 acre homegarden, 3/4 other land
Location	Senerathwela
Altitude	
Aspect	
Occupation	retired teacher & private tuition
Activities	
Homegarden age	Sept 1991, replacing previous from 1930s
Livestock	10 chickens
How long here?	generations

Notes

other land somewhere nearby

Sort Cards   OK

When

Interview Details

Informant(s)	Abeysinghe
Interviewers(s)	Alison and Niyantha
Date	18.6.92
Time & Duration	9.00am, about 1.5 hrs
Location	Senarathwela
Main topics	Soil classification
Techniques	slipsorts, interview and garden walk

Notes

This was less successful than anticipated. Abeysinghe seemed to lose interest after a while and failed to completely allocate soils to like groups. We felt that perhaps he was under-confident of the subject

OK

Figure 3.5 An example of user-customised source information about the informant (top) and the interview (bottom).

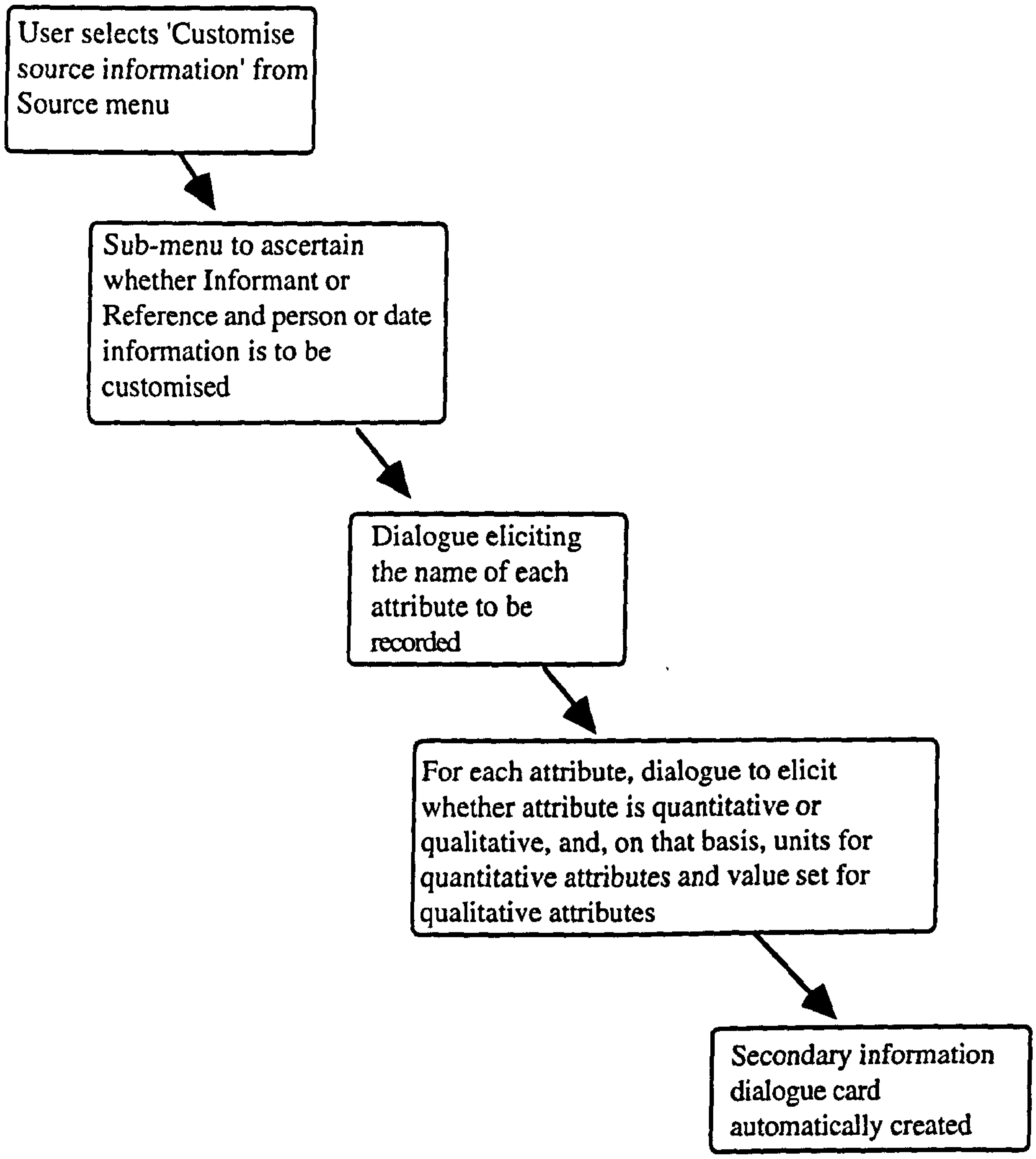


Figure 3.6 Creating a secondary information dialogue

3.8 A SUMMARY OF THE CONTENTS OF THE INTERMEDIATE KNOWLEDGE BASE

The specifications developed in this chapter result in an intermediate knowledge base comprising six components :

- a set of intermediate statements of fact tagged with keywords, source and conditions;
- a set of keyword glossaries containing at least one reference to every keyword;
- a definition associated with each keyword;
- synonymous terms identified for each keyword;

- hierarchical information for object keywords and, where appropriate, process keywords; and
- primary and (optionally) secondary information associated with each source in the knowledge base.

The relationship between these components is summarised as an example in Figure 3.7.

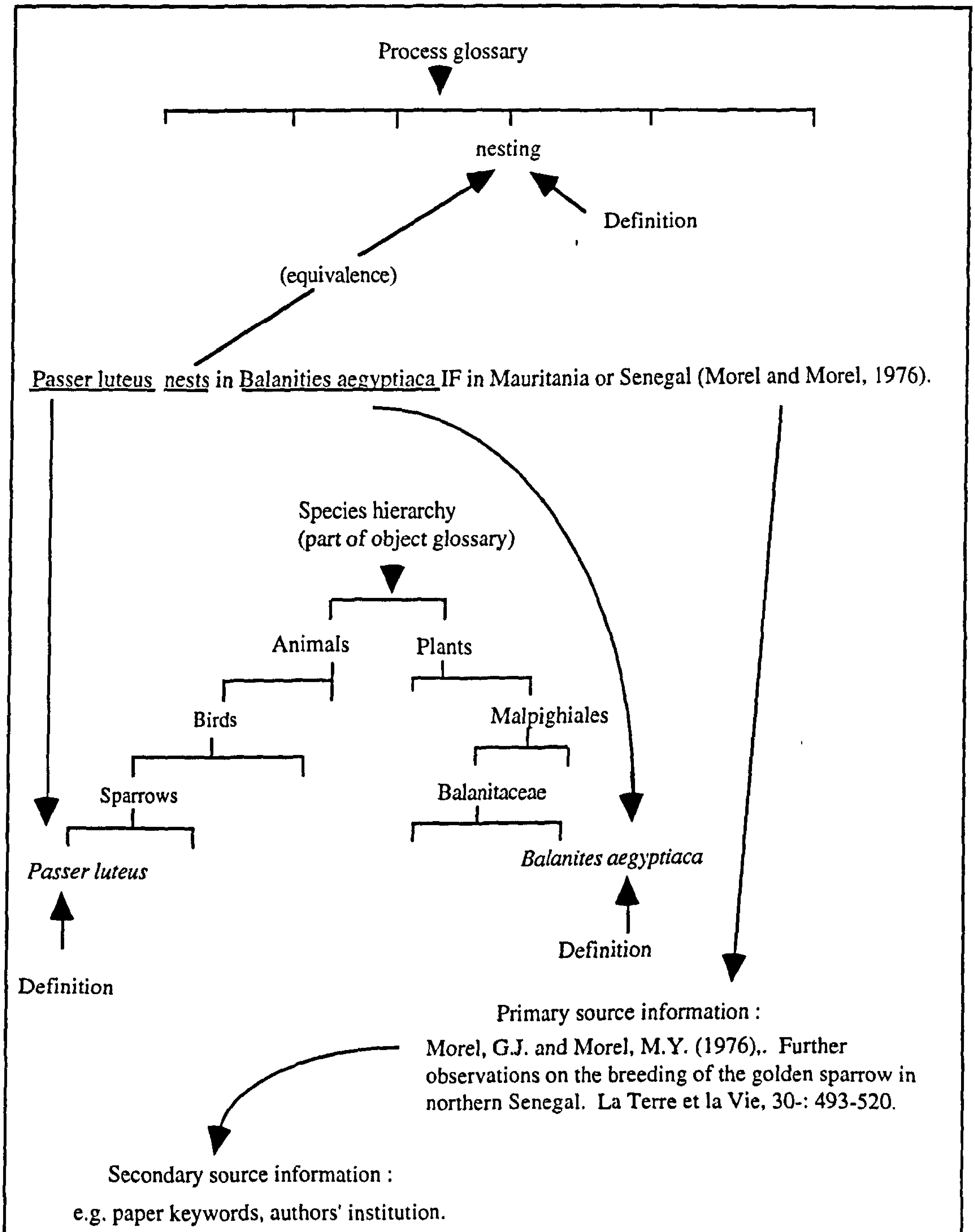


Figure 3.7 A summary of the structure of the intermediate knowledge base. A single statement and the information associated with that statement are shown

3.9 ACCESSING THE CONTENTS OF THE INTERMEDIATE KNOWLEDGE BASE

An intermediate knowledge base is only of any value if its contents can be efficiently and flexibly accessed to facilitate further knowledge elicitation on the basis of consideration of the knowledge already elicited (§7.4) and to enable formal representation (Chapter 4).

At the most basic level, the contents of the knowledge base may be accessed without any synthesis of, or reasoning with, that knowledge by the system. This is referred to here as 'read back' and may occur on-line or through the generation of printed output.

On-line read back can be achieved both through access to disaggregated statements of fact. A list of the individual statements of fact in the knowledge base and text listings of hierarchies and glossaries is of relatively limited utility in accessing the content effectively and efficiently. However, experience has shown that this is an important mechanism because it helps the user to get an impression of the size of the knowledge base and the range of its content. This overview demystifies the content of the knowledge base, which is an important process in encouraging rational use.

Read back is, in practice, an inefficient means of accessing the content of the intermediate knowledge base. The knowledge bases created to date have contained several hundred statements of fact, indeed, up to 3000 in some cases. As a result, automatic search mechanisms were required for the user to access statements relevant to a particular purpose. Keywords and source information provide useful criteria for selecting subsets of the knowledge base. Combinations of these characteristics ('What knowledge elicited from women about leaf age and tree fodder value is held in the knowledge base?') provides a mechanism for further discrimination.

Where keywords and source information are tagged to each statement of fact, automatic searches can be implemented by selecting a combination of these terms linked by the operators AND and OR.

The representational advantages of defining the relationship between keywords have already been discussed (§3.5). This definition also has practical advantages for searches because it enables the use of properties of inheritance in searches.

Three search strategies are possible where the hierarchical relationships between keywords are known: keyword-only searches, keyword and descendants, and keyword and family.

The keyword-only search does not make use of the hierarchical relationships identified. Only statements in which a keyword occurs are abstracted.

The keyword and descendant search is justified by the premise that statements containing keywords below a specified keyword in a hierarchy will contain information that is relevant to that keyword. A keyword and descendants search on 'fodder trees' in Figure 3.8 will abstract statements about Badahar, Tulo pate dudhilo and so on. Because these are species of fodder tree, statements about these species may be relevant to a consideration of fodder trees.

A keyword and family search is based on the premise that statements containing 'parent' as well as 'daughter' keywords may contain relevant information. A keyword and family search on millet in Figure 3.8 will abstract statements referring to Mugale and Chhakre but also statements referring to crops, on the basis that generic information on grain crops also provides information about millet.

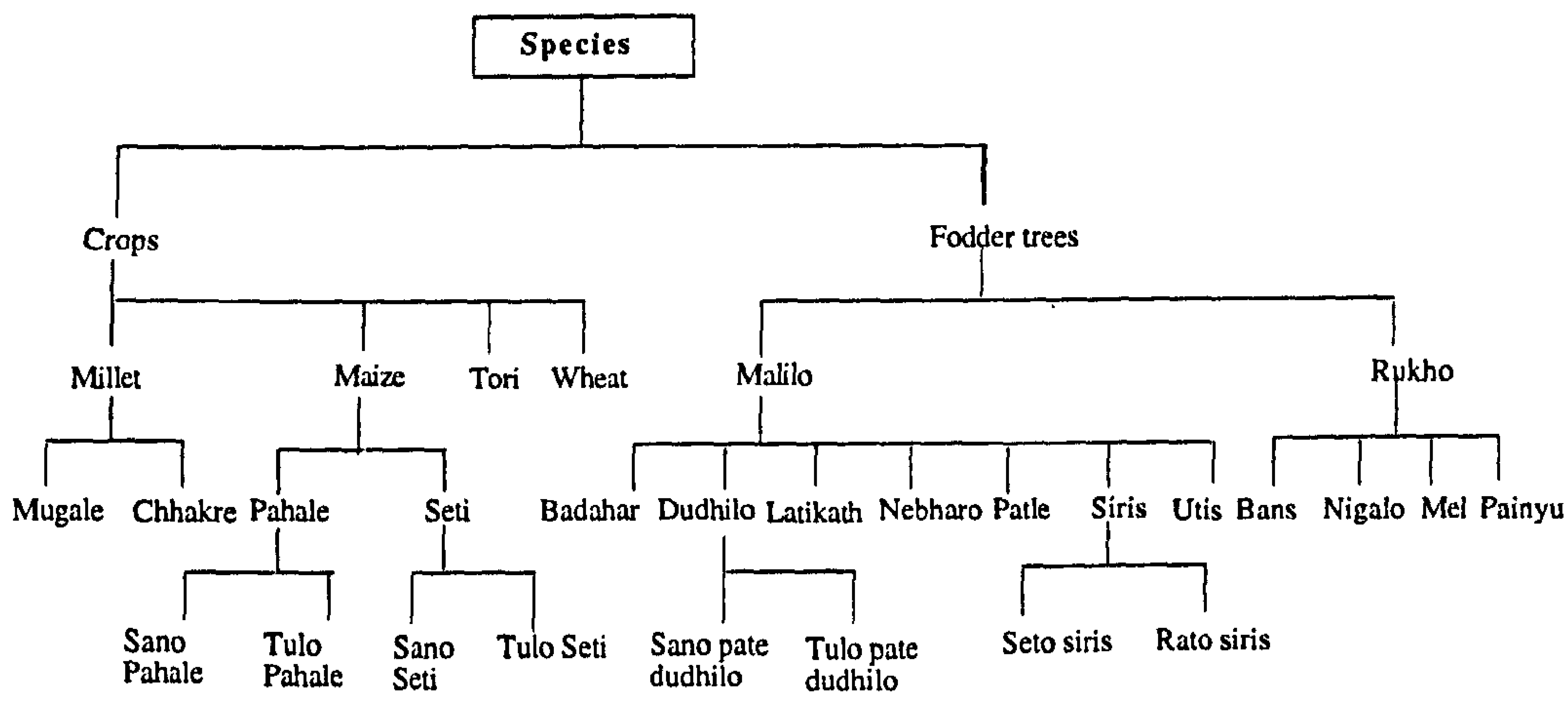


Figure 3.8 An example of a simple, hierarchical, object glossary, derived from knowledge acquisition in Nepal. Adapted from Balaram Thapa (in preparation).

A further refinement to the simple AND and OR search mechanism takes advantage of equivalence. So, a search on the keyword ' pepper' will result in the abstraction of statements containing the terms gammiris or *Piper nigrum* (the Sinhala and binomial names for pepper respectively) where these have been identified as equivalents.

Accessing the contents of the intermediate knowledge bases through the use of keyword searches has proved to be a powerful and successful utility.

3.10 EVALUATION AND CONCLUSIONS

The intermediate knowledge base structure developed in this chapter has resulted in the creation of knowledge bases of disaggregated statements of fact which can be flexibly explored and provide a sound basis for subsequent formal representation.

While it has been stressed that the intermediate knowledge base is a step in the process of the creation of a formal knowledge base, successful investigation of local knowledge systems has been undertaken by three postgraduate research students (da Costa, 1993; Holmes, 1993; Southern in preparation) by developing intermediate knowledge bases on the basis of the specifications developed in this chapter (as implemented in TEAK, Appendix A, and AKT, Chapter 7). The specifications not only provide a sound basis for subsequent formal representation but also the capture of ecological knowledge about agroforestry in a form in which it can be rigorously evaluated.

The identification and separate treatment of statements of fact defining the hierarchical relationship between objects or processes, statements of definition and the relationship between attributes and values provides some structure to the knowledge base above the level of the individual statement of fact and encourages the development of a coherent and consistent knowledge base. Linkages between many statements of fact within the knowledge base can be deduced making the knowledge base more than a database.

The resulting resource may be used for generating a description or explanation of the behaviour of agroforestry practices integrating a large number of statements of fact. The understanding generated by the creation of a description or explanation can be appropriately viewed as an articulation of an argument about how the system functions. Each argument represents a particular model of the topic in question. Creating a knowledge base of disaggregated statements of fact therefore provides a resource for, amongst other things, generating arguments or models. Any approach to the creation of knowledge bases without disaggregating the knowledge content into discrete units would result in the creation of a single model or argument of the domain. As a result, disaggregation of statements of fact is an inescapable necessity. However, disaggregation does lead to loss of information (§3.2.5).

The degree to which sets of individual statements of fact can be automatically reasoned with depends on the degree to which the deducible linkages between knowledge (§5.1) are captured. Constructing a useful knowledge base by entering a set of disaggregated statements of fact into the computer relies on the knowledge base developer's ability to :

- structure statements of fact consistently and in such a way that implicit linkages are captured (§5.1); and
- remember statements of fact already entered that provide explanation for an individual statement of fact.

Humans are not particularly adept at tasks which involve remembering and sorting large amounts of repetitive information. As a result, constructing knowledge bases in this way can lead to the creation of knowledge bases of a relatively low quality. However, humans are good at constructing syntheses. This skill provides an alternative means of achieving the same aims.

If the user is able to incrementally develop an argument about the domain on the basis of the knowledge collected for entry in the knowledge base, linkages will be captured in a useful way in the argument. Furthermore, considerable assessment and interpretation of the source knowledge, leading to further knowledge elicitation to fill in gaps and resolve ambiguities, will be demanded by the need to develop a coherent argument. The resulting set of statements of fact is, therefore, likely to be of a higher quality and the knowledge base may be more useful. As a result, mechanisms that allow knowledge to be entered into a knowledge base through the creation of an argument or set of arguments based on the synthesis of knowledge from a number of sources provide an alternative approach to the creation of knowledge bases as disaggregated statements of fact.

The specification described has provided a framework for the development of intermediate knowledge bases that provide a sound basis for subsequent formal representation. The process of formal representation is discussed in the next chapter. Approaches to capturing arguments as linked sets of statements of fact are discussed in Chapter 5.

CHAPTER 4 FORMAL REPRESENTATION AND AUTOMATIC REASONING

4.1 INTRODUCTION

Formal representation of a statement of fact involves expressing the content of that statement according to syntactic rules which precisely define the relationship between terms. Such statements can be compiled on a computer and, in combination with appropriate inference mechanisms, used in automatic reasoning. While formal statements are less expressive than natural language, they are more tractable for automatic reasoning. Justification for the process of formal representation of knowledge depends, therefore, on the utility of the automated reasoning that it allows. However, experience has also demonstrated an incidental but valuable utility in that the process of formal representation significantly improves the rigour with which knowledge is analysed and expressed in the creation of knowledge bases (Chapter 7).

The restricted syntax of an approach to formal representation is codified through the creation of a formal grammar. This chapter describes the development of a grammar for the formal representation of intermediate statements of fact about the ecology of agroforestry. The grammar was developed by the Department of Artificial Intelligence of the University of Edinburgh as a result of collaboration with the author and colleagues from the University of Wales, Bangor. Specification of the grammar was based on assessment of intermediate knowledge bases and consideration of appropriate automated reasoning tasks for use with that knowledge.

4.2 DEVELOPMENT OF THE FORMAL GRAMMAR

Development of the formal grammar demanded consideration of :

- the intermediate knowledge bases already created (in order to facilitate the development of an *expressive grammar* that allowed the majority of ecological knowledge about agroforestry to be formally represented);

- the implications of considerations in the design and implementation of inference mechanisms for the structure of formal statements of fact (to facilitate the development of a grammar allowing the creation of *useful formal knowledge bases*); and

- the ability of potential users to apply the formal grammar (in order to develop a *practical grammar*).

(i) An expressive grammar

A primary requirement of the formal grammar was to allow the formal representation of at least a significant proportion of ecological knowledge about agroforestry. For this reason, the development of the grammar was based on detailed consideration of sets of ecological knowledge represented in intermediate knowledge bases.

A total of eight different research projects involving the creation of intermediate knowledge bases have been undertaken by postgraduate research students with support from the author[3]. Two of these, Southern (in preparation) and Thapa (in preparation), were used as a primary resource in the development of the grammar.

The knowledge base developed by Southern (in preparation) was created during fieldwork through intermediate representation within a word processing package (Walker and Southern, 1991). The knowledge base developed by Thapa (in preparation) made use of TEAK (Thapa and Walker, 1992 and Appendix A).

Initial evaluation of the content of these knowledge bases in generating a specification for the formal grammar was strongly influenced by the previous development of template sentence structures (§3.2.2). Classifying the contents of these knowledge bases according to the template sentence structural types demonstrated that all the intermediate statements of fact could be classified as falling into at least one of the template sentence categories (causality, definition, taxonomic status, spatial relationship, temporal relationship, attribute statement, comparative statement, statement of certainty, rationale, location for action, function, or means of achieving a goal).

This process indicated that all these categories might usefully be supported by the formal grammar. However, several categories were either already effectively formalised in the intermediate knowledge base structure (Chapter 3) or deemed irrelevant to the creation of useful knowledge bases about the ecology of agroforestry. Taxonomic status was already catered for in the creation of object sort hierarchies (§3.5), with the result that they do not occur in the main body of the knowledge base. Similarly, in principle, statements of definition (which in practice tend to overlap with taxonomic statements and attribute statements) are also catered for in the intermediate knowledge base structure.

[3] Garde (1992), da Costa (1993), Holmes (1993), Jinadasa (in preparation), Kilahama (in preparation), Preechanpanya (in preparation), Southern (in preparation), Thapa (in preparation).

The problems associated with the elicitation of statements of certainty have already been discussed and the decision not to include specific statements of certainty for each statement of fact justified (§3.2.4). For this reason it was decided that it was not necessary for the grammar to support the expression of statements of certainty.

The last four categories in the list (rationale, location for action, function, and means of achieving a goal) all relate to management of the agroforestry practice. While management type statements had been elicited during fieldwork, it was argued that the knowledge bases should be restricted to containing strictly explanatory ecological knowledge, on the basis that this was a more flexible and generally applicable resource than management or technical knowledge (Walker *et al.*, 1991) and that mixing the two in the same knowledge base would be confusing, resulting in a less tractable and useful formal knowledge base. As a result it was decided that the formal grammar would not be required to express management type statements. This decision has had implications for the use of the approach as a whole (§7.7.3).

In summary, causal statements, statements of spatial relationship, statements of temporal relationship, attribute statements, and statements of comparison were identified as the set of statement types to be supported in the formal grammar. Evaluation existing knowledge bases showed that causal statements and attribute statements predominated and were thus considered to be particularly important.

(ii) Useful formal knowledge bases

The usefulness of the formal knowledge base depends on the extent to which it can be used, in combination with inference mechanisms, in automated reasoning to fulfil useful tasks. This in turn depends on :

- what tasks are useful;
- what tasks can usefully be facilitated by automatic reasoning; and
- what inference mechanisms can be designed and implemented that facilitate these tasks.

In the current context, the primary concern is with the last. The development of potentially useful formal knowledge bases has been primarily constrained by the specification and implementation of inference mechanisms. While the range of inference mechanisms that can be envisaged is large, the range of practicably implementable mechanisms is small. This is because both the conceptual adaptation of existing algorithms (which are much fewer in number than hypothetical algorithms) to a new context, and practical implementation, are demanding of limited research resources.

The remaining set of statement types (causal relationship, spatial relationship, temporal relationship, attribute statement, comparative statement) were, therefore, considered in relation to the reasoning tasks for which implemented inference mechanisms could be developed with the resources available. It was proposed by researchers at the Department of Artificial Intelligence that while approaches to temporal reasoning are available (Shaon and Goyal, 1988), they are both conceptually demanding and, as yet, of limited practical utility. However, inference mechanisms allowing reasoning with causal statements, spatial relationships, attribute statements and statements of comparison were considered to be both useful and feasible.

As well as providing a means of identifying the statement types that could be used in automated reasoning and, therefore, usefully supported in the formal grammar, detailed consideration of the design of inference mechanisms had an impact on the detailed specification of the grammar.

(iii) A practical grammar

The approach developed in this thesis is intended for use by agroforestry professionals. As a result, the formal grammar had to be sufficiently straightforward for use by researchers with a minimum training requirement. Experience in training researchers in the use of template sentences (§3.2.4) had already illustrated some of the conceptual difficulties encountered by some users in structuring knowledge. While the practicability of the grammar could only be evaluated after it had been specified, the need to develop a relatively simple grammar was understood throughout the process of specification.

The detailed criteria that were used in the development of the formal grammar were generated on the basis of an evaluation of the content of intermediate knowledge bases and a consideration of the impact that the structure of the grammar has on the implementation of inference mechanisms and the of impact that the structure and function of proposed inference mechanisms has for the development of the grammar.

The impact of the structure and function of inference mechanisms on the grammar is considered in more detail in the next section through description of the development and evaluation of a prototype inference mechanism. The formal grammar resulting from these deliberations is described and discussed in §4.4 to §4.6.

4.3 DEVELOPMENT AND EVALUATION OF A PROTOTYPE INFERENCE MECHANISM

Prior to the development of the grammar, a prototype inference mechanism (the HyperNet causal reasoning tool) was developed. The development and evaluation of this mechanism, described in this section, provided a means of informing and evaluating the decision-making process in the specification of the formal grammar.

The HyperNet causal reasoning tool was implemented within HyperNet (Walker *et al.*, 1993), a software toolkit developed to facilitate the diagrammatic generation of knowledge bases (Chapter 5) and now incorporated into the AKT software (Appendix A).

4.3.1 Task specification

Consideration of the content of the intermediate knowledge bases developed through fieldwork (§4.1) has shown that ecological phenomena can often be effectively described through the creation of linked sets of causal statements of fact. These sets may often be complex. As a result, the provision of inference mechanisms that elucidate the chain of causal relations between events is a powerful means of enabling a user to explore the content of a knowledge base.

Given a large and complex set of formally represented causal statements of fact about ecological phenomena, the user may wish, for example, to pose questions such as :

- What can cause a change in pest population levels? (Or what different causes of a change in pest population level are there?)
- What are the consequences of increased soil fertility?
- What causal chain(s) exist between increased soil fertility and plant species diversity?

The HyperNet causal reasoning tool was designed to facilitate the user in addressing these questions through reference to a knowledge base comprising a set of linked causal statements.

4.3.2 Inference mechanism

The three questions above are examples of the three generic types of question supported by the causal reasoning tool. These involve identifying, respectively :

- the set of causal chains in a linked set of causal statements resulting in a user-selected event;
- the set of causal chains in a linked set of causal statements resulting from a user-selected event; and
- the set of causal chains in a linked set of causal statements connecting two user-selected events.

These three types of question were all be addressed by using a 'depth first' search, using backward-chaining in the first instance, forward-chaining in the second and either in the third (Bratko, 1990). This section illustrates the depth first search for elucidating the consequences of a user-selected event. This same procedure was slightly adapted in order to provide mechanisms for identifying the causes of an event or the causal relationship between two events.

In identifying all the recorded consequences of an event, forward chaining was used to find and print all the causal chains starting at the user-specified event (i.e. an increase or decrease in a particular attribute, or the occurrence of a particular event) and end at the end of a causal chain (i.e. where a change in an attribute is recorded but is not recorded as having any consequences itself). The 'depth first search' approach is analogous to finding a path through a maze by making a decision at any junction and continuing down that path until a dead end is met and then returning to the last junction at which an untested decision can be made. This will eventually result in the identification of the path through the maze or, in this case, the causal chain linking the user-specified start and the particular chain end. To be useful it is necessary: that all the causal chains between the start and all chain ends are identified; that it is not possible to get stuck going around a loop in the causal network; and that the causal chains are only printed once identified (instead of printing out all failed explorations as well). This was achieved using the following algorithm :

```
search([Cause|Path], Final):-
  causal_link(Cause, Caused),
  not on(Caused, Path),
  search([Caused, Cause|Path], Final).

search([Cause|Path],[Cause|Path]):-
  not causal_link(Cause,_).
```

This algorithm builds up a record of the causal chain ('Path') as a list as it is elucidated. The user-selected starting point is initially set as 'Cause'. The knowledge base is searched to find the first recorded causal linkage which starts at 'Cause' and causes 'Caused'. This is represented here using the structure :

```
causal_link(Cause, Caused),
```

for example :

```
causal_link('increase in soil temperature', 'increase in
germination rate').
```

It is then checked that this causal link has not already been explored by ensuring that caused is not already in the Path that is being incrementally built up. If it has not been then Cause is added to the recorded Path and the cycle is repeated by converting Caused into a Cause and trying to identify a link in which it is a cause.

If no link in which Cause is a cause is found or if Caused is already in the recorded Path (i.e. the search is going around a loop in the causal network) then the second 'search' clause is called. Here, provided that there is no link recorded in which Cause is a cause (i.e. the Path is not a loop), the Path is handed over to Final. This a path between a user-specified event and the end of a chain in the causal network and can now be displayed to the user.

This algorithm in itself only identifies the first causal chain between the user-selected event and the first chain end in the recorded network. To identify all the chains between the user-selected event and all the chain ends, the Prolog 'findall' facility (Johns, 1991) is used :

```
findall(Final, search([selected_start], Final), List_of_paths).
```

This call generates the complete set of causal chains that start at the user-selected event and end at a chain end. This set of chains can then be reported to the user.

4.3.3 An example of application

Figure 4.1 represents a simple set of causal knowledge (the use of diagramming as a means of representing knowledge is discussed in Chapter 5). Each link (arrow) within this diagram corresponds to a causal linkage. As a result, each link can be captured as a statement of fact with each node (box) in the diagram representing a keyword. The causal reasoning tool can be used to explore a knowledge base capturing these causal linkages.

Figure 4.2 shows the primary user interface for the causal reasoning tool, prompting the users to specify the causal linkages that they wish to explore.

The options selected in the dialogue shown in Figure 4.2 can be read as :

```
In what way(s) does an increase in tobacco growing influence the
number of acacia grown?
```

Once the user has specified the question, the algorithms described above are applied. Once all linkages have been elucidated, the tool reports to the user (Figure 4.3).

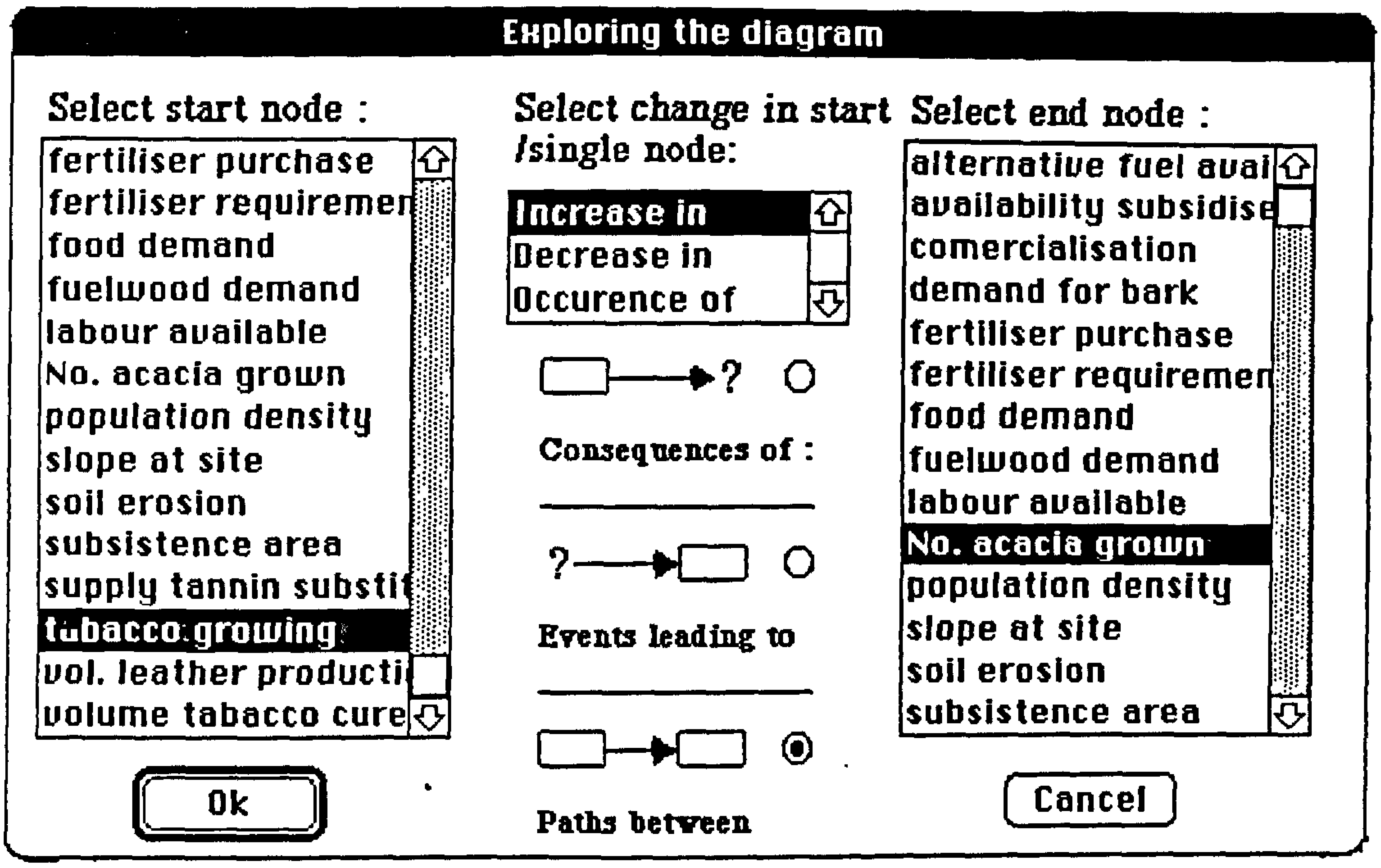


Figure 4.2 The primary causal reasoning interface. The column on the left lists all the keywords (nodes) that act as a causal agent in any of the statements of fact. The column on the right lists all those keywords (nodes) that are acted on in a causal statement of fact[4]. One of the reasoning options (Consequences of, Events leading to, Paths between) , a keyword from the left-hand column, a keyword from the right-hand column or one from both (as appropriate) and (where appropriate) the change in the causal agent are selected.

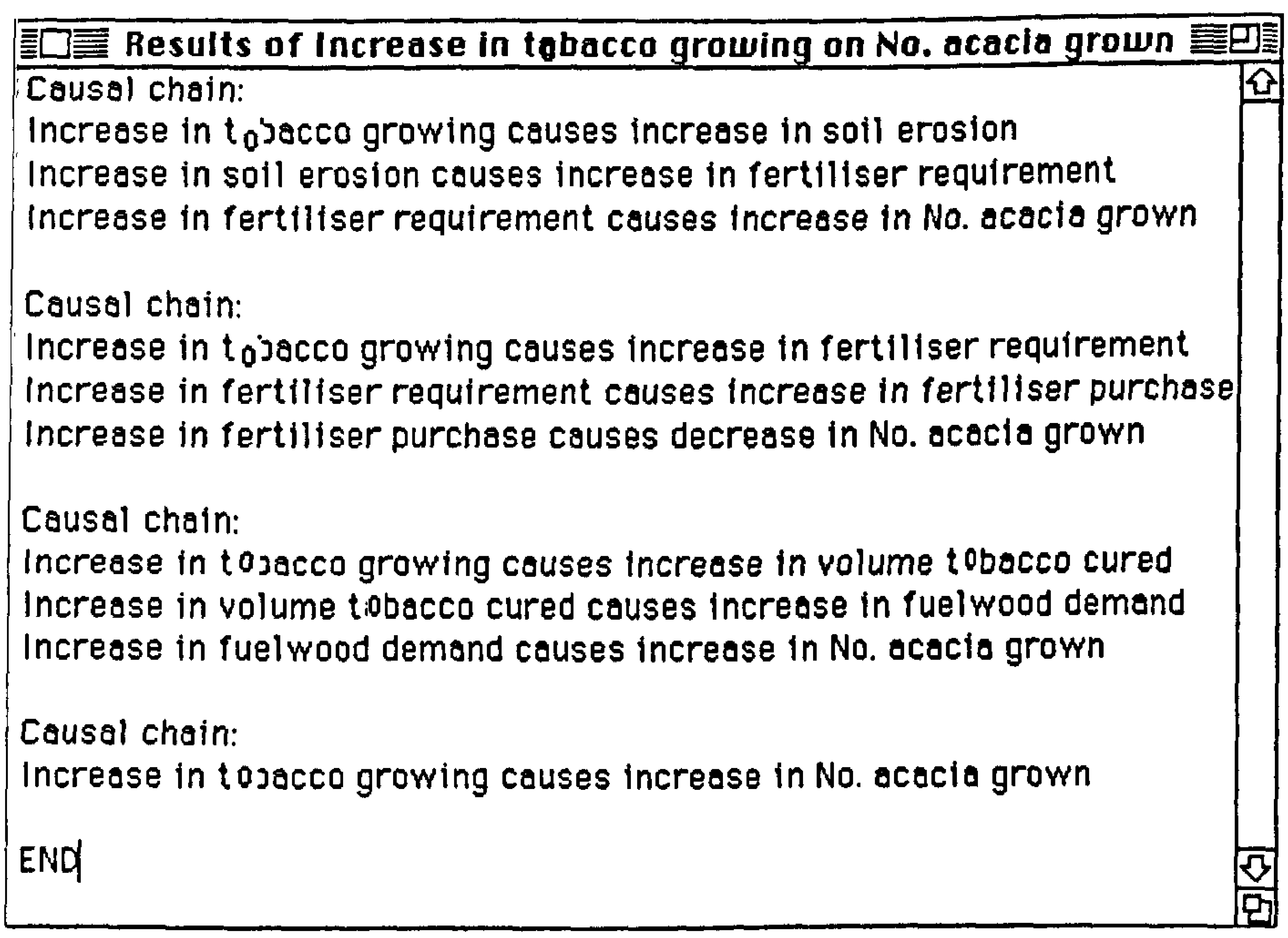


Figure 4.3 Causal reasoning - reporting to the user.

[4]As the list of keywords for each is longer than the space provided, only a portion of each list is visible.

4.3.4 Limitations to the causal reasoning tool.

The causal reasoning tool provides a successful means of exploring causal chains of events. Furthermore, the approach used is based on linkages, which may be recorded in statements of fact or deducible between statements of fact.

However, there are limitations to this trial implementation, in particular it does not take conditional information into account during reasoning. The importance of conditional information associated with statements of fact has already been discussed (§3.2.4). Causal relationships will usually occur only under defined circumstances. Furthermore, different causal relationships between the same objects or processes may occur under different circumstances. The implementation described makes no use of conditional information. However, a more useful reasoning process is provided by using causal information to assess whether a relationship will occur, either by reference to other recorded knowledge or by eliciting conditional information from the user, particularly in order to distinguish between alternative events.

Furthermore, the causal reasoning tool is limited to explicitly recorded relationships and cannot use those that can be deduced from them. The statements :

```
An increase in the rate of soil erosion causes an increase in
the rate of decline of soil fertility
```

and

```
A decrease in the rate of soil erosion causes a decrease in the
rate of decline of soil fertility
```

mean exactly the same thing. Failing to recognise this equivalence, as in the current implementation, means that linkages may be missed. So, for example, the fact that an increased soil organic matter reduces the decline in soil fertility by reducing soil erosion would not be deduced from the statements of fact :

```
An increase in the organic matter content of the soil causes a
decrease in the rate of soil erosion.

An increase in the rate of soil erosion causes an increase in
the rate of decline of soil fertility.
```

This means that an arbitrary decision must be made on the way in which a causal relationship is stated and consistently applied across the complete set of statements of fact if these are to be successfully reasoned with.

4.3.5 Conclusions

The work described in this section provided valuable experience in the design and implementation of an inference mechanism of the basis of existing algorithms. Given that causal relationships have been found to play a central role in the representation of ecological knowledge, this experience facilitated the development of specifications for the representation of causal statements in the grammar and the development of a navigation tool as a component of the AKT software (Appendix B). It also highlighted further requirements of an approach to formal representation, for example the need for conditional information to be formally represented and therefore available for incorporation into the reasoning process.

4.4 THE FORMAL GRAMMAR

On the basis of the considerations outlined in §4.2 and §4.3 a formal grammar was development by the Department of Artificial Intelligence of the University of Edinburgh. This was then tested for expressiveness and practicality in application by a joint group of researchers from the University of Wales, Bangor, and the Department of Artificial Intelligence through the formal representation of a set of 50 intermediate statements of fact selected from the intermediate knowledge bases developed by Thapa and Southern (§4.2). Each of these 50 statements was selected on the basis that formal representation using the grammar was likely to be particularly demanding.

This process resulted in minor modifications to the grammar and a set of guidelines on its application. It also provided a basis for the specification of inference mechanisms by the Department of Artificial Intelligence (Appendix B).

The resulting grammar is stated in Table 4.1. A summary explanation of the grammar is provided in this section before a more detailed exploration of its application in §4.6.

Formal statements comprise a set of terms linked by, and ordered in relation to, other special terms that form part of the formal language. This linkage and ordering provides information on the way in which the elements of the statement are related and therefore enables a semantic interpretation of the syntax of the statement. All the terms in the statement relate to four elements : objects, processes, attributes and values. The representation of knowledge in terms of this set of elements was largely based on experience in the creation of node-and-link diagrams as an approach to knowledge representation (Chapter 5).

Table 4.1 The grammar of the formal language : The grammar defines the range of forms that a formal statement of fact can take. The symbol '==>' means 'can take the form', so the structure on the left-hand side of this symbol can be developed by creating the structure on the right-hand side of the symbol. Capitalised words are defined elsewhere in the grammar, words in **bold** form part of the language and therefore can only be used for the purposes for which they are used in the grammar, X and Y represent atoms that are defined by the user

Sentence ==> Statement **if** Conditions
Sentence ==> Statement
Statement ==> Causal_statement
Statement ==> Attribute_statement
Statement ==> Negative_attribute_statement
Statement ==> Link_statement
Statement ==> Comparison_statement
Negative_attribute_statement==> **not** (Attribute_statement)
Link_statement ==> **link**(Link_type, Object, Object)
Link_statement ==> **link(influence, Thing, Thing)**
Thing ==> Process
Thing ==> Object
Comparison_statement ==> **comparison** (Attribute, Object1, Comparison_type, Object2)
Comparison_statement ==> **comparison**(Attribute, Process1, Comparison_type, Process2)
Conditions ==> Attribute_statement
Conditions ==> Negative_attribute_statement
Conditions ==> Link_statement
Conditions ==> Comparison_statement
Conditions ==> Conditions **or** Conditions
Conditions ==> Conditions **and** Conditions
Attribute_statement ==> **att_value**(Object, Attribute, Value)
Attribute_statement ==> **att_value**(Process, Attribute, Value)
Attribute ==> X *{atom appearing in Attribute glossary}*
Value ==> **increase** | **decrease** | **change** | **nochange**
Value ==> **range**(X,Y) *{ X and Y are atoms }*
Value ==> > | < | = | <= | >= X *{ constant or variable appearing elsewhere in statement }*
Value ==> X { atom }
Comparison_type ==> **greater_than** | **less_than** | **same_as**
Causal_statement ==> Cause **causes** Attribute_statement
Cause ==> Attribute_statement
Cause ==> Process
Cause ==> Object
Process ==> **process**(X) *{X is an atom - the name of a process }*
Process ==> **process**(Object,X) *{X is an atom and name of a process }*
Process ==> **process**(Object,X,Object2) *{X is an atom and name of a process }*
Object ==> X *{X is a variable appearing in the sort specifications}*
Object ==> **part**(Object,Part)

Objects, processes, attributes and values are the four elements of a formal statement of fact that can be defined by the user. The combination of these four elements with a set of special terms used in the formal language (for example 'causes', 'if', 'and', 'or', 'comparison' and 'link') provides the basis for the formal representation.

Objects and processes can be considered to be the fundamental elements around which the formal statement is structured. Objects are normally physical items in the real world like trees and crops, but may be conceptual, for example, niche or wet season. Processes (or events) describe changes or fluxes in the real world, for example the process of soil erosion describes the loss of soil, and the process of germination describes the change in a seed from dormancy to active growth. The objects in a formal statement are represented as atoms. Atoms are the name of the object; in Prolog atoms are represented in lower case only. So, 'cow', 'tree' and 'hill' are objects represented as atoms. In fact, objects in formal statements are usually classes of items (for example 'cows', 'trees' and 'hills'). Like objects, processes are named (represented as atoms). In some circumstances a process is not associated with any particular objects in the statement; alternatively, it may be related to an identified object or may provide linkage between two objects.

Statements will frequently incorporate measurable attributes of objects or processes, for example the height of a tree or the rate of soil erosion. Attributes have particular values, for example the height of a tree might be 'tall' or '5 m' and the rate of soil erosion might be 'high' or '40 kg ha^{-1} yr^{-1}'. As with objects and processes, attributes are represented as atoms. Values are generally represented as atoms as well, although some special representations of values are included for particular purposes.

The most basic combination of these elements occurs in an attribute-value statement. An attribute-value statement for an object takes the form :

```
att_value(Object, Attribute, Value),
```

for example:

```
att_value(tree, height, tall).
```

An attribute-value statement for a process takes the form:

```
att_value(Process, Attribute, Value),
```

for example:

```
att_value(process(decomposition, leaf), rate, slow).
```

The term 'comparison' enables the comparison of the value of an attribute for two objects or processes. The term 'causes' allows a causal relationship between an attribute-value statement, object or process and another attribute-value statement to be captured. The term 'if' allows conditionality to be captured whereas the terms 'and' and 'or' allow multiple conditions to be specified for a statement of fact. The term 'part' allows a particular part of a specified object (for example the roots of wheat) to be represented. The term 'not' can be used to capture negation (e.g. sycamore is not a native British species). Finally the term 'link' allows relationships between objects or processes other than causal or comparative links to be captured (e.g. cows eat grass).

Some elements of statements may be represented as atoms that have special meanings and can be used in particular ways. For example a value may be represented by the special terms 'increase', 'decrease', 'change', 'no_change' and 'range'. Comparison types must be one of three terms, 'greater_than', 'less_than' or 'same_as'. Finally, under some circumstances a link type may be represented by the special term 'influence'.

The grammar was designed with a particular emphasis on causal, comparison and attribute-value statements. However, careful application means that many statements that might more intuitively be categorised as, for example, spatial, temporal or management type statements can still be formally represented.

4.5 FORMAL STATEMENTS AND THE SORT HIERARCHY

The inference mechanisms developed for use with the formal knowledge base make use of a sorted order logic (Robertson *et al.*, 1991). For this to be possible, all objects occurring in a formal statement must also appear in the sort hierarchy of the knowledge base. The sort hierarchy is derived from those hierarchies of objects in the intermediate knowledge base in which the links in the hierarchy mean 'is a type of'.

4.6 FORMAL REPRESENTATION OF INTERMEDIATE STATEMENTS OF FACT

A detailed description of the application of the formal grammar and justification for its structure is best achieved through example.

The grammar in itself provides the user with little guidance in the process of formal representation. For this reason, a set of guidelines was created (Walker, 1993, in Sinclair *et al.*, 1993) on the basis of trial application of the grammar on some 50 intermediate statements of fact. This section follows the structure of those guidelines, being divided into four parts :

- re-evaluation of the intermediate statement;
- identification of fundamental elements;

- identification of statement structural type; and
- creation of the formal statement.

The examples used are derived from the intermediate knowledge bases created by Southern (in preparation) and Thapa (in preparation).

(i) Reassessing the intermediate statement

Experience shows that statements of fact in the intermediate knowledge base are often ambiguous, incomplete, or actually represent more than a single statement of fact. The first step in the process of formal representation is to reassess the individual statement of fact, ensuring its meaning is clear and sensible. In particular, it has frequently been found that statements are compound and such that it is necessary to break them down into a set of individual statements of fact.

In principle, it is only necessary to alter the intermediate statement of fact if it is inaccurate or if it is compound. The intermediate statement and formal statements remain linked in the knowledge base, but the formal statement will not always capture all of the meaning of the intermediate statement. For this reason, so long as it is not actually wrong, the intermediate statement is kept intact.

While it is stressed that formal representation is justified by the automated reasoning that it makes possible, the rigorous consideration of the meaning of an intermediate statement that is demanded by the process proves to be of very significant utility in its own right (§7.7.3).

(ii) Identifying the elements in the statement

The second step in formal representation is to identify the objects (and parts of objects), processes, attributes and values in the statement.

This step is important because knowledge must be explicitly and unambiguously stated in creating formal statements that provide a robust resource for automatic reasoning. By contrast, natural language and, by extension, intermediate statements, tend to contain implicit elements. A tentative distinction can be drawn distinguishing (a) an implication resulting from inadequate intermediate representation (where domain-specific knowledge may be needed to achieve more explicit representation) and (b) an implication that, it might be assumed, the user of the system would understand but which is unacceptable for formal representation.

The fact that the statement :

```
Karsu causes sickness in cattle
```

means that cattle become sick if they eat Karsu is an example of the former because the user needs to know that Karsu is eaten. By contrast, and as an example of the latter, it might be assumed that any user of the system can interpret the statement that :

```
Utis is a tall tree
```

as being about the height of the tree, while formal representation demands that this be explicitly stated.

Attributes have been found usually to be implicit. For example in the statement :

```
Loam is very fertile
```

'loam' (a type of soil) is an object, 'very fertile' is a value but it is not immediately clear what this value is a measure of (i.e. the attribute to which it refers). However, the attribute must be identified in formal representation. 'Soil fertility' might be suitable in this case.

By contrast, objects, processes and values have only occasionally been found to be implicit, usually only where intermediate representation is inadequate (as with the implicit process in the first example above) or where representing statements that stretch the use of the grammar. For example, the statement :

```
Rainfall is maximum in January
```

does not contain an explicitly stated object. Rainfall is taken to be a process. In other circumstances, it might be viewed as an object but 'maximum in January' implies that the attribute is rate rather than volume. Rates can only be associated with processes. So, 'maximum' is a value for the attribute 'rate' of the process 'rainfall'. 'January' is a value for an attribute time, or maybe time of year. The attribute 'time' refers to the time at which the event occurs. For the purposes of representation an object must be identified, 'system' is proposed in this case (i.e. the system is the object that experiences the process rainfall).

The statement :

```
Soil erosion reduces soil fertility
```

contains implicit values. Fertility is an attribute of an object 'soil' while soil erosion is a process with an implicit attribute 'rate'; both attributes have implicit values, in this case 'increase' and 'decrease' respectively.

'Part' relationships between objects are also identified at this stage. For example, in the statement :

```
Siris has a light crown
```

Siris (a tree species) and crown are objects, however they are further related in that the crown is a particular part of the object Siris. The identification of 'part of' relationships is similar to the identification of 'type of' relationships between objects in the sort hierarchy and can similarly be used in reasoning.

(iii) Identifying the statement type

The third step in the process of formal representation is to identify which type of formal statement best captures the meaning of the intermediate statement. The identification of structural type has already been discussed (§3.2.2). However, the process can be more demanding during formal representation because some statements which cannot directly be represented in the formal grammar (for example temporal statements) can be represented using the other, existing, structural types.

Formal statements may be of one of five types : a causal statement, a comparison statement, a link statement, an attribute statement or a negative attribute statement. Some intermediate statements may be captured by more than one formal statement type. However, the different types have a differing utility in reasoning with the knowledge base. In general, causal statements are more useful than comparison statements which are more useful than link statements. This is principally because the navigation tool (Appendix B) is particularly designed for use with causal statements but the other two types can, at present, only be reasoned with through use of the generic query tool (Appendix B). Any of these types of statements may additionally have conditional information attached.

Causal statements

In a general sense a causal statement is any in which the value of an attribute is changed.

Given that a casual statement takes the general structure :

```
X causes Y
```

Y will always be a change in the value of an attribute. This change can be captured by using one of the special values 'increase', 'decrease', 'change' or 'no change'. X may also be a change in the value of an attribute (again taking one of the four special values) or may be a process. The grammar also allows objects to be a cause (for example sheep cause soil creep) but this is incomplete, containing implicit information. The complete statement (for example, trampling by sheep causes soil creep) is always preferable, but the feature is retained in the grammar for cases in which it is not known how an object causes a change, but simply that something about its presence does.

So a causal statement can take one of the following three forms:

Attribute statement causes attribute statement, for example :

```
a decrease in stem thickness causes a decrease in stem strength.
```

which can be formally represented as :

```
att_value(stem, thickness, decrease) causes att_value(stem,
strength, decrease).
```

Process causes attribute statement, for example :

```
Soil erosion causes reduced soil fertility
```

which can be formally represented as :

```
process(soil_erosion) causes att_value (soil, fertility,
decrease).
```

Object causes attribute statement, for example:

```
Sheep cause increased soil creep
```

which can be formally represented as :

```
sheep cause att_value(soil_creep, rate, increase)
```

Comparison statements

Comparison statements compare the relative value of either a pair of objects or a pair of processes.

As a result the comparison statement can take one of two forms :

```
comparison (Attribute, Object1, Comparison_type, Object2).

comparison (Attribute, Process1, Comparison_type, Process2).
```

Comparison statements may be self evident, for example :

```
Bamboo grows faster than fruit trees.
```

Frequently, however, comparison is implicit, usually against an implicit 'norm'. The statement

```
Bans leaves decompose slowly
```

can be interpreted as being a comparison with the average rate of leaf decomposition. This kind of implicit comparison is, however, best captured as an attribute value statement, in this case :

```
att_value(part(bans, leaf), 'decomposition rate', slow).
```

The only instances in which implicit comparison may best be represented as comparative statements are those in which there are clearly only two possible circumstances. So, for example, the statement

'Forests with closed canopies cast deeper shade' is a genuinely comparative statement and might be more explicitly stated as :

```
Forests with closed canopies cast more shade than forests with
open canopies.

comparison (depth of shade, closed canopy forest, more than,
open canopy forest).
```

Link statements

Link statements take the basic form

```
link(Link_type, Object1, Object2).
```

Ecological relationships such as :

```
link(eat, cows, grass)

link(pollinate, bees, clover)
```

are good examples of link statements.

The grammar includes one special type of link statement, in which the link type is 'influences'. In this instance the link may be between any combination of objects and processes (normally links are only allowed between objects). Influence relationships are very closely related to causal relationships. However, in an influence relationship there is no information on what attribute of the object or process impacted on, is changed or how. Where there is information on the result of the influence, this should be captured as a causal statement.

Attribute and negative attribute statements

Entire formal statements of fact can be captured as attribute or negative attribute statements where the statement consists of a single object or process and information about the value of an attribute of that object or process. These statements may also occur within causal statements.

For example :

```
Siris has small leaves

att_value(part(siris, leaf), size, small);
```

and

```
Siris does not have big leaves

not(att_value(part(siris, leaf), size, big)).
```

Conditional statements

Causal, comparison, link, attribute and negative attribute statements may all be conditional. Conditions in the formal language can take the form of attribute or negative attribute statements, link statements and comparison statements. Conditions may be linked by AND and OR. Conditions can not be causal.

The structure of intermediate statements allows conditions to be identified (§3.2.4). However, experience shows that they frequently are not, often because conditionality is obscured by the representation as a compound statement. So, for example the statement :

```
Trampling by goats causes soil creep on slopes.
```

is compound and needs to be treated as being conditional :

```
Trampling by goats causes soil creep if the trampling occurs on
slopes.
```

or:

```
Soil creep occurs on slopes if there are goats walking on the
slopes
```

This example illustrates a further point. It is clear that goats cause increased soil creep only if those same goats are on the slope. This is recognised in the formal grammar provided that the name given in each instance is the same; if, however, the statements refers to goats and the condition to goat this linkage is lost.

4.7 CONCLUSIONS

The development of the formal grammar described in this chapter balanced requirements for :

- an expressive grammar;
- a tractable grammar (i.e. suitable for automated reasoning tasks through the development and application of appropriate inference mechanisms); and
- a practical grammar (i.e. one that can be used by agroforestry professionals in creating knowledge bases).

The development of the grammar was iterative and, therefore, provided an indication of its expressiveness. Development was based on testing and modification through the formal representation of 50 intermediate statements of fact from two intermediate knowledge bases. These statements were purposively selected from the total of some 5 000 statements available to provide a rigorous test of the grammar. It was found that all of these statements could be formally represented, although in some cases the correspondence between the intermediate and formal statement was considered to be weak, while in others the formal statements were so complex that inference mechanisms making use of such statements were hard to envisage. Nevertheless, on the basis of this experience it was proposed that the grammar could be used for the formal representation of a significant proportion of ecological knowledge about agroforestry. Further evaluation of the expressiveness of the formal grammar is undertaken in §7.7.3.

The specification of the formal grammar enabled the specification and implementation of the two reasoning tools available in current versions of AKT by Gill Kendon of the Department of Artificial Intelligence of the University of Edinburgh (Appendix B). The utility of these tools, and the tractability of the formal grammar for automatic reasoning are discussed in §7.7.

The involvement of University of Wales researchers in testing and modifying the formal grammar presented here indicated that it is practical and can be successfully applied by professionals without an artificial intelligence background within hours of introduction. The practicality of the process of formal representation with this formal grammar by agroforestry professionals is returned to in §7.7.3.

CHAPTER 5 REPRESENTING LINKED SETS OF STATEMENTS OF FACT

5.1 INTRODUCTION

The intermediate and formal knowledge base specifications developed in Chapters 3 and 4 provide a means of creating knowledge bases of disaggregated statements of fact. These disaggregated statements of fact can be flexibly used in combination through deduction of the linkages between statements. However, these specifications do not provide a means by which the knowledge base developer or user can view the linkages between statements of fact during the creation or exploration of knowledge bases. An effective overview of sets of statements and the linkages between them:

- facilitates the effective elicitation of coherent sets of knowledge;
- facilitates synthesis of knowledge about a system or part of a system;
- facilitates the dissemination of knowledge by providing an effective means of allowing other people access to the knowledge that has been recorded; and
- facilitates decision support by providing a structured means of exploring the implications of options.

Linked sets of statements of fact may be represented textually, but this approach is limited by the fundamental linearity of text. By contrast, diagram-based approaches to the representation of linked sets of statements are not constrained by linearity. In the research described in this chapter, an approach to creating node-and-link diagrams conforming to a defined protocol was developed as a means of achieving an explicit representation with multiple linkages between statements.

5.2 REPRESENTING LINKED SETS OF STATEMENTS OF FACT

5.2.1 Text-based approaches

Linkages between statements of fact may be deducible from the contents of a pair of statements of fact. This may be achieved for any linked pair of statements that state either a binary relationship between two terms or a combination of a binary and attribute-value statement. The deduction of the linkage involves matching the 'tail' of one statement of fact to the 'head' of another. The head of the statement is all that that occurs before the term that defines the relationship (for example 'causes', 'eats', or 'happens after'); the tail is all that occurs after it. Linkage may also be achieved by matching one of a set of conditional statements associated with a statement of fact with the head of a binary relationship or with an attribute-value relationship.

So, for the statement :

```
An increased soil fertility causes an increased weed population
if weed populations are not controlled and crop canopy is not
closed.
```

the text before 'causes' is the head, and the text between 'causes' and 'if' is the tail and may provide linkage with the statement :

```
An increased weed population causes increased evapotranspiration
```

However, linkage may also be achieved through conditions. For example, with the conditional attribute-value statement :

```
Crops do not have a closed canopy if crops are young or crop
density is low.
```

This matching process is used in forward- and backward-chaining inference mechanisms and is a special example of the links between units of information used in HyperText systems. The navigation tool (Appendix B) makes use of this mechanism. In the current context, however, the interest is in providing the user with an overview of the linkages between a set of statements of fact. Linked sets of statements of fact may be represented textually by developing chains of statements of fact linked through tail and head matching, or condition and head matching[5].

Tables 5.1 - 5.3 illustrate this process. Table 5.1 is a natural-language text description of the consequences of deforestation. This explanation is qualitative and descriptive and is representative of the type of knowledge that might be entered into a knowledge base. This argument can be manually disaggregated into the set of statements of fact shown in Table 5.2. The natural-language statement of the argument in Table 5.1 is too complex for an automated disaggregation into a sets of statements of fact. The knowledge can then be represented as a linked sets of statements of fact by identifying a set of chains of explanation with deducible linkages (Table 5.3).

---

[5]Ecological knowledge is often not articulated as a chain of logically linked statemnts. The potential use of the operator BECAUSE in representing explanation, by linking a statement to a set of statements that are asserted to provide an explanation (themselves linked by the operators AND and OR) was discussed in Chapter 3. While these operators could be used in representing linked sets of statements of fact, it has been found that this discourages the rigorous and explicit statement of knowledge and results in specific detailed explanations, but a loss of generality and, therefore, flexibility.

Table 5.1 A text description of some of the consequences of deforestation

Site deforestation results in a decrease in the rate of evapotranspiration, and an increase in the rate of convection. Decreased evapotranspiration and increased convection result in reduced rainfall.

Deforestation also increases both wind and insolation, resulting in hotter days and cooler nights.

By removing the canopy, and, therefore, shade, deforestation means that sun and rain reach the soil unimpeded. This results in 'baking' of the soil and in induration. Furthermore it results in an increase in both the rate of leaching and the rate of soil erosion.

Deforestation also dries the site, resulting in increased chances of burning. Burning results in the destruction of leaf litter, and a resulting reduction in water retention by the soil.

This reduction in soil moisture retention combines with induration and soil baking, hotter days and cooler nights and reduced rainfall to result in a harsher micro climate.

Burning also causes nitrogen and sulphur to be volatilised, nutrients to be made soluble and animal populations to decline, resulting in reduced nutrient cycling. These processes, in combination with increasing leaching and erosion rates serve to reduce soil fertility. A reduction in animal populations also reduces rates of predation.

Decreased predation, reduced fertility and a harsher micro climate all result in pasture degradation.

Decreased predation causes an increase in pest population densities. The plants remaining on the site have little defence against pests, in combination with increased pest densities, pest damage is therefore increased. The pasture is degraded as a result.

Similarly oligoculture resulting from reduced soil fertility both encourages weeds and results in nutrient deficiency. Nutrient deficiency results in pasture degradation.

An increase in weeds increases the competition for nutrients, light and water, resulting in pasture degradation. A harsher environment decreases the number of palatable plants available to cattle. This in turn encourages overgrazing, trampling and soil compaction, all of which result in pasture degradation.

Table 5.2 Disaggregated statements of fact capturing the knowledge in Table 5.1

Deforestation causes a decrease in rate of evapotranspiration.
Deforestation causes an increase in rate of convection.
Deforestation causes a decrease in percentage canopy cover.
A decrease in percentage canopy cover causes decrease in percentage shade.
Deforestation causes drying of the site.
A decrease in rate of evapotranspiration causes a decrease in volume of rainfall.
An increase in rate of convection causes decrease in volume of rainfall.
Deforestation causes an increase in amount of wind.
Deforestation causes an increase in rate of insolation.
An increase in amount of wind causes increase in daytime temperature.
An increase in amount of wind causes decrease in night-time temperature.
An increase in rate of insolation causes increase in daytime temperature.
An increase in rate of insolation causes decrease in night-time temperature.
A decrease in percentage shade causes increase in percentage sunlight to ground.
A decrease in percentage canopy cover causes increase in percentage rain to ground.
Drying causes burning.
Increase in percentage rain to ground causes increase in rate of leaching.
Increase in percentage rain to ground causes increase in rate of erosion.
Burning causes destruction of micro organisms.
Burning causes volatilisation of soil nitrogen.
Burning causes volatilisation of soil sulphur.
An increased percentage rain to ground causes increased induration.
Burning causes an increase in nutrient solubility.
Burning causes a decrease in animal population densities.
Decrease in animal population densities causes decrease in rate of nutrient cycling.
Burning causes destruction of leaf litter.
Destruction of leaf litter causes a decrease in percentage water retained in soil.
Decrease in percentage water retained in soil causes increase in harshness of micro climate.
Induration causes an increase in harshness of micro climate.
Increase in amount of wind causes increase in harshness of micro climate.
Increase in rate of insolation causes increase in harshness of micro climate.
Decrease in volume of rainfall causes increase in harshness of micro climate.
Increase in rate of leaching causes decrease in soil fertility.
Increase in rate of erosion causes decrease in soil fertility.
Increase in nutrient solubility causes decrease in soil fertility.
Volatilisation of nitrogen causes a decrease in soil fertility.
Volatilisation of sulphur causes a decrease in soil fertility.
Destruction of micro organisms causes a decrease in soil fertility.
Decrease in animal population densities causes decrease in rate of predation.
Decrease in rate of predation causes increase in pest numbers.
Increase in harshness of micro climate causes decrease in number of palatable plant species.
Increase in harshness of micro climate causes increase in weed population size.
Decrease in soil fertility causes oligoculture.
Decrease in soil fertility causes nutrient deficiency.
Oligoculture causes an increase in weed population size.
Increase in weed population size causes increase in rate of competition for nutrients.
Increase in weed population size causes increase in rate of competition for water.
Increase in weed population size causes increase in rate of competition for light.
Increase in pest numbers causes increase in rate of pest attack.
Oligoculture causes a decrease in plant defences against attack.
Decrease in plant defences against attack causes increase in rate of pest attack.
Decrease in number of palatable plant species causes increase in grazing rate.
Decrease in number of palatable plant species causes increase in compaction rate.
Decrease in number of palatable plant species causes increase in trampling rate.
Increase in grazing rate causes pasture degradation.
Increase in compaction rate causes pasture degradation.
Increase in trampling rate causes pasture degradation.
Increase in rate of competition for water causes pasture degradation.
Increase in rate of competition for nutrients causes pasture degradation.
Increase in rate of competition for light causes pasture degradation.
Increase in rate of pest attack causes pasture degradation.

Table 5.3 The textual statements of fact capturing the knowledge in Table 5.1 represented as sets of statements linked as chains

Chain 1. Deforestation causes a decrease in the rate of evapotranspiration. -> A decrease in the rate of evapotranspiration causes a decrease in rainfall. -> A decrease in rainfall cause an increase in the harshness of the micro climate

Chain 2. Deforestation causes an increase in the rate of convection. -> An increase in the rate of convection causes a decrease in rainfall. -> A decrease in rainfall causes an increase in the harshness of the micro climate.

Chain 3. Deforestation causes an increase in the amount of wind. -> An increase in the amount of wind causes an increase in daytime temperature. AND An increase in the amount of wind causes a decrease in night-time temperatures AND An increase in the amount of wind causes an increase in the harshness of the micro climate

Chain 4. Deforestation causes an increase in the rate of insolation. -> An increase in the rate of insolation causes an increase in daytime temperature AND An increase in the rate of insolation causes a decrease in night-time temperatures. AND An increase in the rate of insolation causes an increase in the harshness of the micro climate.

Chain 5. Deforestation causes a decrease in canopy cover. -> A decrease in canopy cover causes a decrease in shade. -> A decrease in shade causes an increase in the sunlight reaching the ground

Chain 6. Deforestation causes a decrease in canopy cover. -> A decrease in canopy cover causes an increase in rainfall reaching the ground. -> A decrease in the rainfall reaching the ground causes an increase in induration. -> An increase in induration causes an increase in the harshness of the micro climate.

Chain 7. Deforestation causes a decrease in canopy cover. -> A decrease in canopy cover causes an increase in rain reaching the ground. -> An increase in rain reaching the ground causes an increase in the rate of erosion. -> An increase in the rate of erosion causes a decrease in soil fertility.

Chain 8. An increase in the rain reaching the ground causes an increase in the rate of leaching. -> An increase in the rate of leaching causes a decrease in soil fertility.

Chain 9. Deforestation causes vegetation drying. -> Vegetation drying causes an increased risk of burning.

Chain 10. Burning causes destruction of leaf litter. -> Destruction of leaf litter causes a decrease in the water retained in the soil. -> A decrease in the water retained in the soil causes an increase in the harshness of the micro climate.

Chain 11. Burning causes the volatilisation of nitrogen. -> Volatilisation of nitrogen causes a decrease in soil fertility. **Chain 12.** Burning causes a volatilisation of sulphur. -> Volatilisation of sulphur causes a decrease in soil fertility.

Chain 13. Burning causes the destruction of micro-organisms. -> The destruction of micro-organisms causes a reduction in soil fertility.

Chain 14. Burning causes an increase in nutrient solubility. -> Increase in nutrient solubility causes a decrease in soil fertility.

Chain 15. Burning causes a decrease in animal population density. -> A decrease in animal population density causes a decrease in the rate of nutrient cycling.

Chain 16. Burning causes a decrease in animal population density. -> A decrease in animal population density causes a decrease in the rate of predation. -> A decrease in the rate of predation causes an increase in pest numbers. -> An increase in pest number causes an increase in the rate of pest attack. -> An increase in the rate of pest attack causes pasture degradation

Chain 17. A decrease in soil fertility causes nutrient deficiency.

Chain 18. A decrease in soil fertility causes the occurrence of oligoculture. -> Oligoculture causes reduced plant defences against attack. -> Reduced plant defences against attack result in an increased rate of pest attack. -> An increased rate of pest attack causes pasture degradation.

Chain 19. A decrease in soil fertility causes the occurrence of oligoculture. -> Oligoculture causes an increase in weed population sizes.

Chain 20. An increase in the harshness of the micro climate causes an increase in weed population size.

Table 5.3 continued

Chain 21. An increase in the harshness of the micro climate causes a decrease in the number of palatable plant species.
Chain 22. An increase in the weed population size causes an increase in the rate of competition for light. -> An increase in the rate of competition for light causes pasture degradation.
Chain 23. An increase in the weed population size causes an increase in the rate of competition for nutrients. -> An increase in the rate of competition for nutrients causes pasture degradation.
Chain 24. An increase in the weed population size causes an increase in the rate of competition for water. -> An increase in the rate of competition for water causes pasture degradation.
Chain 25. A decrease in the number of palatable species causes an increase in grazing pressure. -> An increased grazing pressure causes pasture degradation.
Chain 26. A decrease in the number of palatable species causes an increase in the compaction rate. -> An increase in the compaction rate causes pasture degradation. -> A decrease in the number of palatable species causes an increase in the trampling rate. -> An increase in the trampling rate causes pasture degradation.

The length of the chains of statements of fact used in Table 5.3 is a compromise between capturing connectivity and avoiding repetition. This same process might have made use of much longer chains, thereby capturing more of the connections between of statements of fact. Where chains start with terms that do not occur as the tails of any other statements of fact and do not end with any terms that represent a head of another statement of fact this has the added advantage that there is no linkage between chains to be captured, but this is not normally be the case. In Table 5.3, five chains end with 'an increased harshness to the micro climate' and two start with the same. These linkages are deducible but are not explicitly represented. The long chain approach would also have resulted in a much larger set of chains and, therefore, much more repetition. Equally, shorter chains could have been used. Table 5.2 represents this approach in its extreme with no linkages explicitly identified. This demands a significant effort on the knowledge base developer's part to ensure that all useful linkages are indeed identified. The use of the infix operator 'AND' serves to reduce repetition. Use might be made of further such operators, including OR, THEREFORE and BECAUSE. However, reliance on a matrix of such operators to link statements of fact begins to obscure the individual statements of fact. The intention here was to provide an overview of the linkage between a set of statements of fact, rather than a means of representing an argument.

In summary, text interfaces have been found not provide an effective means of providing an overview of the linkages between a set of statements of fact.

5.2.2 Diagrammatic approaches

In contrast to textual representation, diagrammatic approaches to representing linked sets of statements of fact, which do not contain an innate linearity, have been demonstrated to provide an effective overview.

Diagrams are widely used by ecologists and resource managers as a means of summarising information. The use of diagramming as a means of summarising the behaviour of a system or a portion of a system under particular circumstances is of particular relevance in the current context. This may be very informal or may be rigorously applied as in the development of a diagram as a basis for compartment flow modelling in the development of computer simulation models (Dent and Blackie, 1979). One of the most widely used approaches is referred to here as node-and-link diagramming. Node-and-link diagrams are diagrams in which pieces of text (nodes) are joined by lines, usually arrows, (links). Common examples include taxonomic trees, causal or influence diagrams and compartment-flow diagrams. This section explores the utility of a node-and-link diagramming approach to representing linked sets of statement of fact.

Figure 5.1 is a diagrammatic representation of the knowledge in Table 5.1. This figure provides a structure for explicitly recording all the knowledge captured in Table 5.1 and can be disaggregated into the same set of statements of fact as in Table 5.2. It is proposed that this provides a much more effective overview of the linkages between these statements of fact than Table 5.3. Figure 5.1 is constructed using a convention for producing diagrammatic representations of linked sets of statements of fact developed in this chapter.

5.3 NODE-AND-LINK DIAGRAMS AS A MEANS OF REPRESENTING LINKED SETS OF STATEMENTS

5.3.1 Types of node-and-link diagrams

The meaning of a node-and-link diagram is dictated by the meanings of the links between nodes. Consideration of 30 published node-and-link diagrams (Appendix C) showed that the meanings of links within diagrams are frequently not specified, making such diagrams uninterpretable. Nevertheless, five clear categories of node-and-link diagrams can be identified on the basis of the type of relationship represented by the link. Of these only one category is relevant to the representation of explanatory ecological knowledge as undertaken in this thesis.

(i) Decision trees

In decision trees, nodes represent questions or options and links represent answers (Yes or No) or selection. Decision trees provide a mechanism for making a decision rather than representing an explanatory set of knowledge.

Figure 5.1 Diagrammatic representation of a linked set of statements of fact.

(ii) Management-flow diagrams

The term 'management-flow diagram' is used to describe diagrams that outline the stages in a process, sometimes including objectives and decisions, and indicate the order in which these occur. Although widely used, such diagrams are generally informal and idiosyncratic and frequently have a poorly defined syntax. Management flow diagrams do not provide a means of explicit representation of knowledge.

(iii) Compartment-flow diagrams

Compartment-flow diagrams are widely used to describe systems, including ecological systems. Exacting protocols exist, particularly in relation to using such diagrams as a basis for mathematical or simulation compartment - flow modelling (Dent and Blackie,1979).

The links in flow diagrams represent flows from one compartment (node) to another. A second link type to represent an influence on a flow or a compartment may sometimes be used.

Compartment-flow diagramming is used as a software interface in modelling applications such as Stella (Lewis, 1986). However, description is limited to statements about flows between entities and the influencing factors on those flows.

(iv) Association

The term 'association diagram' is used here to describe diagrams that identify an unspecified association between two nodes. The development of association diagrams may help an individual in thinking about a system but, because the meaning of the links is not explicit, the diagram serves no other useful purpose. Nevertheless, association diagrams are surprisingly widespread in the literature.

(v) Explanatory node-and-link diagrams

Explanatory node-and-link diagrams have nodes which represent entities and links which represent a named relationship between those entities, for example *'influences'*, *'is a type of'* or *'is followed by'*. Because such diagrams capture descriptions of the relationships between entities they are suitable for communicating explanatory and descriptive knowledge about systems or parts of systems.

Consideration of the diagrams listed in Appendix C shows that different links with different meanings are encountered in published diagrams. In principle, different links may be used within the same diagram. In practice, diagrams tend to contain a single link type. Link types frequently encountered in published diagrams include :

- influence links, of which causal links are a particular type;
- taxonomic links; and
- temporal links, of which there are several types including :

  (a) those describing the order in which events (which may be unrelated) occur (e.g. going to work happens after breakfast);

  (b) those describing sequences in which each step is the necessary next step in a process (e.g. sowing happens after ploughing); and

  (c) those describing an evolutionary sequence in which objects or events change into other objects or events (e.g. a caterpillar becomes a butterfly).

5.3.2 The utility of a diagrammatic representation of linked sets of statements of fact

The diagrammatic representation of linked sets of statements of fact can facilitate :

- knowledge elicitation;
- knowledge synthesis;
- knowledge dissemination; and
- decision support.

Textual and diagrammatic approaches to representing linked sets of statements of fact have been discussed in §5.2.1 and §5.2.2. These linked sets may then be disaggregated into individual statements of fact. Because these statements of fact are abstracted from a coherent set, the connectivity between pairs of statements is likely to be higher than for an equivalent set of statements of fact entered individually.

(i) Eliciting knowledge

The representation of knowledge as a linked set of statements of fact during the process of knowledge elicitation provides a powerful means of helping to ensure that the knowledge set that is elicited is comprehensive and coherent, and therefore useful.

The knowledge synthesis inherent in diagramming can provide a useful vehicle for knowledge elicitation (Chambers *et al.*, 1989; Lightfoot, 1989). In particular, diagrams can provide a useful medium for reaching consensus, because, where based on a defined set of symbols and conventions, they can represent an unambiguous explanation of a system. As a result, using node-and-link diagrams in knowledge elicitation can result in a set of knowledge that is significantly more comprehensive and coherent than that which results from other approaches to elicitation. Effective elicitation requires the informant(s) and the individual developing the diagram to use an agreed set of conventions in developing the diagrams.

(ii) Synthesising knowledge

Producing a diagram is a powerful means of enabling the developer to synthesise available knowledge on a particular topic and, as a result, to increase his or her understanding of that topic. Diagrams can be used to make an explicit statement of what is known about the topic. To make this explicit statement, developers must systematically clarify their understanding of the existing knowledge, identifying and dealing with ambiguity, and organising the knowledge accordingly. This process forces a systematic and rigorous approach to considering the available knowledge. Furthermore, it enables developers to assess the completeness of their current understanding or of the available knowledge through the identification of missing linkages in the diagram.

The diagramming approach to representation is not constrained by linearity. A set of statements of fact may be legitimately explored from any point in any direction. As a result, diagrams provide a more succinct representation of knowledge than textual approaches.

(iii) Disseminating knowledge

Diagrams are widely used as a means of conveying knowledge, for example in publications to illustrate and enhance information described in the text and may be appropriate as a tool for extension (Conway, 1989). A carefully structured diagram can succinctly and accurately convey a large amount of complex information as a concise statement.

To act as an effective means of conveying information the conventions used in constructing the diagram must be clear and, preferably, familiar to the readers. Diagrams which do not conform to a stated convention may have been useful for the developer in synthesising the knowledge, but are not necessarily suitable as a means of conveying that knowledge to other people.

(iv) Decision support

Carefully drawn diagrams can be used to assess the implications or impact of management actions because they provide a succinct summary of what is known.

Knowledge based probabilistic systems are used in a range of decision-making roles such as medical diagnosis(Pereira and Barlow, 1990) and military decision making (Shachter, 1988). These systems are based on node-and-link-type diagrams based on a tightly defined logic and are mathematically precise (Shachter, 1990a and b).

In the context of decision making in relation to ecological systems, knowledge is frequently not precise enough to enable such a rigorous approach. However, diagrams can successfully structure available knowledge even where it is qualitative and incomplete. This structured knowledge can facilitate a more systematic approach to decision-making, allowing a more accurate assessment of the implications of those decisions on the basis of available knowledge than might otherwise be possible.

5.4 A CONVENTION FOR PRODUCING NODE-AND-LINK DIAGRAMS TO REPRESENT ECOLOGICAL KNOWLEDGE

In §5.3 explanatory node-and-link diagrams were identified as providing a means of representing knowledge. Consideration of published diagrams and experience in the use of a diagramming approach to knowledge representation has shown that diagrams only provide a useful intermediate representation of knowledge where they are developed in accordance with an explicit set of conventions.

This section presents a set of conventions for the development of explanatory node-and-link diagrams as a means of intermediate knowledge representation.

Diagrams developed on the basis of these conventions display the following features :

- every node in the diagram is fully labelled;
- every node represents an object, a process or an attribute;
- information is primarily attached to links rather than nodes;
- the meaning of every link is explicitly stated;
- linked pairs of nodes represent information that corresponds to a statement of fact;
- the labels attached to links are consistently used;
- complex sets of knowledge are represented as hierarchically linked sets of diagrams;
- conditional information is stated for each link; and
- all links are tagged according to the source of the knowledge they represent.

(i) Labelling nodes

In published diagrams, node names are frequently ambiguous as they contain implicit meaning (§5.3.1). For example, node labels in diagrams that represent influence are frequently the names of objects (e.g. pests) or processes (e.g. grain set), while the links mean 'influence(s)'. However, 'pests influence grain set' is incomplete information: all that it explicitly states is that some attribute of pests influences some attribute of grain set.

Where this attribute is known, it has to be stated rather than implied, e.g. 'pest population size' 'influences' 'timing of grain set' if knowledge is to be explicitly stated. Complete information about the meaning of each node is critical for explicit representation of knowledge in the diagram.

(ii) Classification of nodes

Classifying each node according to type encourages a full statement of meaning and results in diagrams that are consistent with the specifications for intermediate and formal statements of fact developed in Chapters 3 and 4.

Three types of node are identified:

Object nodes represent things, or more commonly, groups of things. An object is something that occurs physically e.g. an oak tree, or in an abstract sense, e.g. a niche.

Process nodes represent things that happen e.g. seed germination or soil erosion. Depending on the time scale involved, a process might be called an event (it is clearer to refer to the death of an organism as an event rather than a process) but an event is simply a special type of process that occurs instantaneously. From the point of view of the semantics of a diagram, they are equivalent.

Attribute nodes refer to a particular attribute of a process or an object. An attribute is something relating to an object or process that might be measured, e.g. height of oak tree, rate of germination, colour of flower. Experience shows that attribute nodes are more frequently used than object or process nodes.

This set covers all the node types that are used in the creation of explanatory node-and-link diagrams. Retrospective classification of nodes in a set of diagrams created without the use of a computer and summarising knowledge elicited during six months of fieldwork in India (Garde, 1992) shows that of the 102 nodes included (counting each node only once, i.e. ignoring duplication in sub-diagrams and parallel diagrams), all fitted into the set of node types given. There were 87 attribute nodes, ten object nodes and five process nodes. The classification also conforms to the fundamental components of statements of fact represented using the formal grammar (Chapter 5).

Classification may depend on context. For example, 'shade' might be represented as an attribute (of a site, for example), an object (in the sense that shade is cast or falls) or a process (shading) in different diagrams. This again illustrates the need for consistent and precise use of terms (Chapter 3).

Classification has been particularly problematic where an entity was most appropriately represented as one type of node in one relationship but another type in another relationship. So, in the reimplementation of the diagram set developed by Garde (1992) in HyperNet (Appendix A) 'location' was considered to be an attribute of 'forest' but 'aspect' represented an attribute of 'location' (which, therefore, becomes an object in relation to aspect). Such problems can invariably be resolved, in this case by viewing 'aspect' as an attribute of 'forest' rather than 'location'.

(iii) Attaching information to links

In all the published node-and-link diagrams considered in §5.3.1, the majority of information was attached to nodes. Links were rarely explicitly labelled. This approach may be intuitive but imposes constraints on the diagram. When this approach is used in constructing, for example, causal diagrams the links simply mean 'causes', while exactly what is caused is specified in the nodes. So, for example, the node 'increase in leaf area' might be linked by a 'causes' link to 'increased crop yield'. This severely constrains further additions to the diagram. The fact that pest attack reduces leaf area might need to be included. This would only now be possible by adding a further node to the diagram 'decreased leaf area'. It is not desirable to have two nodes referring to leaf area, one to increase and one to decrease. Changes do not occur only in terms of the quantity of an attribute but also in the existence of an object or process. Under this scheme there may be clumsily labelled nodes e.g. 'occurrence of ovulation' or 'disappearance of frost'.

These problems are overcome by attaching the information about changes occurring to the link rather than the nodes. So 'increase in leaf area' 'causes' 'increased crop yield' becomes 'leaf area' 'increase in causes increase in' 'crop yield' (See Figure 5.2). Over a large diagram the discipline of attaching the majority of information to the link simplifies diagram construction.

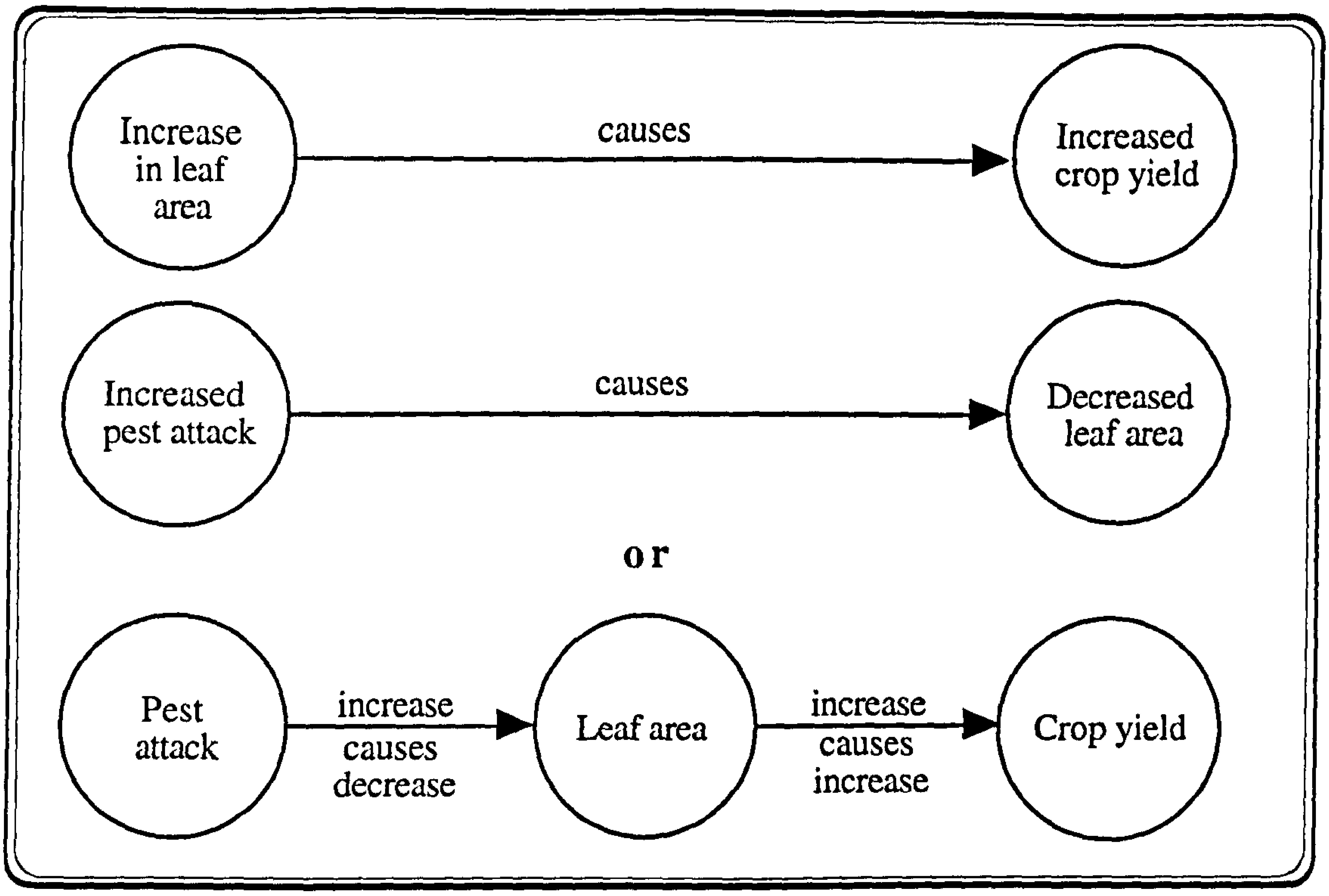


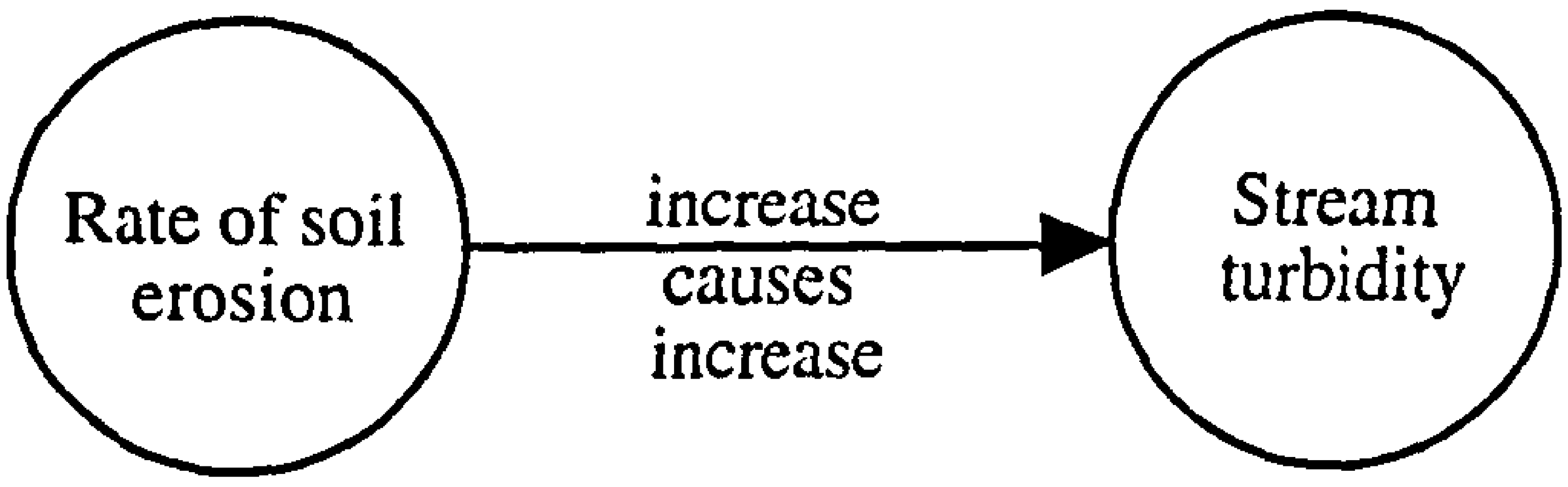
Figure 5.2 Attaching information to links rather than nodes.

(iv) Stating the meaning of links

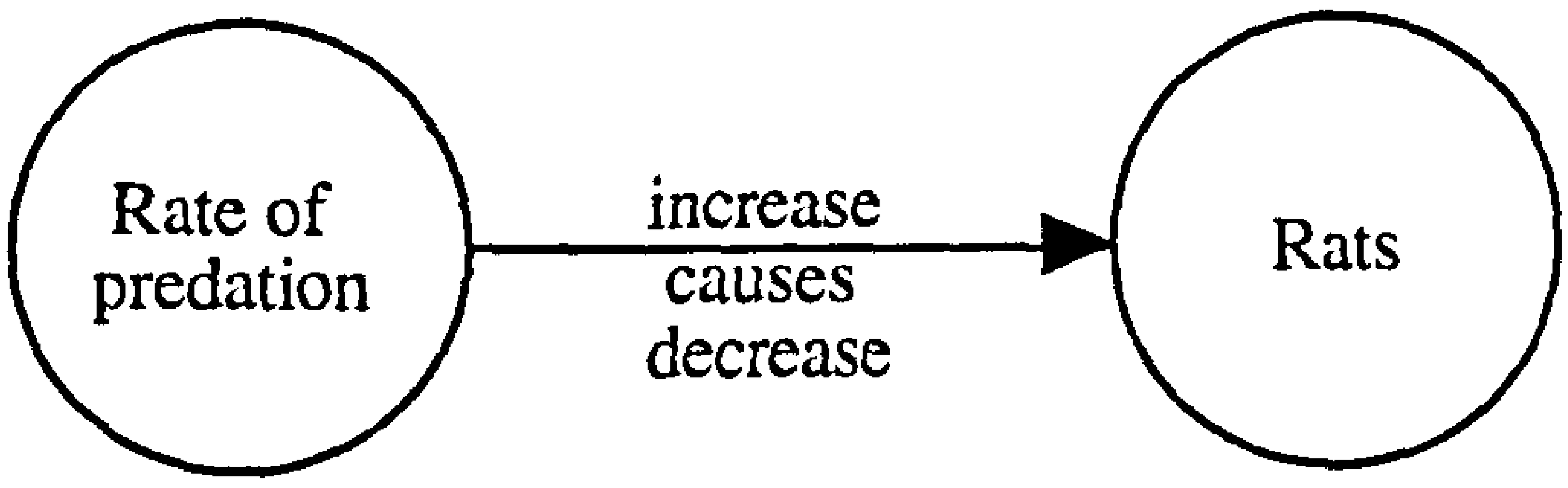
If a diagram is to represent an explicit statement of knowledge, links can only remain unlabelled if they all have exactly the same meaning and this meaning is clearly defined. This makes the mixing of different types of link in a single diagram impossible and is extremely restrictive where the links are considered to be the primary source of information. The explicit statement of the meaning of each link facilitates the flexible development of a representative diagram.

(v) Linguistic correspondence of linked pairs of nodes

If node-and-link pairs represent useful knowledge, there will be a correspondence between the knowledge represented by each pair (i.e. nodes and link) and a meaningful English sentence. This sentence will, in fact, comply with the definition of a statement of fact (§5.3.1). Where this correspondence does not occur, one or more of node pair and link must be inappropriately labelled.



can be read as 'An increase in the rate of soil erosion' 'causes' 'an increase in stream turbidity', which makes sense in English whereas :



would be read as 'An increase in the rate of predation' causes a decrease in 'Rats'. This is understandable but incomplete: it clearly refers to a decrease in the rat population size. This should be explicitly stated by relabelling 'Rats' as 'Rat population size'.

The node 'Rate of predation' is also incompletely labelled and clearly refers to the rate of predation on the rat population. It can be argued that this can be assumed from the context of the diagram, but, where a diagram is created as a means of generating a coherent knowledge base, this cannot be assumed. Predation of other species may occur elsewhere in the knowledge base which can probably not be assumed as being of the same nature as, and therefore equivalent to, the predation of rats.

(vi) Labelling links

It will often be possible to express the meaning of a particular link in different ways. However, using different terms for the same concept reduces the comparability of the resulting diagrams, particularly given that differing terms for the same concept will rarely be exactly equivalent.

```
Fungal wilt causes yield loss
```

is similar to :

```
Fungal wilt results in yield loss
```

and

```
Fungal wilt leads to yield loss
```

and not very different from

```
Fungal wilt is followed by yield loss
```

which might be restated as

```
Fungal wilt happens before yield loss.
```

Ensuring a consistent use of terms both within a diagram and between diagrams is desirable in trying to attain unambiguous comparability of links.

Application of a diagramming approach for representing knowledge as implemented in a trial software package, HyperNet (Walker *et al.*, 1993), made use of the link types shown in Table 5.4.

Table 5.4 A restricted set of diagram link types

_ influences _,
Increase in _ causes increase in _,
Decrease in _ causes increase in _,
Increase in _ causes decrease in _,
_causes an increase in _,
_ causes a decrease in _,
Increase in _ causes_,
Decrease in _ causes_,
_ results in _,

_ is a sort of_,
is a part of,
is an example of,
has a type,
has a part,

happens before,
happens at the same time as,
happens after,
leads to,
changes to.

These have been found to enable the representation of a useful proportion of ecological knowledge. It will be noted that these link labels correspond closely to the template sentences discussed in Chapter 3. Although useful, this set of options is not exhaustive; it has frequently proved necessary to specify special links for use in particular circumstances. In §5.5, the modification of this set of link labels to comply with the specifications for formal representation developed in Chapter 4 is discussed.

(vii) Hierarchically linked sets of diagrams

Experience in the creation of diagrams as a means of producing coherent knowledge bases suggests that diagrams containing more than 40 nodes tend to be increasingly difficult to create and interpret. This is in part the result of the balance that must be achieved between the size, and therefore, legibility of nodes and the text associated with links, and standard paper or computer screen size. However, even where large boards or paper are used, diagrams rapidly become unsustainably complex. It is the clear representation of links rather than nodes that presents difficulties. Typically a node will be linked to between two and five other nodes. With an increasing number of nodes it becomes increasingly difficult to place nodes near to all the other nodes to which they are linked: links have to travel further across the diagram, crossing increasing numbers of nodes and links until the diagram becomes impenetrable. The placement of nodes, technique used for representing the meaning of a link and arrow type used to link nodes all have an impact on the number of nodes and links that can be successfully represented, nevertheless, fundamental limitations will still be reached. However, the contents of a single diagram can be split into a set of hierarchically linked diagrams and each diagram in the set can be relatively simple while the knowledge represented by the whole set is complex.

(a) Disaggregating complex sets of knowledge into subsets for representation

Where using a diagramming approach to represent a complex set of knowledge, that knowledge can be broken down into hierarchically structured subsets of knowledge where the hierarchy represents a rational classification of knowledge in increasing levels of detail (with the most generic knowledge in the top-level diagram and the most specific knowledge in the bottom level of diagrams). This hierarchy can be captured in the diagram by expanding a link or node in the parent diagram into a sub-diagram.

Expanding links into a sub-diagrams

A link in a diagram may be expanded into a sub-diagram, producing a meaningful link between sub and parent diagram. By default, the nodes at each end of the link in the parent diagram should appear in the sub-diagram.

Nodes can be expanded into sub-diagrams, or areas of the diagram can be collapsed into a single nodes.

A node in a diagram may be expanded into a sub-diagram, producing a meaningful link between sub and parent diagram. For example the object node 'soil fauna' may be expanded into a diagram representing the interactions between the various components of soil fauna. The process node 'photosynthesis' might be expanded into a more detailed representation of photosynthesis while the attribute 'rate of soil erosion' might be expanded to represent the factors influencing the rate of soil erosion.

(b) Diaggregating complex diagrams into hierarchically structured sets of diagrams

Where a diagram progressively becomes too complex to be useful or where a clearly modular structure appears within a diagram, sets of nodes and associated links representing a coherent portion of the diagram can be collapsed down to produce a single node or link, with the original nodes and links becoming a sub-diagram linked to the new node or link.

The end result is the same as the application of the previous principle, but, in the former the hierarchical structure of the knowledge is defined before the diagram set is produced. In this instance the hierarchical nature is identified as the structure of the knowledge on the topic under consideration is elucidated.

Figure 5.3 illustrates the hierarchical expansion of a diagram link.

(viii) Recording conditional information about links

There may frequently be more than one relationship between two nodes in a diagram and therefore more than one link between two nodes. Such multiple relationships can only occur under mutually exclusive circumstances. Each link in a set of multiple links between two nodes in a diagram must be validated by recording the unique circumstances under which the relationship that that link represents applies.

For example, the nodes 'soil nitrogen levels' and 'crop growth rate' may be linked by both the apparently contradictory links 'increase in _ causes increase _ in' and 'increase in_ causes decrease in _'. The additional conditional information that the former applies below certain nitrogen levels and the latter above those levels is required if this knowledge is to be interpretable and useful.

(ix) Tagging links according to source

All the links in a diagram should be attributable to a source. This allows the user to :

- make an assessment of the reliability of the knowledge summarised in the diagram (particularly where the diagram is to be used in decision support), by referencing information about the source; and

- upgrade and improve the diagram by returning to the information source for clarification of knowledge already represented and addition of further knowledge.

Nodes on their own have very limited meaning and, therefore, do not require tagging.

This section has developed a convention for the diagrammatic representation of sets of statements of fact. The following sections discuss the application and modification of this convention in the use of a diagrammatic approach as an interface in creating and viewing knowledge bases.

5.5 DIAGRAMMING AND INTERMEDIATE REPRESENTATION

The diagramming convention detailed in §5.4 provides a basis for the representation of linked sets of statements of fact that can be disaggregated into a list of statements of fact that correspond to node pairs and the links between them. Node pairs and a link joining them therefore correspond to an intermediate statement of fact, as specified in Chapter 3. Moreover, with some further restrictions on the diagram syntax, the intermediate statement created by the node pair and link can, because of its syntax, be automatically formalised. Diagramming also provides an interface for the specification of the hierarchical relationship between process and object keywords. The use of a diagramming convention as an interface for intermediate representation therefore demanded the development of specifications for two separate types of diagrams: core diagrams and keyword diagrams.

5.5.1 Core diagrams

Core diagrams correspond to the conventions developed in §5.4. Two further restrictions (restriction on permissible links and full specification of the meaning of attribute nodes) are imposed to enable automatic formal representation.

(i) Restriction on permissible links

The permissible link types in core diagrans were restricted to the three basic binary relationships supported by the formal grammar: causal, link and comparison statements.

Causal links were restricted to :

Increase in _ causes increase in _,
Decrease in _ causes increase in _,
Increase in _ causes decrease in _,
_causes an increase in _,
_ causes a decrease in _,
Increase in _ causes_,
Decrease in _ causes_,
_ causes _,

The 'link' links that may be used were not specified, but, a 'link' link was only permissible between objects and must be a verb.

Permissible comparison links were restricted to : greater than; equal to; less than. Comparison links are always between attribute nodes as any comparison of two objects or two nodes will, inevitably, be a comparison of any attribute of those objects or processes.

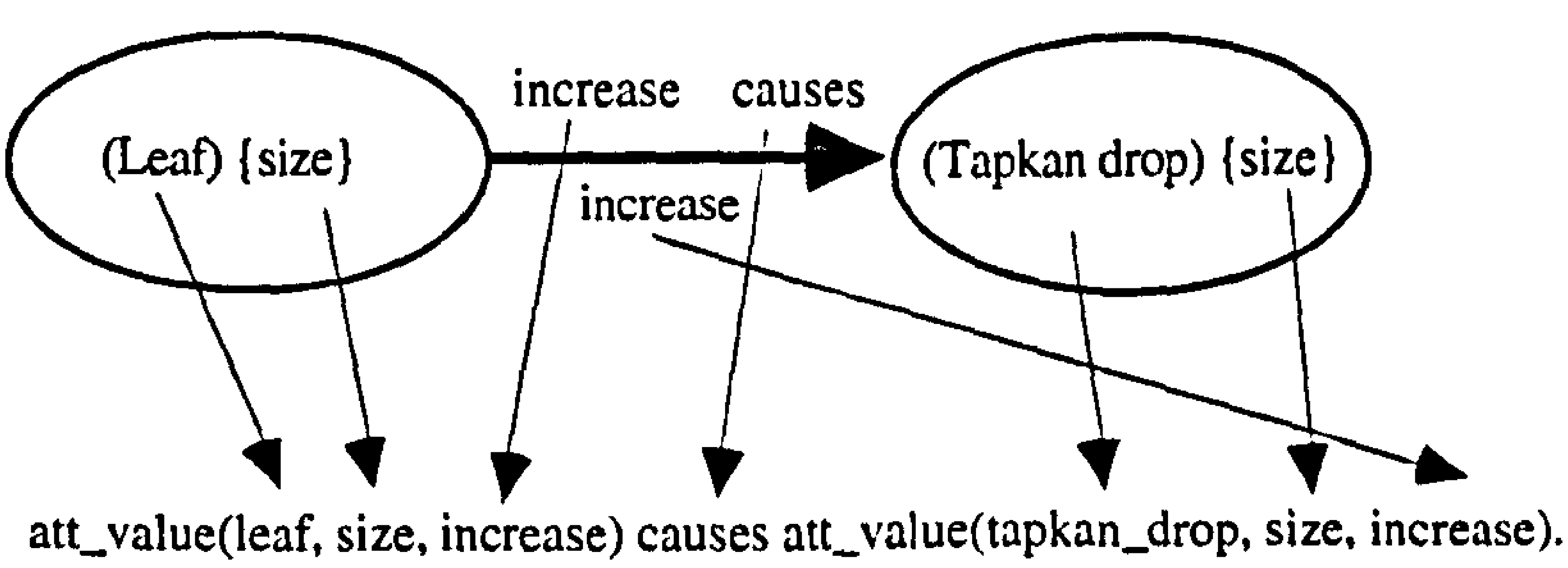
In principle these link types may be used in combination within a diagram. In practice they have usually been used in separate diagrams. Causal and link diagrams (for example a food web) represent fundamental ecological knowledge and are frequently complex, making the diagrams useful in the representation of knowledge of both types[6]. By contrast large, complex and highly connected comparison diagrams are very unusual. Comparisons are usually made between two objects, attributes or processes. Where more than two entities have been compared (e.g. in a ranking of soils in terms of fertility), comparison is often linear. As a result, the value of using a diagram as opposed to textual approach to represent comparative knowledge is marginal. The diagramming approach does, however, restrict the permitted statement structure, leading to less ambiguity in representation, and enables automatic formal representation.

As with the diagramming conventions developed in §5.4 and the requirements for an intermediate statement of fact (§3.2), each link may be tagged with conditions and source.

(ii) Full specification of the meaning of attribute nodes

In the formal grammar, attribute statements identify both the attribute and the object or process to which that attribute applies. Explicitly identifying the object or process and the attribute of that object or process under consideration for every attribute node provides a more robust intermediate representation than not doing so.

The resulting node pair and link represent an intermediate statement of fact with a known structure which can, therefore, be automatically formalised. For example :



---

[6]Because current reasoning tools support the use of causal statements but not link statements, experience to date has been almost entirely in the creation of causal diagrams.

The core diagrams provide a means of representing sets of binary statements of fact. However, the knowledge base will also contain attribute-value statements, information about the hierarchical relationships between object and process keywords and definitions of objects, processes and attributes. These may be captured in, or accessed through, keyword diagrams.

(iii) Keyword diagrams

Hierarchy diagrams provide a means of representing object and process glossaries as sets of hierarchies, a framework for recording definitions and a framework for recording attribute - value statements.

These hierarchies may be one of four types :

- objects linked by 'is a type of' links;
- processes linked by 'is a type of' links;
- objects linked by 'is a part of' links; and
- processes linked by 'is a part of ' links.

Using separate hierarchies to represent these different types of relationship will tend to result in a clearer representation of the knowledge, particularly because nodes at the same level in a hierarchy will tend to be interpreted as in some sense of equivalent status which will not normally be the case where link types are mixed. As discussed in Chapter 4, the diagrammatic representation of keyword hierarchies provides a useful overview of relationships. Following a standard convention such that the most general objects or processes occur at the top of the diagram and are progressively expanded down through the diagram helps to ensure that the hierarchies are correctly interpreted. Each node in the diagram may then be tagged with a list of statements that provide a definition for that term (§3.4) and a list of attribute-value statements that are associated with that term (clearly there will be overlap between the two lists).

As a result, while statements of definition and attribute-value statements are not, in themselves, diagrammatically represented, they are ordered and accessed through a diagrammatic framework.

5.5.2 Comparative utility of a diagramming approach in creating knowledge bases

In response to a perceived need for a means of viewing linked sets of statements of fact, a diagramming interface has been implemented within knowledge acquisition software (§6.5.3). Both this approach and the direct entry of disaggregated statements of fact as text have been and continue to be used in the creation of knowledge bases. This section compares experience in these two approaches to knowledge base construction.

(i) Coherence, completeness and utility

Entering disaggregated statements of fact abstracted from interview material has proved to be advantageous in that :

- statements of fact can be entered into the knowledge base rapidly, by contrast, diagram construction is time consuming;
- identifying and entering statements of fact requires less skill than the construction of a diagram;
- a greater percentage of the knowledge derived from a source (e.g. an interview) is likely to be entered, because in linking a set of statements of fact into diagrams, some knowledge is likely to be omitted; and
- statements of fact entered into the knowledge base will be closer to source knowledge in form and wording than in cases where the knowledge has been modified to fit into a linked set of statements of fact.

However, knowledge entered through the construction of a diagram has tended to be critically assessed and clarified (possibly through further knowledge elicitation) before entry. This assessment has been particularly undertaken in relation to the knowledge already entered, thereby ensuring linkages between new knowledge and knowledge already entered, allowing both to be modified in developing a more coherent representation of source knowledge. By contrast, knowledge entered through a direct statement entry approach has often only been superficially assessed, leading to the inclusion of ambiguity and inconsistency. In particular it has been hard to assess the implications of new knowledge for existing knowledge. So, construction of diagrams in knowledge base creation has generally led to knowledge bases of a higher 'quality' (i.e. fitness for purpose). The critical assessment inherent in the construction of a diagram has also reduced the amount of irrelevant knowledge entered into the knowledge base. Furthermore, it has facilitated the identification of gaps in the knowledge, encouraging more directed fieldwork and so led to a more comprehensive knowledge base. By contrast, experience has shown that it is extremely difficult to identify gaps in knowledge treated as discrete statements of fact (Thapa and Walker, 1992). The net result is that, if properly used, the construction of a diagram will result in a more cohesive, coherent and comprehensive knowledge base comprising fewer statements of fact of higher individual utility.

(ii) Expressiveness

The diagramming approach proposed can be used to represent the four basic structures supported by the formal grammar. However, while experience suggests that almost all textual intermediate statements of fact can be formalised, not all the statements of fact that can be represented through the formal representation of intermediate statements of fact can be captured by the diagramming interface. The diagramming interface does not allow the full range of combinations that can be achieved through the formal grammar itself (Figure 5.4). While there are many intermediate statements of fact that cannot be explicitly represented diagrammatically, it is often the case that the same meaning can be captured by representing a completely different statement in the diagram. Because sense is subjective the implications of this restricted expressiveness have proved hard to evaluate. The importance of knowledge not represented has been equally hard to evaluate, as has the relative importance of these omissions in comparison to the improved coherence and utility of the knowledge base developed through the construction of a diagram. On balance, experience suggests that an interface requiring construction of diagrams has advantages over direct entry of disaggregated text in producing a useful knowledge base and is, therefore, recommended. However, direct entry provides a faster and more flexible approach. The two approaches are not mutually exclusive.

Knowledge bases created using the text route represent the source ecological knowledge more completely, but knowledge bases created through the diagrammatic route tend to be smaller, more coherent, better connected and, therefore, less ambiguous and more tractable (Figure 5.4). A comparative evaluation of the two approaches demands an evaluation of the proportion of ecological knowledge that can be represented with the diagramming conventions developed, the importance of the knowledge that is not represented and the relative value of an increased coherence and precision of representation.

On the basis of experience, it is asserted that all ecological knowledge can be represented as textual intermediate statements of fact, while applications of the formal grammar to date have suggested that some 95% of textual intermediate statements of fact can be formalised. While all diagrammatic intermediate representation can be formalised, it is the percentage of ecological knowledge that can be diagrammatically represented that remains unknown, nor is it apparent how this could be tested.

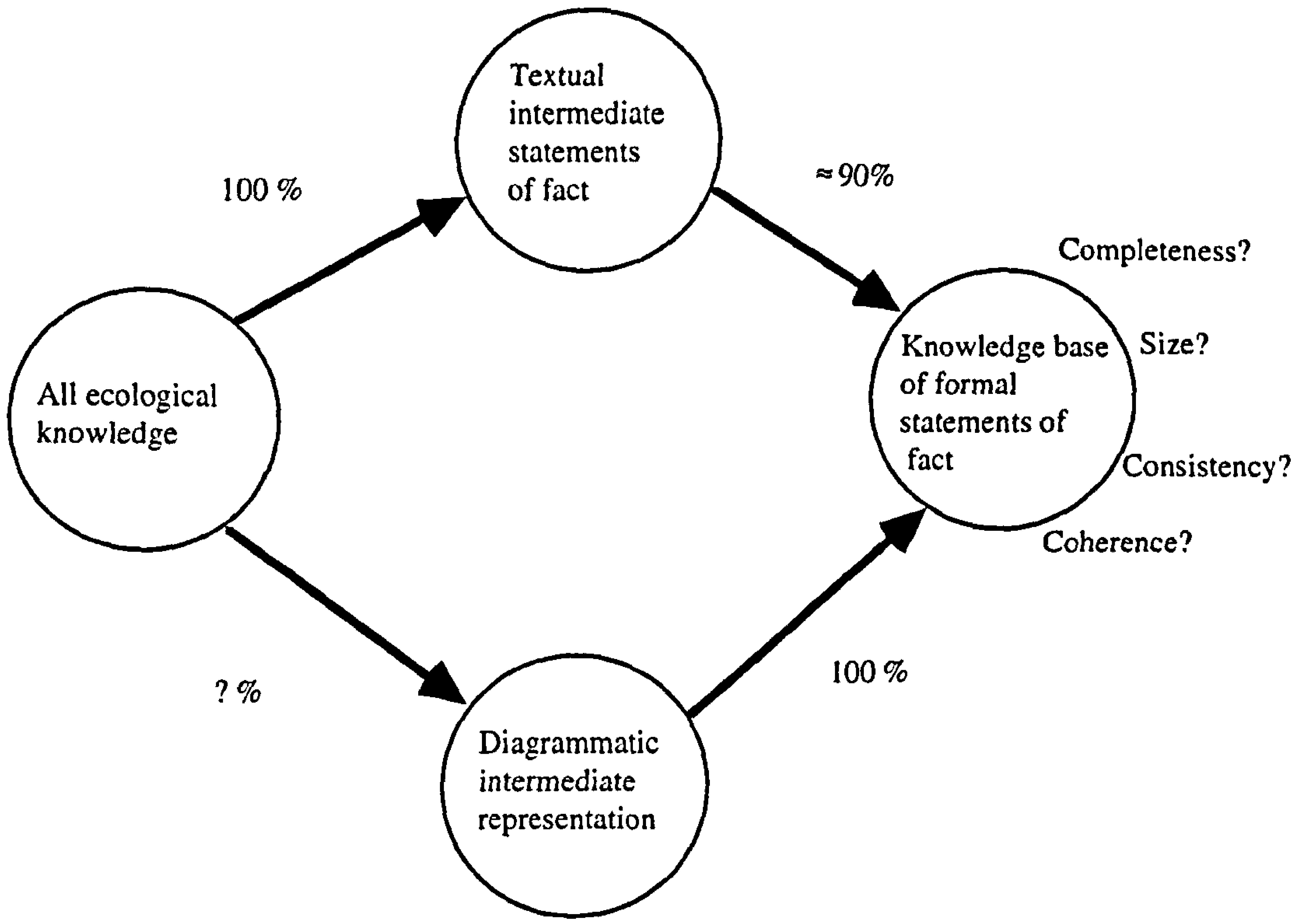


Figure 5.4 Two routes to the creation of a formal knowledge base. Arrows represent intermediate and formal representation, percentages indicate the proportion of the information that can be transformed. Some attributes of the resulting formal knowledge base that may vary according to route of representation are indicated.

5.6 THE DIAGRAM AS A KNOWLEDGE BASE INTERFACE

Diagram structures used in entering knowledge into the knowledge base may be stored and recalled as a means of providing a diagrammatic overview of the contents of the knowledge base (§6.5.3). However, this provides only a static view of a portion of the knowledge base. Of more interest is the automatic generation of diagrams.

The diagram interface developed provides a means of diagrammatically representing the binary statements of fact supported by the grammar and supporting a framework for attribute-value statements through attachment to hierarchical glossaries. As a result, all the necessary information is stored within a formal knowledge base to allow that knowledge base, or some part of it, to be represented as a computer-generated diagram. This provides a means of creating diagrams that reflect the evolving content of the knowledge base, for example, automatically combining statements of fact derived from two or more diagrams into a single diagram (Figure 5.5).

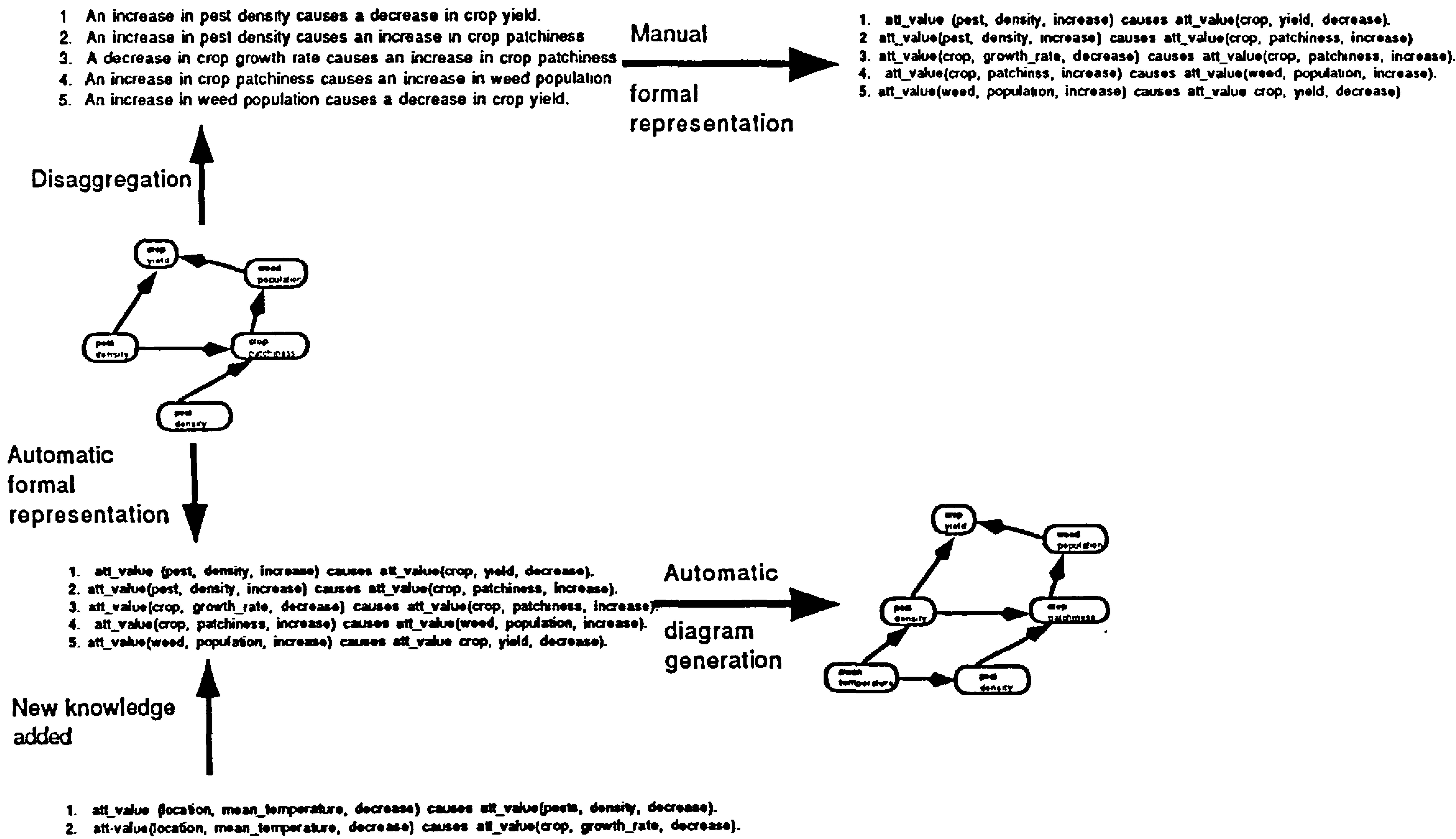


Figure 5.5 Diagram interface implementations. Both the current implementation and use of the diagramming interface (disaggregation and manual formal representation) and the proposed implementation (automatic formal representation and automatic diagram generation) are shown. The ability to combine knowledge from different sources in automatic diagram generation is indicated.

Formal statements of fact and hierarchical statements contain all the necessary information for the creation of the diagram. As a result automatic generation is, in principle, straight forward. Problems are, however, encountered in generating diagrams that are visually effective.

Three categories of binary relationship are supported in the formal language - causal, comparison and link. It is suggested that binary statements in any one category will tend only to be linked to others in the same category. As a result it is reasonable to restrict automatically generated diagrams such that they only contain one category of link.

Automatic creation of a visually effective taxonomic diagram is not overly demanding because there is a clear convention for the diagram to be vertically structured such that terms at the top of the hierarchy occur at the top of the diagram and so on down the hierarchy such that terms at equivalent level in the hierarchy occur at an equivalent level in the diagram. In this instance decisions are only required on the relative horizontal position of terms at a specified level on the hierarchy. This again is relatively straight forward given that terms will normally be clustered under their parent term.

Much more serious problems are, however, encountered in the automatic generation of other diagrams in which there is no clear structure. The computer-based generation of optimised node-and-link diagrams is a well-known problem, being encountered in many applications, including the design computer software itself (Protsko *et al.*, 1991). The optimisation strategies are complex because the recognition of an optimal layout is problematic. However, for the current purposes an acceptable rather than optimal layout is required, making the task significantly more tractable. An automatically generated diagram may then be manipulated by the user to improve its layout and interpretability (including 'collapsing' areas of a diagram to create a hierarchical diagram set).

The flexible creation of diagrams from selected subsets of the knowledge base will, therefore, contribute to an iterative improvement of the content of the knowledge base and provide a means of generating targeted output from the knowledge base.

5.7 CONCLUSIONS

Application of specifications developed in Chapters 3 and 4 in the creation of useful knowledge bases was constrained by the problems encountered in the consistent use of terminology. Furthermore, it is difficult to provide the user with an immediate overview of the resulting knowledge base. These problems have been addressed through the generation of mechanisms for representing the linkages between statements of fact. Representing linked sets of statements of fact as text is constrained by the innate linearity of text. By contrast, node-and-link diagrams provide an effective mechanism for representing the linkages between statements of fact.

The generation of conventions for the creation of node-and-link diagrams in knowledge representation and its implementation has provided an effective interface. This diagramming interface has been constrained such that the resulting knowledge representation corresponds exactly to the specifications developed in the Chapter 3 and Chapter 4. Indeed the correspondence is such that a formal statement of fact may be automatically generated from a pair of nodes and the link between them.

As well as providing a means of creating a knowledge base, diagrams may be used as a means of viewing the contents of an existing knowledge base. Automatic diagram generation would make this approach particularly powerful.

CHAPTER 6 IMPLEMENTATION - THE AKT1 SOFTWARE

6.1 INTRODUCTION

The research described in Chapters 3 to 5 generated specifications for the development of knowledge-based system software. This chapter details the implementation of these specifications into a software toolkit (AKT1 for Agroforestry Knowledge Toolkit version 1). The process of implementation provided a powerful means of evaluating the practicability of the specifications developed.

The design and implementation of AKT1 was based on the creation of a series of knowledge acquisition and manipulation software prototypes. The development of, and relationship between, these prototypes is detailed in Appendix A.

6.2 SUMMARY OF METHODOLOGICAL OBJECTIVES AND IMPLEMENTATIONAL REQUIREMENTS

The objectives developed through the preceding chapters are summarised below as a set of task objectives and consequent implementational requirements.

6.2.1 Task objectives

The software toolkit and associated approach are intended to facilitate the user in :

- the elicitation of coherent sets of knowledge from verbal or written sources;
- recording of sets of knowledge in a rigorous and complete form; and
- the evaluation of sets of knowledge.

6.2.2 Implementational requirements

Software requirements were identified as the provision of:

- a knowledge base structure for storing ecological statements of fact as intermediate and formal representations;

- a set of interfaces for entering knowledge into the knowledge base;
- a set of interfaces for the on-line access to the contents of the knowledge base;
- a set of mechanisms for reasoning with the contents of the knowledge base; and
- a set of mechanisms for producing printed output from the knowledge base.

6.3 IMPLEMENTATION

6.3.1 Programming environment

The research described has been undertaken using Apple Macintosh computers. Alternative programming environments considered included a UNIX or an MS-DOS platform. The Apple Macintosh was chosen for two reasons.

First, it was considered essential that developmental programming was performed on the same system used for implemented software. At the outset of software implementation, the UNIX platform was not sufficiently widely used or available in relevant institutions in developing countries to justify its use in software development.

Secondly, the software applications developed are for use by development professionals from developing countries. Many such users have had limited computer experience, some none at all. For this reason the 'user-friendliness' of the system used was identified as a key factor in determining the uptake of software developed. At the outset of the research, the windows environment available in Apple Macintosh computers was considered to be the most intuitively attractive and easily used platform (Kay and Goldberg, 1977; English *et al.*, 1977; Goldberg and Robson, 1979). Since then, developments in PC operating environments, notably the introduction of Windows, have considerably narrowed the gap between the two systems.

Concern that Macintosh machines would not be compatible with existing computing facilities at collaborating institutions, which are predominantly PC based, have proved to be unfounded. Several of the institutions involved (including Pakhribas Agricultural Centre, the University of Sri Jayewardenpura and the University of Peradeniya) now operate joint PC and Macintosh facilities. The continuous improvement in communications between the two systems have also helped to forestall potential problems. It is intended that AKT will be translated for use in a PC Windows environment in 1994.

AKT1 and its antecedents, HyperNet and the Prolog version of TEAK (Appendix A), are written in LPA MacProlog 4.5.

LPA MacProlog Version 4.5 runs on the Macintosh System 7 operating system (Johns, 1991).

6.3.2 Programming procedure

Programming was undertaken as a continuous, iterative process and was used as a means of exploring, developing and consolidating ideas relating to the task as well as producing implemented software. Detailed functionality was incrementally designed and implemented.

This iterative programming procedure has enabled an appropriate balance between the conceptual development and the implementation of outputs to be maintained.

6.3.3 Testing, evaluation and feedback

Testing the AKT1 software toolkit was achieved in three ways.

1. Testing by the programmer. The iterative program development resulted in continuous testing by the programmer and modification accordingly. This has been the most significant means of identifying implementational improvements.

2. Controlled testing. The software was regularly tested by individuals familiar with the task but not involved in implementation, by working through defined tutorials. This has primarily been a means of identifying weak points in interfacing, documentation and other user support that caused the user to precipitate problems.

3. Evaluation and feedback through application. From the early stages of development, researchers in four countries (§7.1) have been using the software in support of their research activities. Continuing feedback from these researchers in relation to their research activities has provided effective evaluation of the software.

6.4 AN OVERVIEW OF TOOLKIT FUNCTIONALITY AND ARCHITECTURE

AKT1 provides means of entering, viewing and editing knowledge, imposing a hierarchical structure on knowledge bases, abstracting subsets of knowledge bases and producing printed output.

There are six distinct sets of interfaces in AKT1 for entering, viewing and editing knowledge; the diagram interfaces (Figures 6.7 to 6.12), the sentences interface (Figure 6.13), the source interfaces (Figures 6.3, 6.4, 6.5) the hierarchy interface (Figure 6.17), the selection interface (Figure 6.18) and the statement set interfaces (Figure 6.15).

Intermediate statements of fact, associated information on the source of that statement, and conditions can be entered using either the diagram interface or the sentence interface. These are alternative approaches; the diagramming approach generates linked sets of statements of fact (Chapter 5), the sentence approach is based on disaggregated statements of fact (Chapter 3). An intermediate statement of fact is automatically generated for every pair of nodes joined by a link created through the diagram interface; this statement then becomes visible through the sentence interface and may be used in the same way as sentences entered through the sentence interface. Knowledge entered through the sentence interface cannot be accessed through the diagram interface.

The source interfaces allow the user to enter, access and edit information about the sources of the knowledge in the knowledge base. The hierarchy interface allows object keywords to be entered into the knowledge base in hierarchies. The selection interfaces allow the user to specify search criteria for abstracting subsets of statements of fact from the knowledge base. The statement set interface is used where the user requires access to sets of informal statements of fact, and, through them, access to full details on associated information on these facts.

6.5 AKT1 FUNCTIONALITY AND FEATURES

6.5.1 Managing knowledge bases

Each knowledge base created with AKT1 contains four sets of information :

- the statements of fact (including intermediate and formal representations);
- source details;
- full information about the structure and content of any diagram sets in the knowledge base; and
- records of any keyword hierarchies developed.

Each knowledge base created is stored as a Prolog text file. Knowledge bases are managed through the file menu (Figure 6.1).

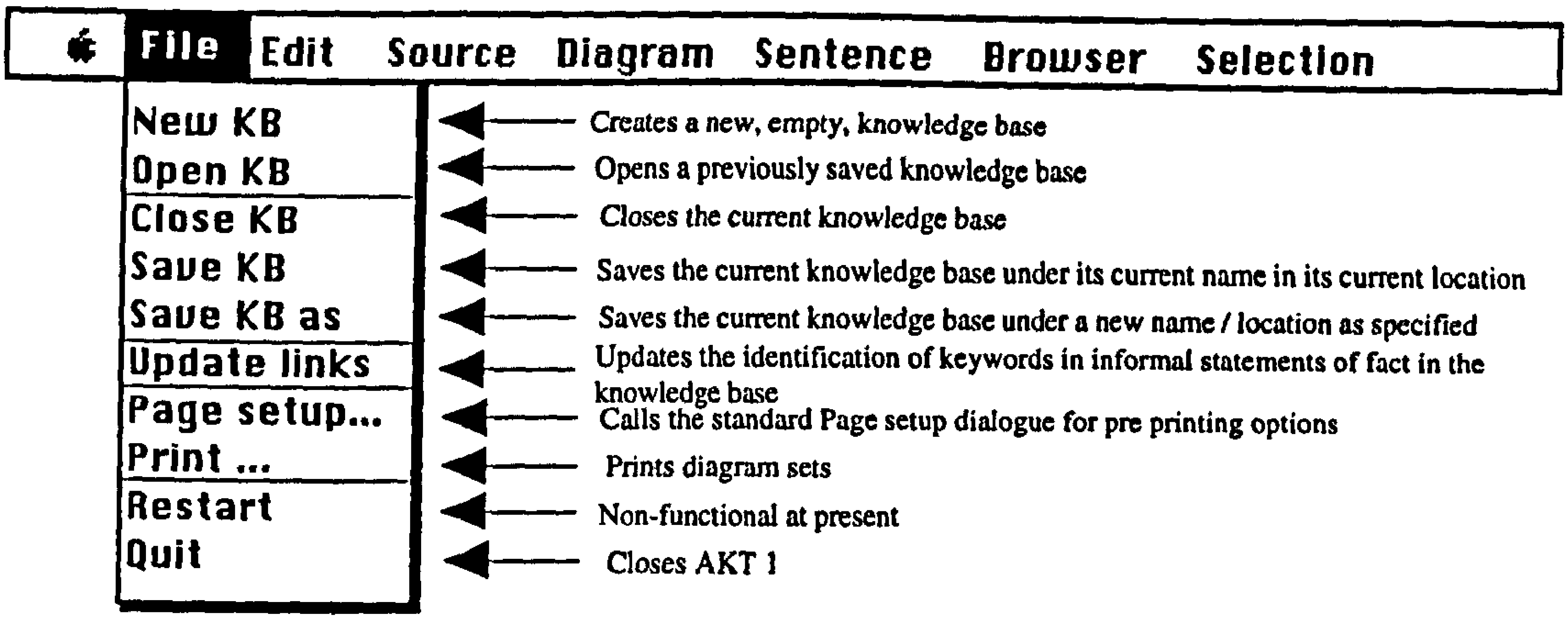


Figure 6.1 The file menu.

A knowledge base file must be opened before work can begin, either by opening an existing file (Open KB) or by creating a new file (New KB). Other file management options allow the current file to be closed (Close KB), the file to be saved (Save KB) or the file to be saved under another name or in another location (Save KB as).

6.5.2 Entering source information

Every statement of fact entered into the AKT1 knowledge base is tagged according to the source(s) to which it is attributed (as specified in §3.2.6). Information about each source is separately stored in the knowledge base. A source may be an informant or a reference. For each informant, the date of the interview, informant's name, age and gender are recorded. For each reference, the name(s) of the author(s), date of publication, title and other information are recorded. At any point in time one knowledge source is set as the 'current source'. All knowledge entered is attributed to this source until it is changed. Alternatively the current source may be added to a list of sources to which a single statement of fact is already attributed. If there are no recorded sources or the current source is not allocated, AKT1 will demand source information or the allocation of the current source before knowledge can be entered. The source information associated with a knowledge base is manipulated by use of the Source menu (Figure 6.2).

To enter source information, 'Record new source' is selected from the Source menu and 'Reference' or 'Interview' from the sub-menu triggering the reference or informant dialogues (Figures 6.3, 6.4).

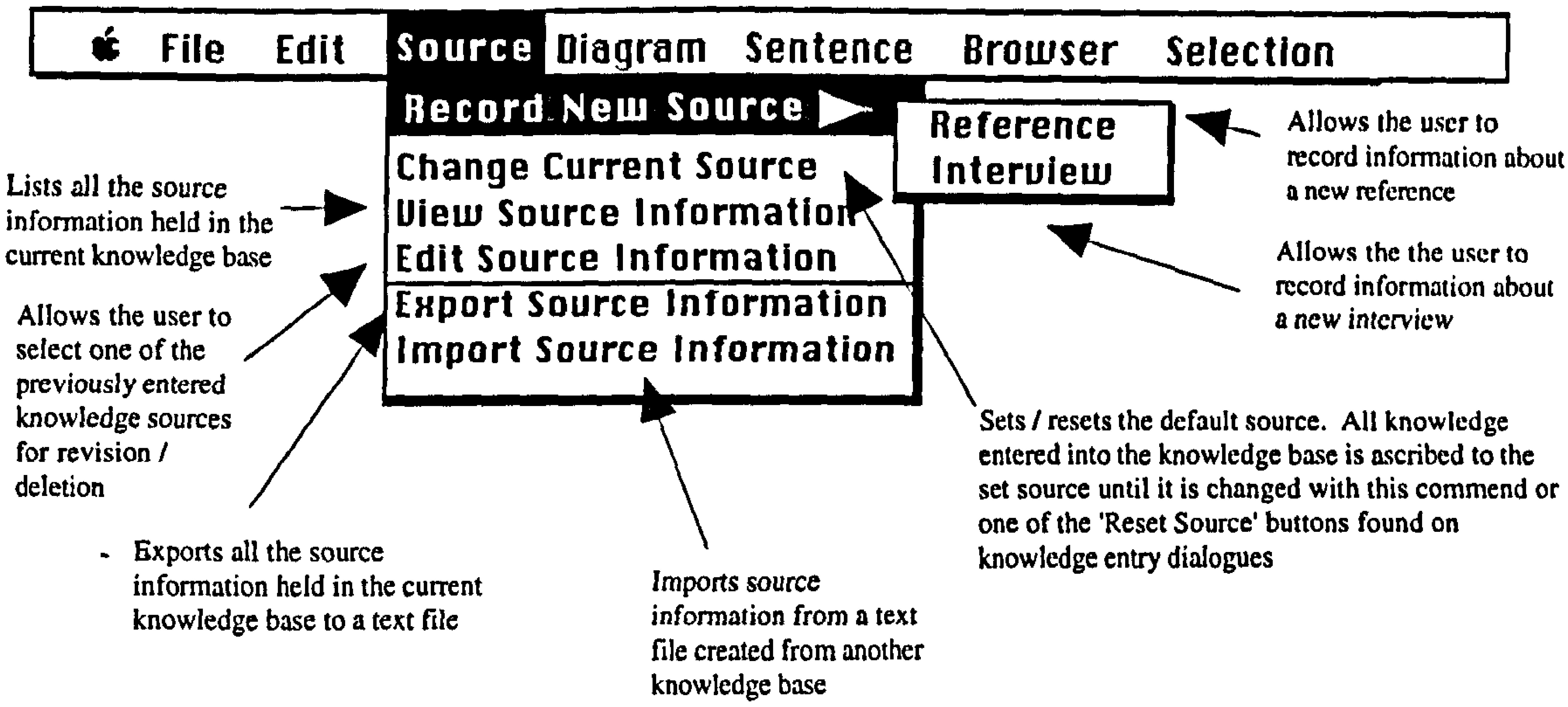


Figure 6.2 The source menu.

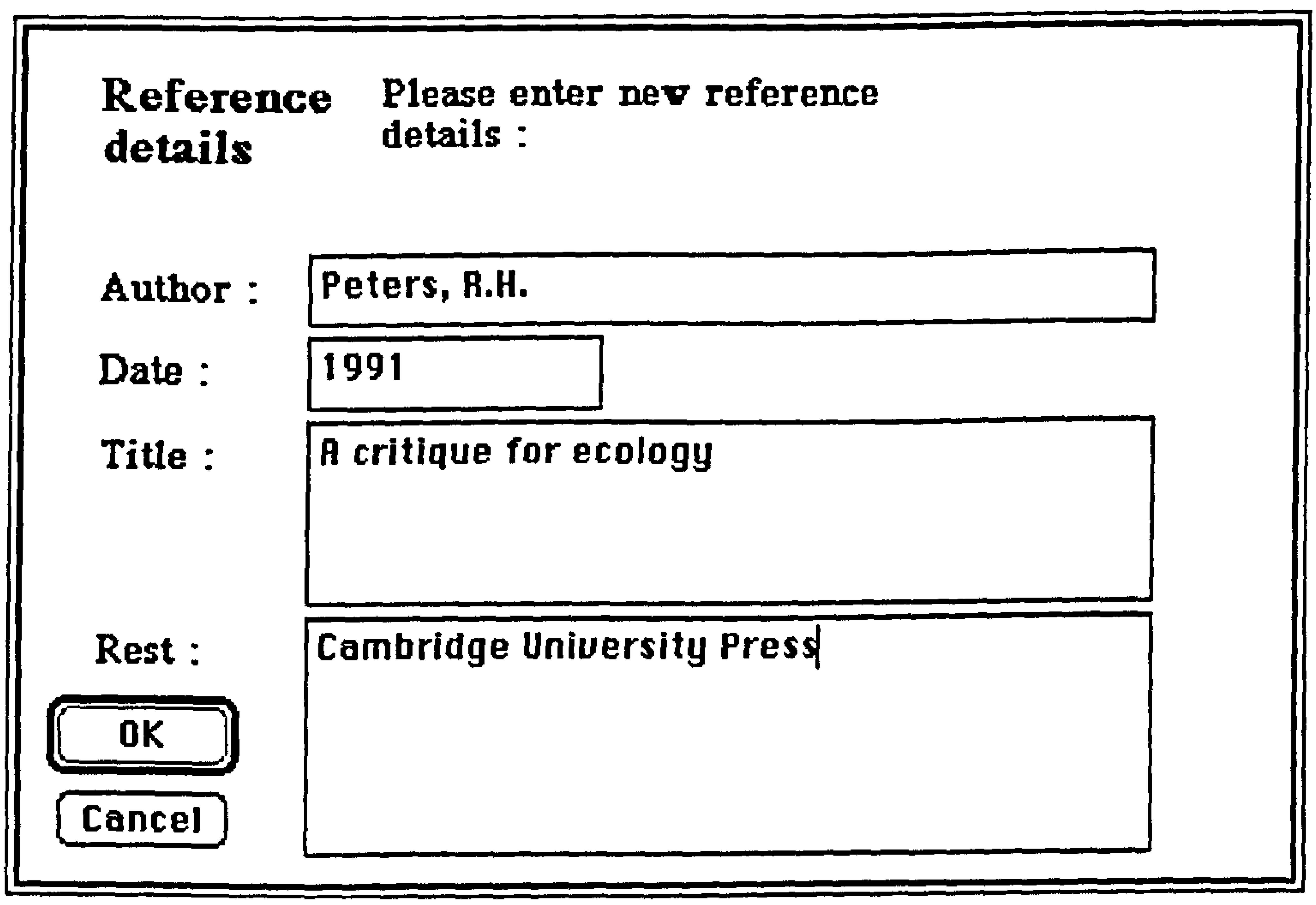


Figure 6.3 The reference dialogue.

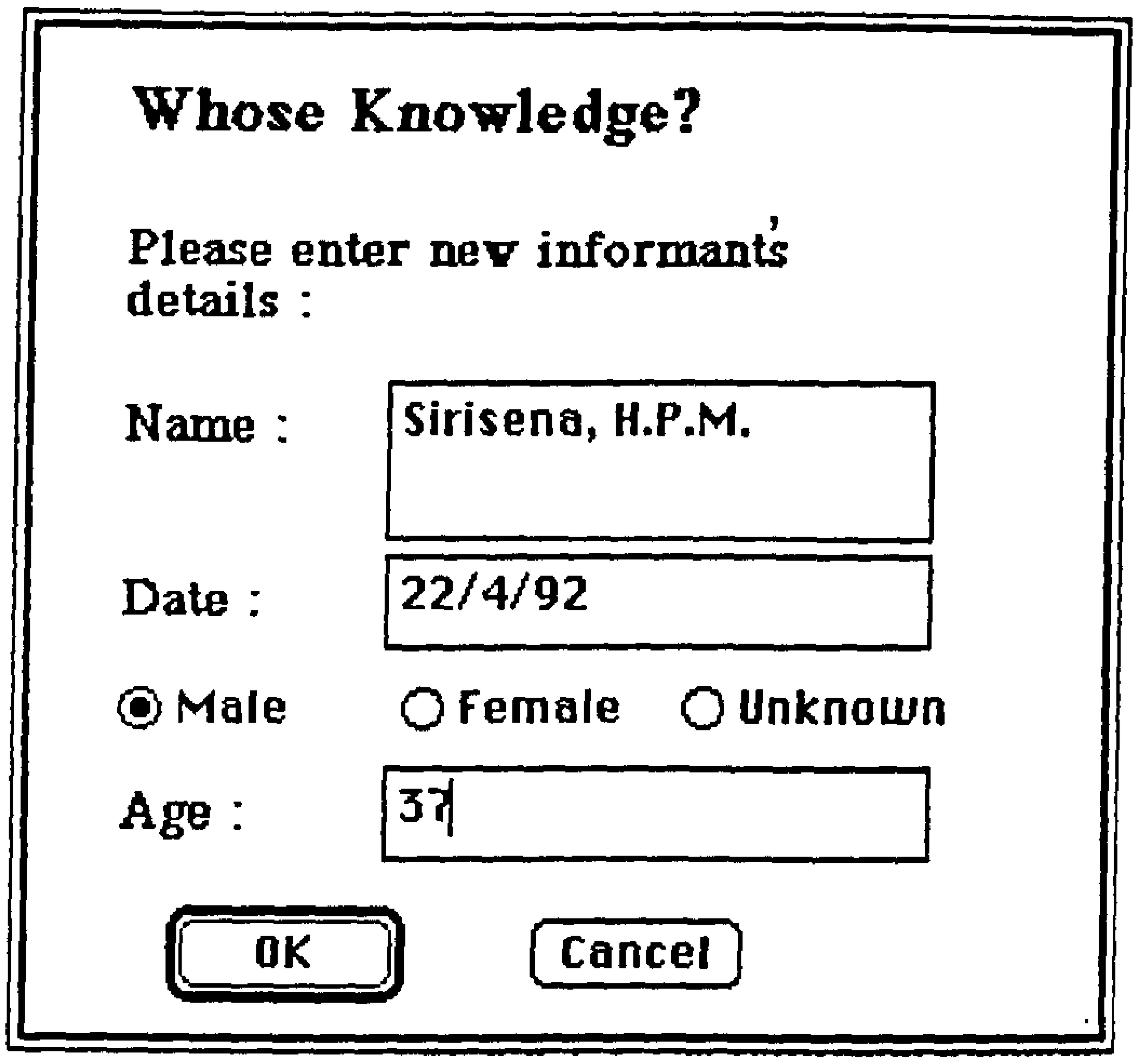


Figure 6.4 The informant dialogue

The current knowledge source is set through the 'Change current source' option.

If the relevant source has not already been entered a new informant or reference may be entered by using the 'New source' button.

'View source information' provides a listing of all the details of all the sources recorded in the knowledge base.

'Edit source information' provides a menu giving the name and date of each source (Figure 6.5). These can then be individually selected for editing or deletion. Choosing to edit brings up the original informant or reference dialogue (Figures 6.3, 6.4) with the current information for editing.

Because sets of information sources may be used in more than one knowledge base, the 'Export source information' and 'Import source information' options allow source information to be transferred between knowledge bases. The export option creates a text file of source information; this file can be taken into another knowledge base by using the import option when that knowledge base is open.

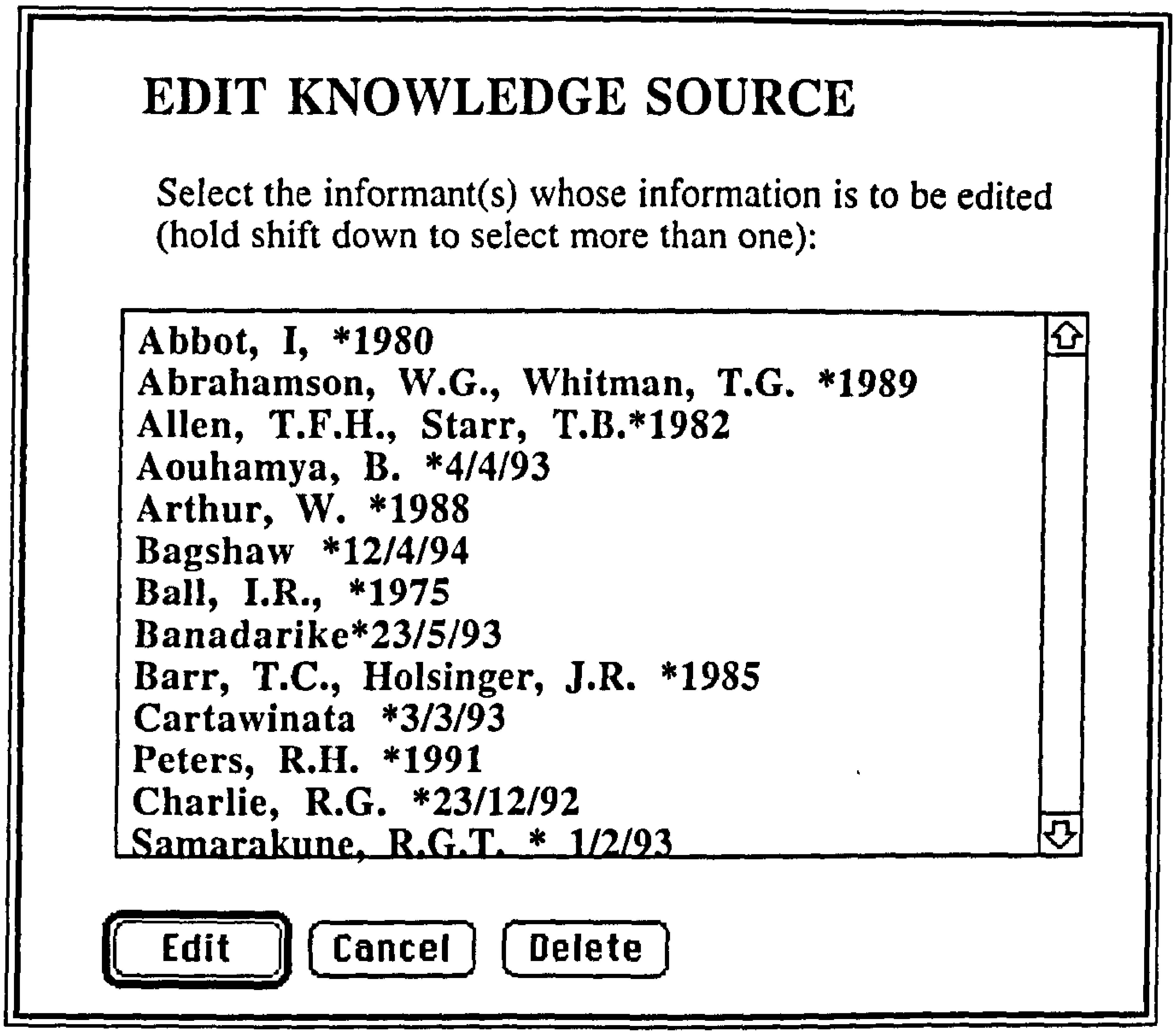


Figure 6.5 The edit source dialogue.

6.5.3 Entering knowledge through the diagram interface

The diagram interface provides a means of entering knowledge into the knowledge base as diagrammatic linked sets of statements of fact (Chapter 5). Whenever two diagram nodes are joined by a link, a corresponding intermediate statement of fact is generated. Additionally, full information on the structure and content of each hierarchically linked set of diagrams is stored such that the diagrams can later be viewed and edited. A single knowledge base may contain a number of separate diagram sets.

The diagram interface is called up by selecting 'New diagram set' from the Diagram menu (Figure 6.6). Each of the diagram sets contained within a knowledge base is appended to the bottom of the Diagram menu; a previous diagram set is opened by selecting its name from the Diagram menu. Diagram sets (and, optionally, associated intermediate statements of fact) are deleted by selecting 'Delete diagram set'. A diagram set is a group of diagrams that are hierarchically linked. A single knowledge base may contain a number of unlinked diagram sets each consisting of a number of linked diagrams.

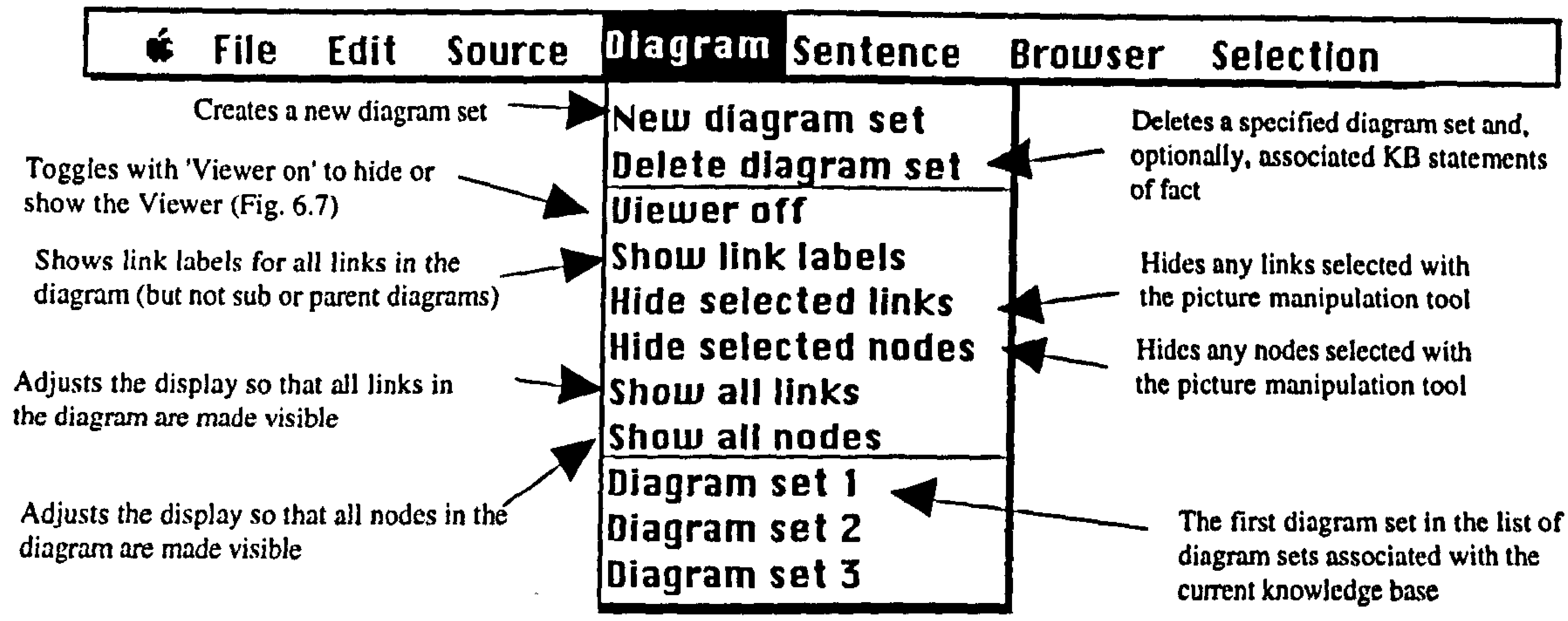


Figure 6.6 The Diagram menu

Figure 6.7 gives an overview of the diagram interface, showing part of the diagram presented in Figure 5.1.

(i) Entering statements of fact - creating nodes and links

The process of entering a statement of fact through the diagram interface is best documented by example. The first sentence of the text presented in Table 5.1 includes : "... deforestation results in a decrease in the rate of evapotranspiration ...." This can be represented as a pair of nodes and a joining link in a diagram. Deforestation is a process, so the process node tool is selected and the cursor is moved to an appropriate position in the drawing area. A node is created and a dialogue triggered which elicits the name of the node from the user. This dialogue also provides an optional opportunity to provide a definition for the meaning of the name of the node. The process is repeated for 'rate of evapotranspiration' except that this is a measurable property and is, therefore, represented using the attribute node tool.

The relationship between the two nodes is diagrammatically represented by creating a link between them. The link tool is selected from the tool window. The cursor is moved to the node at the beginning of the link (Deforestation), the mouse is clicked and held down and the cursor dragged to 'rate of evapotranspiration'. Releasing now creates a link between the two nodes.

The clarity of a diagram has been shown to be significantly influenced by the form of the arrows used as a link (see Figure 6.8). The arrows in the diagram interface are normally compound, comprising a horizontal and a vertical line joined by a curved corner. If the two nodes are, however, on, or close to, the same vertical or horizontal line a simple, straight line arrow is used. Diagrams with compound arrows are more easily interpreted than diagrams employing straight arrows.

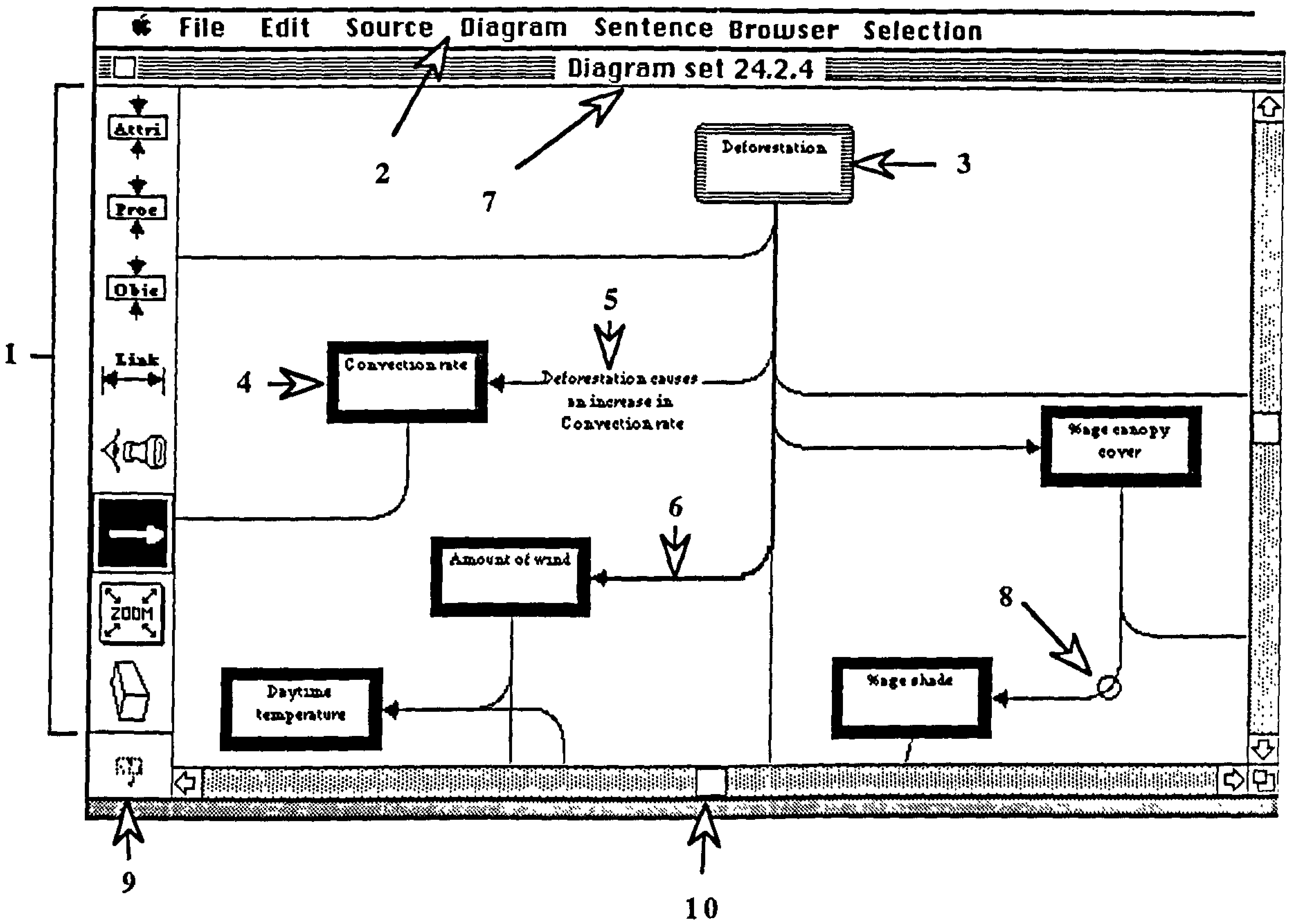


Figure 6.7 An overview of the diagram interface.
1. The nine tools in the tool window (from the top the attribute node tool, the process node tool, the object node tool, the link tool, the see-behind tool, the picture manipulation tool (highlighted as the currently selected tool), the zoom tool, the eraser and the viewer) provide the primary means of creating and manipulating the diagram.
2. The diagram menu also provides means of manipulating the diagram.
3. A process node (contrast with 4).
4. An attribute node (contrast with 3, object nodes are different again).
5. A link label. Link labels may be shown or hidden according to requirement.
6. This double thickness arrow indicates that there are two links from 'Deforestation' to 'Amount of wind'.
7. The name of this diagram. As this is the 'top level' diagram the name corresponds to the name of the diagram set. For subdiagrams the name is automatically generated.
8. A circle on a link indicates that it has been expanded into a subdiagram, accessed with the 'see - behind' tool.
9. The viewer provides a miniaturised overview of the diagram. The user can move around the diagram by dragging a small box within the viewer, i.e. the small icon box on the lower left of the screen.
10. The area visible at any one time is usually only a proportion of the total drawing area. The user can move around the drawing area using standard Macintosh scroll bars.

(a)

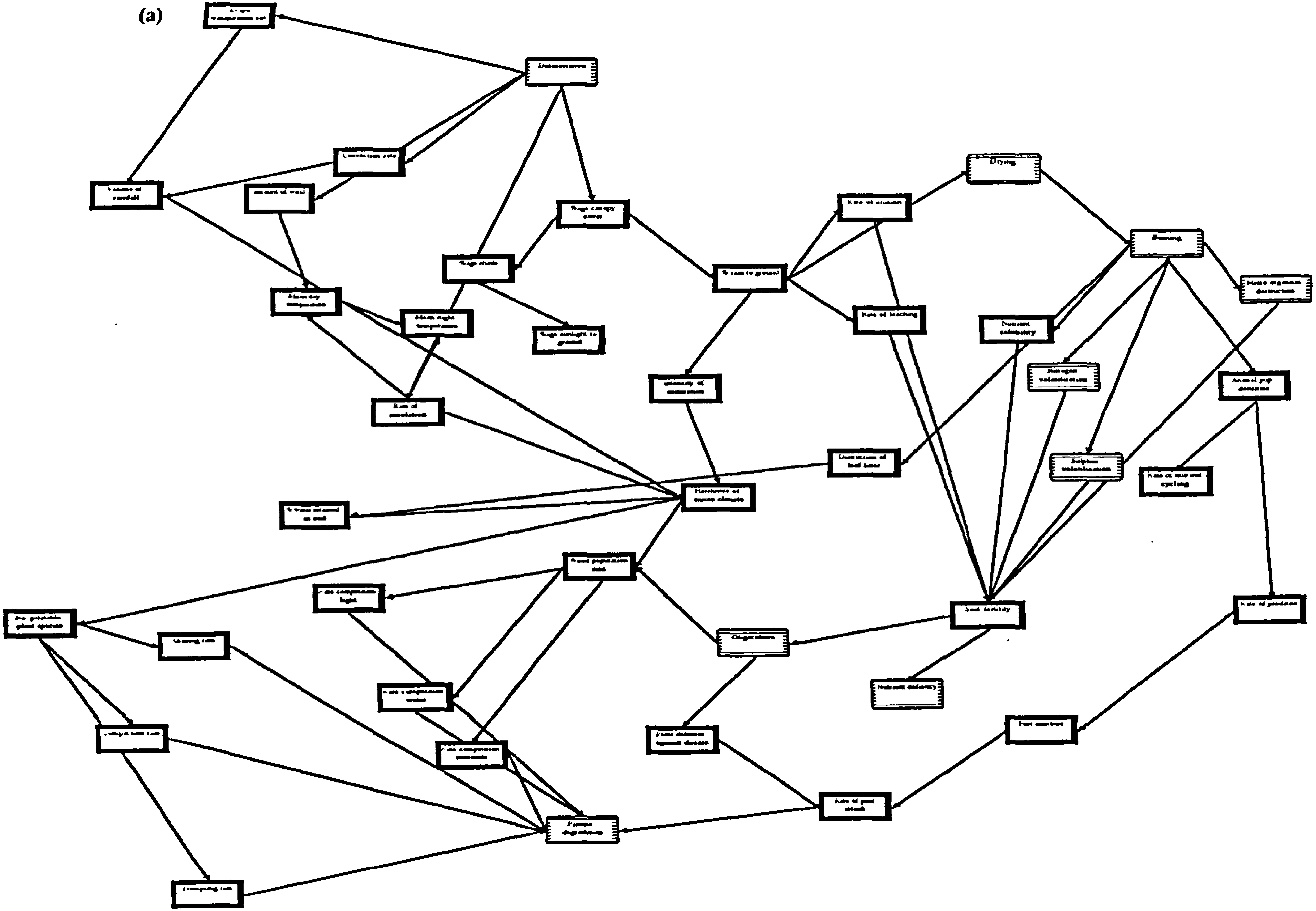


Figure 6.8 A comparison of straight (a) and compound (b) link arrows. Full link details for this diagram can be found in Figure 5.1.

(b)

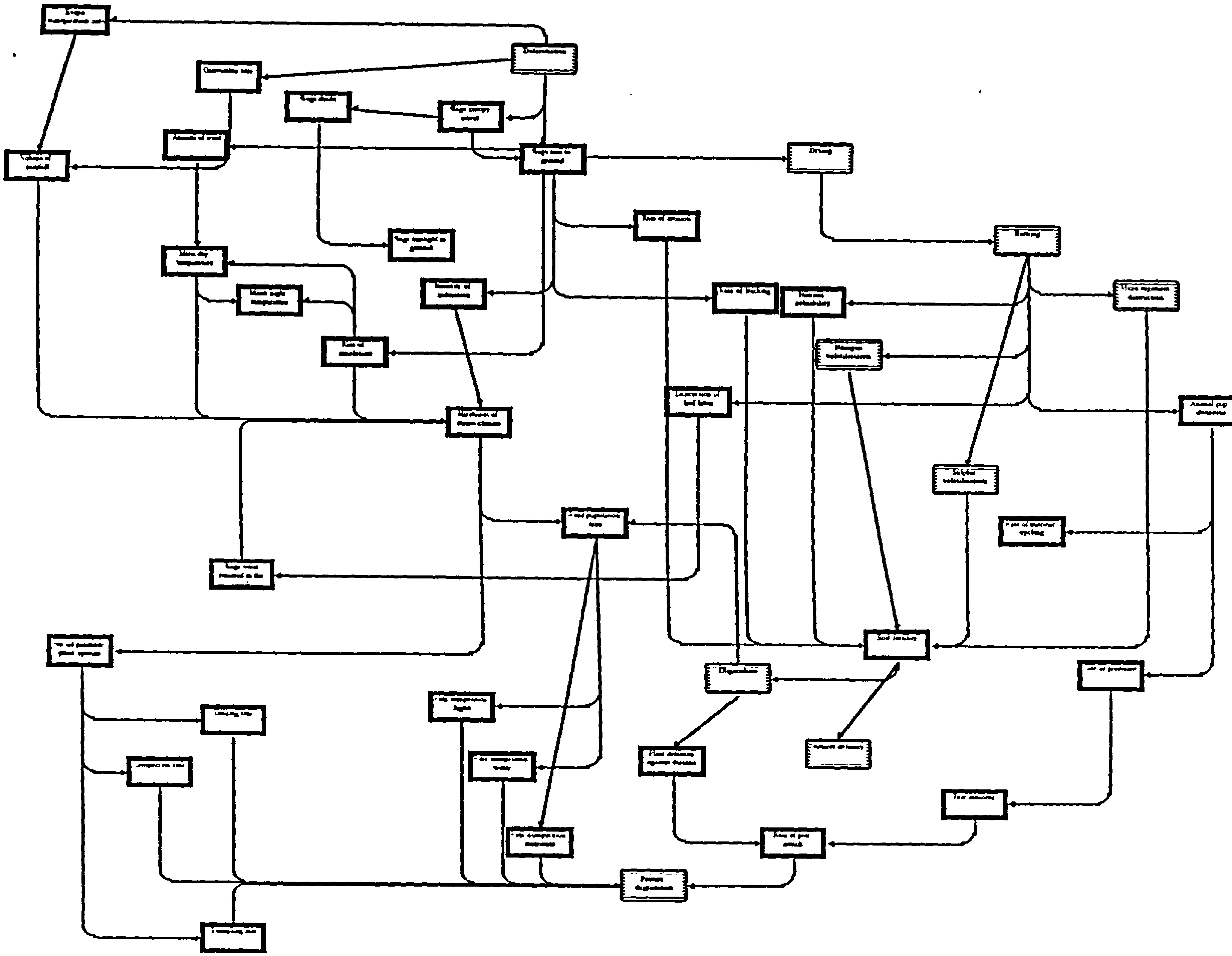
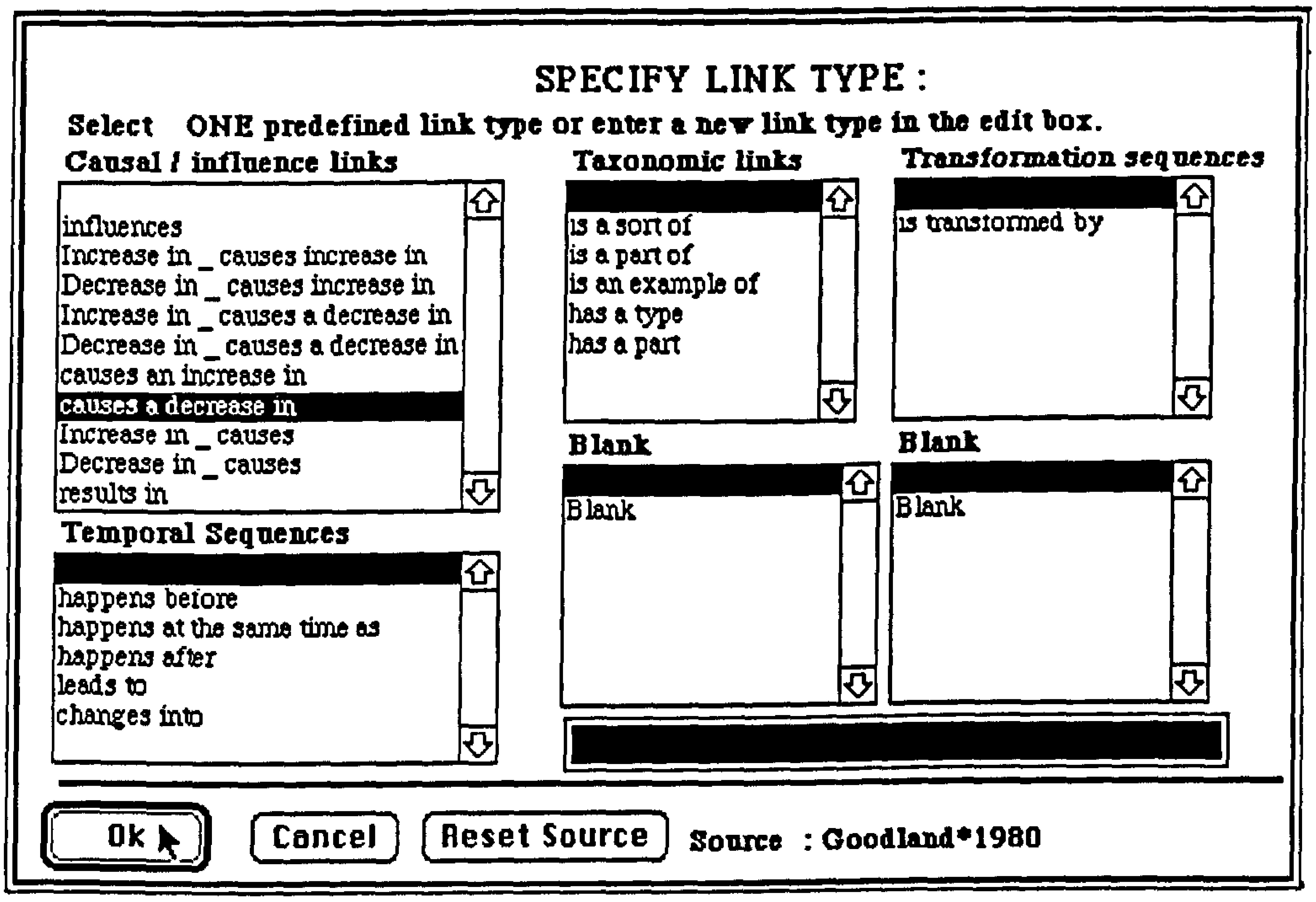


Figure 6.8 cont.

The diagramming paradigm implemented in AKT1 places emphasis on links (Chapter 5). As a result, creating a link triggers a series of dialogues eliciting further information about the link. The first dialogue box prompts the user for the meaning of the link (Figure 6.9a) by providing a predefined set of link types and giving the user the alternative option of creating a new link type. Selection from a predefined list of link types results in a more tractable knowledge base than unconstricted specification (Chapter 3). The second (Figure 6.9b) generates an English statement from the two nodes and link to allow the user to check the sense of the statement of fact entered. A further set of three dialogues then elicits any conditional information associated with the statement of fact (Figure 6.10).

There may frequently be more than one way in which two nodes are linked. For example, a set of different relationships between two entities may occur under differing sets of conditions. Multiple links between two nodes can be created in the diagram interface. The width of the arrow between two nodes is proportional to the number of links that it represents.

a)



b)

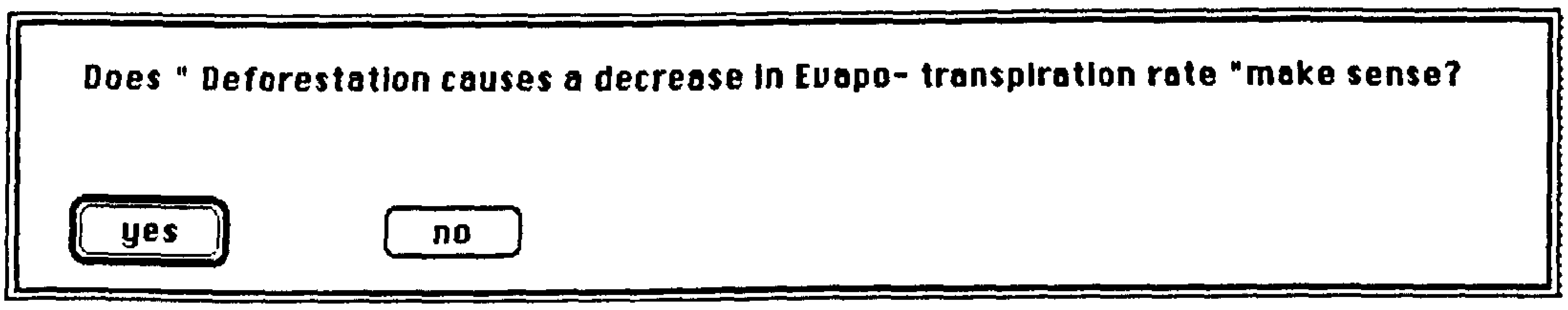


Figure 6.9 The link label (a) and sense check (b) dialogues.

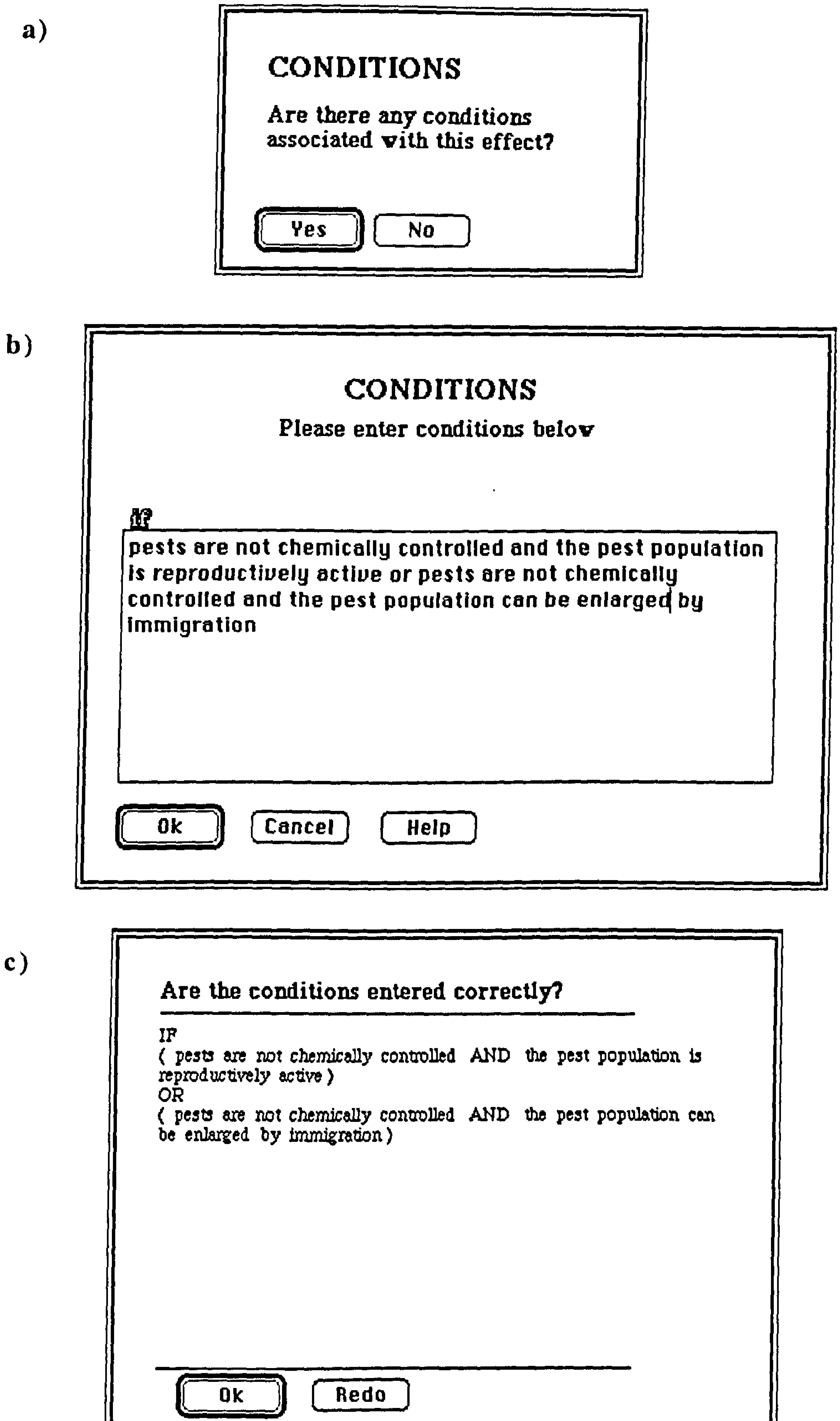


Figure 6.10 The link conditions dialogues. Selecting 'Yes' in dialogue (a) calls dialogue (b). Once conditions have been entered in dialogue (b), selecting 'OK' calls dialogue (c).

(ii) Viewing information associated with nodes and links

Not all the information associated with a link is visible in the diagram. However, this information can be accessed through the diagram. If the see-behind tool (Figure 6.7) is selected and the cursor clicked on a link, a pop-up menu is generated (Figure 6.11). Selecting 'Extra information' from the pop-up menu allows the user to view the knowledge represented by the link through the sentence interface (§6.5.4).

A dialogue giving full information can also be created by clicking on a node.

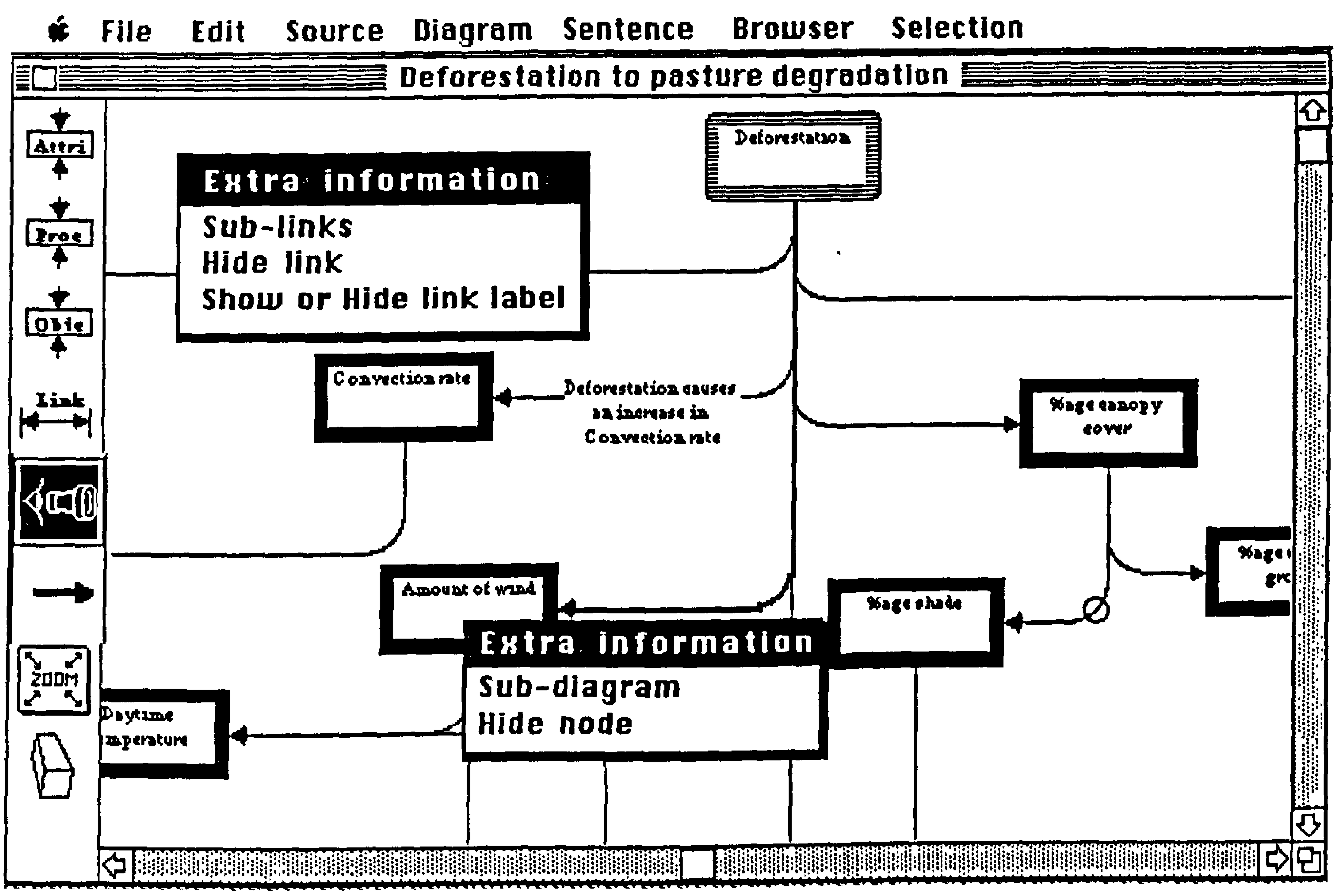


Figure 6.11 The node and link pop up menus - activated by the 'see-behind' tool.

(iii) Hierarchically structured sets of diagrams

The diagram interface provides means of generating hierarchically structured sets of diagrams (Chapter 5). Nodes or links may be selected and expanded into subdiagrams by selecting the see-behind tool, clicking on the node or link and selecting sub-diagram or sub-links respectively from the resulting pop-up menus. Where a link is expanded, the nodes at the start and finish of the link in the parent diagram are automatically generated in the sub diagram, but may be removed. Sub-diagrams are automatically named after their parent node or link. Once created sub-diagrams can be accessed again by selecting the sub-diagram or sub-links options from the node or link pop-up menus. To return to the top-level diagram, sub-diagrams must.be hidden by clicking in the 'go-away' box in the top left hand corner of the diagram.

Sub-diagrams may also be created by selecting a set of nodes (by clicking on each node with the diagram manipulation tool while holding down the shift key), and, thereby, their associated links, and selecting Create sub-diagram from the pull down Diagram menu. The user may then either replace the set of nodes and associated links by an object node or a link in the parent diagram.

(iv) Manipulating diagrams

The diagram interface contains a number of features for manipulating the contents of the diagram.

The position of nodes and links can be changed by clicking and dragging a node using the diagram manipulation tool; links are all automatically adjusted.

Individual links and nodes can be temporarily hidden using the 'Hide node' and 'Hide link' options in the node and link pop-up menus respectively. Alternatively, sets of nodes or links can be hidden by selecting the nodes or links with the diagram manipulation tool while holding down the shift key and then selecting 'Hide selected nodes' or 'Hide selected links' from the Diagram pull down menu. Hidden nodes and links can be made visible by selecting the 'Show all links' or 'Show all nodes' options from the Diagram pull down menu.

The textual statement of fact represented by a link can be shown on the diagram by selecting the 'Show or Hide link label' option from the link pop-up menu. Alternatively all the textual statements can be shown using the 'Show link labels' option under the 'Diagram' pull-down menu and hidden using the 'Hide link labels' option that replaces it. The position of link labels can be altered by clicking and dragging with the diagram manipulation tool.

The viewer provides a miniaturised overview of the diagram as a whole (Figure 6.7). The user can move over the drawing area by dragging a tiny box in the viewer. The continuous updating of the viewer significantly slows the diagramming process, so it can be turned on or off by selecting 'Viewer on' or 'Viewer off' from the 'Diagram' pull-down menu. In Figure 6.11, for example, the viewer is off (compare Figs. 6.7, 6.12).

Alternatively, the user can get an overview of the diagram by selecting the zoom tool and clicking on the diagram (Figure 6.12). The user can work at this zoomed-out scale. Clicking with the zoom tool again returns the diagram scale to the default.

Nodes or links and all associated information in the diagram and the knowledge base can be deleted by selecting the eraser tool from the tool window (Figure 6.7) and clicking on the node or link. Where a node is deleted all associated links and their associated information are automatically deleted.

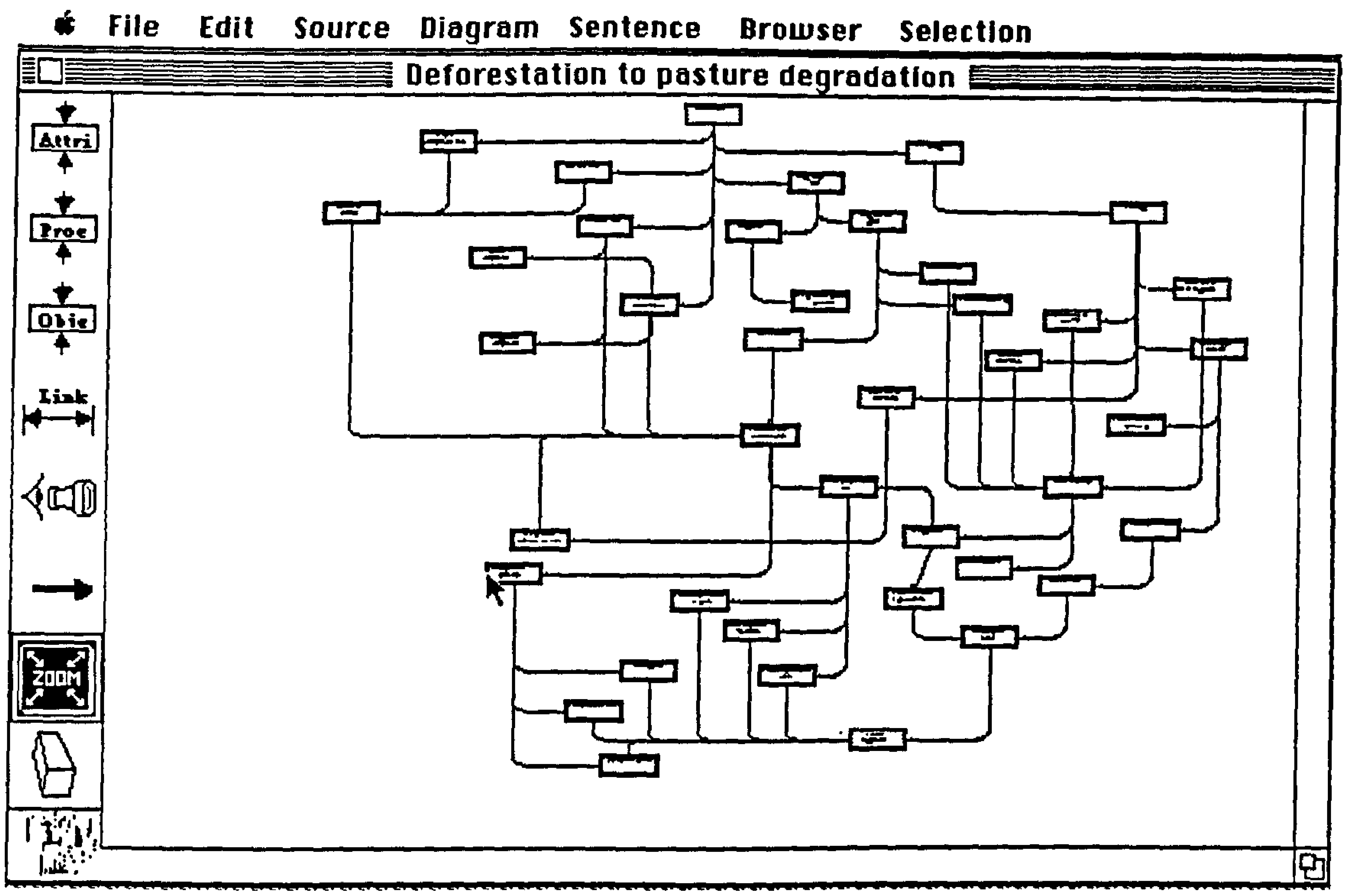


Figure 6.12 The diagram interface in 'zoomed out' mode showing the diagram in Figure 6.11.

6.5.4 Entering knowledge through the sentence interface

The sentence interface provides a means of entering unstructured, disaggregated statements of fact (Chapter 3) into the knowledge base. Other than the requirement that all knowledge be tagged according to source, there are no checks or restrictions on the form of the knowledge entered. Nevertheless, the sentence interface does provide mechanisms for eliciting all the components of the statement of fact. It also acts as the default interface for viewing a statement of fact. Figure 6.13 provides an overview of the sentence interface.

New sentences can be added to the knowledge base by selecting 'New Sentence' from the sentence menu (Figure 6.14) or by selecting OK on sentence interface if already visible. Existing sentences may be accessed by selecting 'View all' - which provides a list of all the statements of fact in the knowledge base (Figure 6.15) and double clicking on the relevant statement or clicking and selecting the Details button. For large knowledge bases this may be more effectively achieved by first doing a search (Chapter 3). Statements of fact entered via the diagram interface may also be accessed via the diagram interface (Chapter 5). The Write button on the View all interface allows the user to write all the statements of fact in the knowledge base to a separate text file.

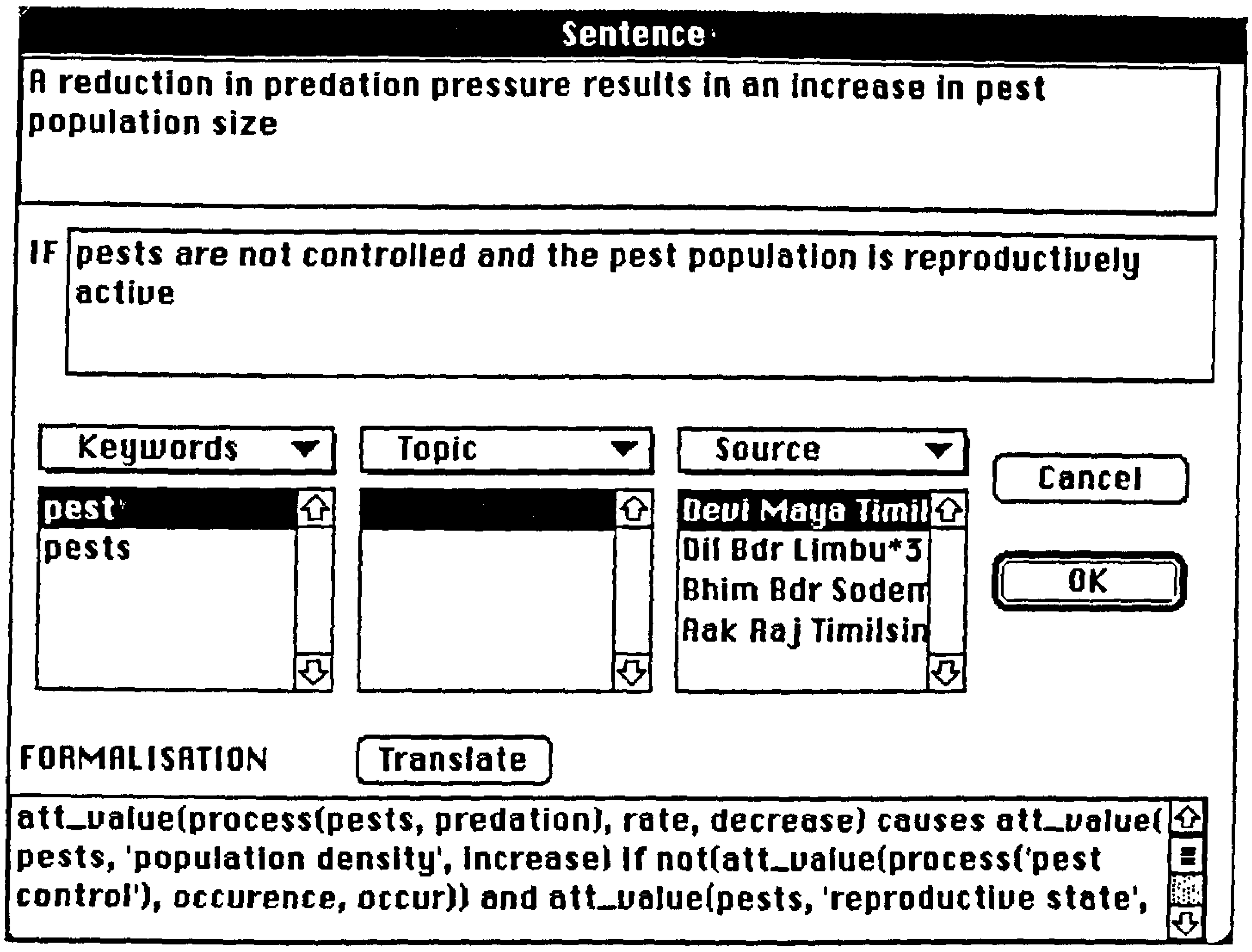


Figure 6.13 An overview of the sentence interface.

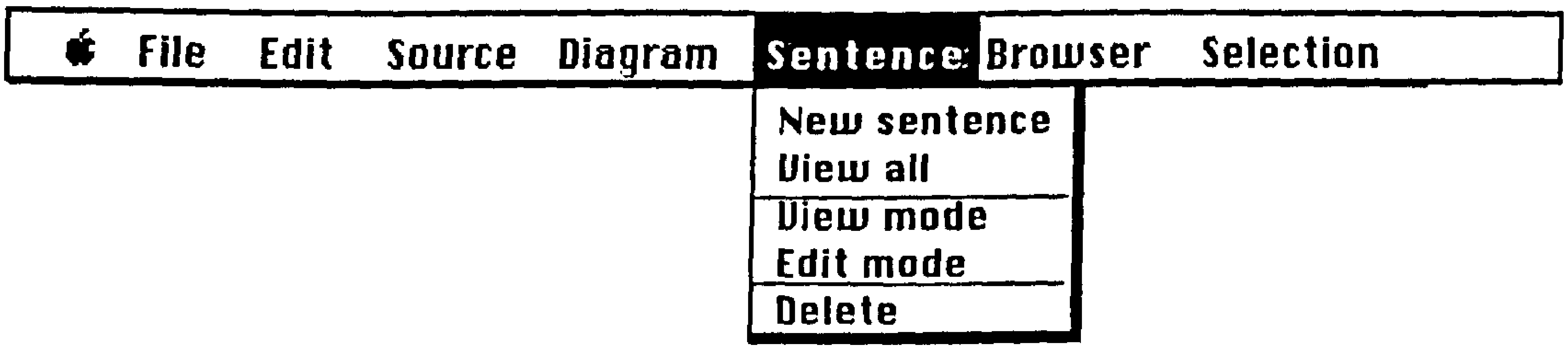


Figure 6.14 The sentence menu.

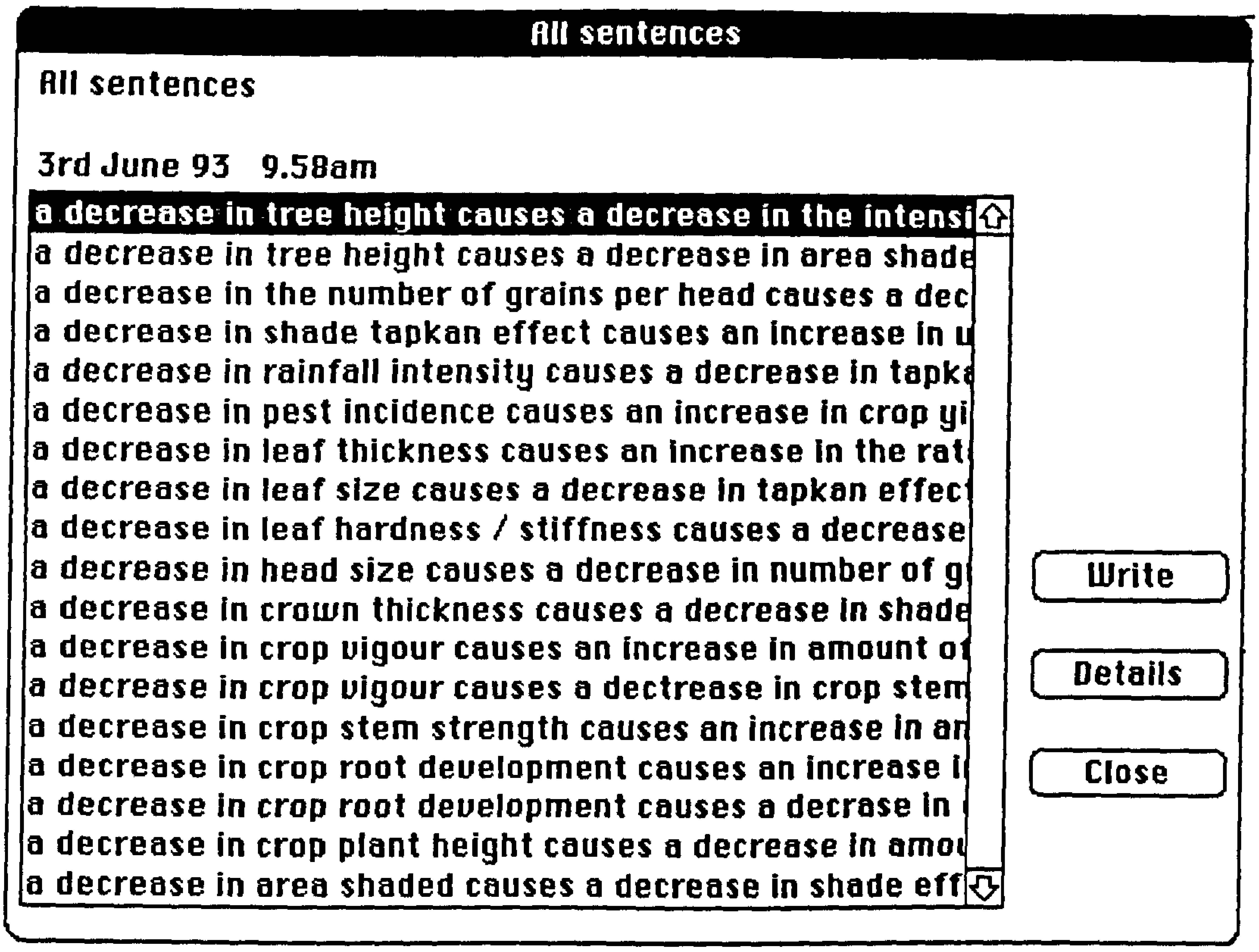


Figure 6.15 Statement set interface. Only a few of the statements are visible at any one time, but the scroll bar on the right-hand side of the list allows the user to view the other statements. In this instance the interface shown is generated by the view all command, but similar interfaces are generated to list subsets of the knowledge base identified through keyword searches.

Statements of fact and associated information may be viewed with the sentence interface in View mode or Edit mode (set in the Sentence menu, Figure 6.14). In the former, the details recorded may not be altered. When using the Edit mode for statements recorded through the diagram interface, only the topic, source and structural type information can be altered.

6.5.5 Creating a keyword hierarchy

AKT1 provides facilities for generating a hierarchy of keywords. This provides an indexing system for the otherwise unsorted statements of fact that constitute the knowledge base (§3.9).

The creation and manipulation of the keyword hierarchies are controlled by the Keyword menu (Figure 6.16). The keyword hierarchy interface is called up by selecting the Keyword option under the Keyword menu. Figure 6.17 provides an overview of the keyword hierarchy interface.

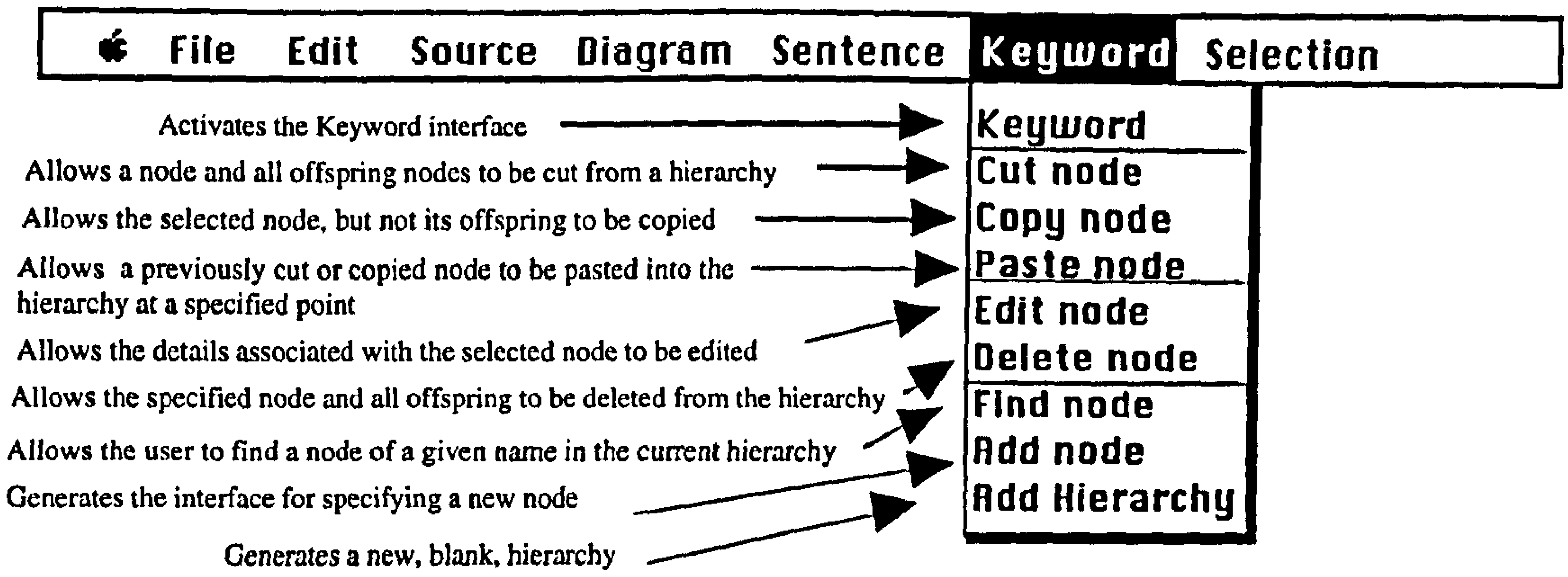


Figure 6.16 The keyword menu.

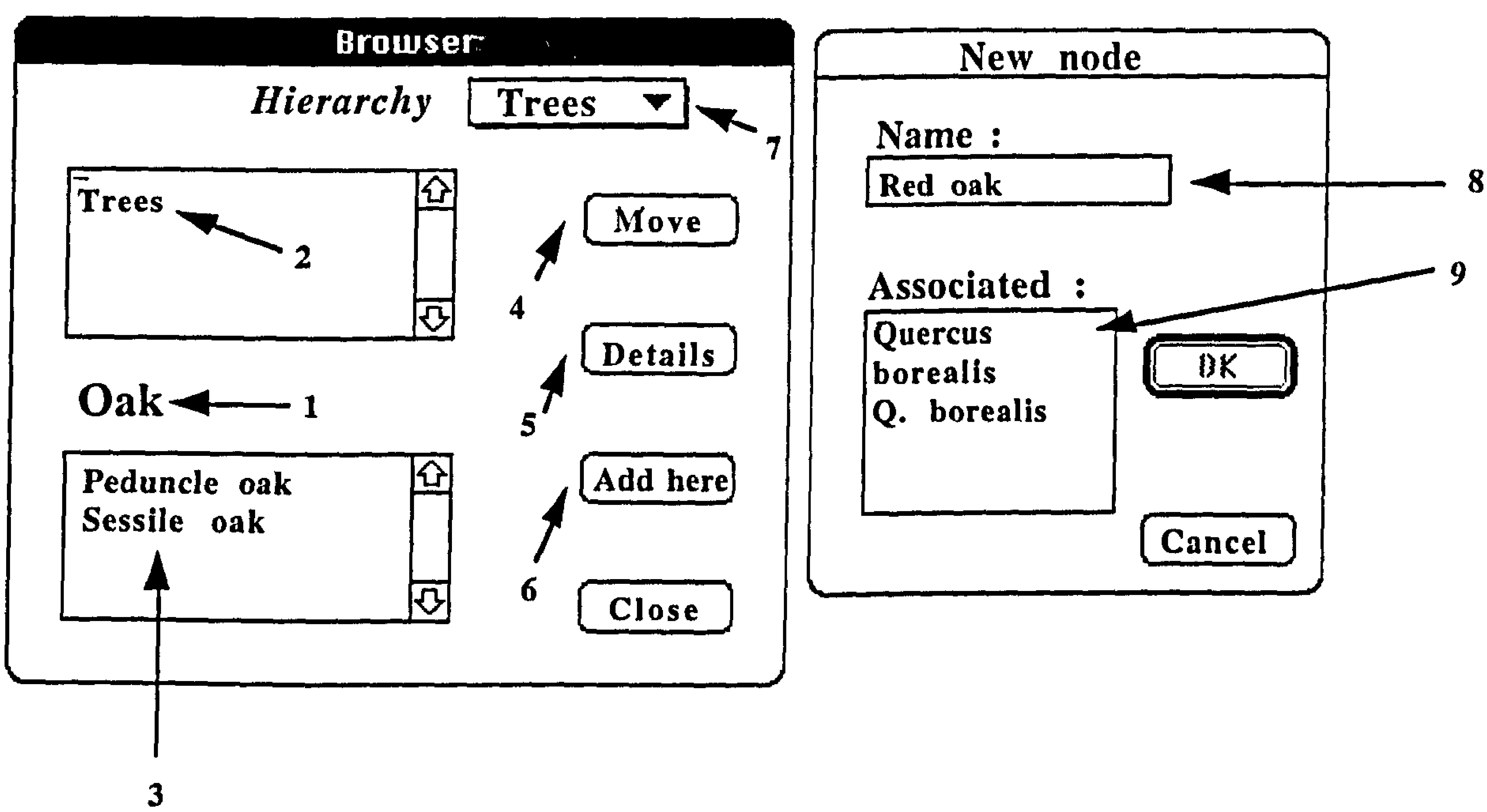


Figure 6.17 The keyword interface.
1. The current node
2. A list of 'parents' of the current node
3. A list of 'children' of the current node
4. The current node is reset by selecting a new node and clicking on Move
5. Details provides further information about the current node
6. Add here places the new node in the list of 'children' of the current node
7. The name of the current hierarchy. Selecting the arrow generates a list of all the hierarchies
8. The name of a new node about to be added
9. Synonyms for 8.

The keywords added into a hierarchy may occur in statements of fact (content keywords), may be attached to statements of fact as topic keywords, or may not relate directly to any statement of fact but provide a means of identifying the relationship between other keywords. A single knowledge base may contain a number of discrete hierarchies, and a single keyword may occur in more than one hierarchy. New hierarchies are added by selecting the 'Add hierarchy' item from the Browser menu.

The Source information in a knowledge base automatically generates a hierarchy with the names of informants or authors at one level and, at the next level down, details of all interview or publication dates.

New keyword nodes in a hierarchy are added using the 'Add node' command. The keywords and synonyms are typed in to the 'New node' dialogue but may be copied from statements of fact using the 'Copy node' (or 'Cut node') and 'Paste node' options.

'Edit node' allows the name of the node and its synonyms to be edited.

'Delete nodes' removes the selected node and all the descendent nodes in the hierarchy.

Because only a small portion of the whole hierarchy is visible at any one time, the 'Find node' option is used to track down keywords not currently visible.

6.5.6 Selecting subsets of the knowledge base

AKT1 enables the user to abstract subsets of a knowledge base according to user-specified criteria. This selection of subsets of the knowledge base may frequently be a part of the process of evaluating and using the knowledge base.

Selection is based on keywords, as identified through inclusion in the keyword hierarchies.

Search criteria are generated through selecting a set of keywords from the Keyword interface, specifying the relationship between them by selecting either AND or OR before selecting a further keyword.

Search criteria can only contain either AND linkages or OR linkages. However, more complex criteria may be generated using the 'alias' facility. Here a combination of keywords linked with, for example, ANDs can be saved as an alias. This alias may then be linked with other keywords using ORs or with other aliases using ANDs or ORs. An alias is created by specifying the search criteria and then saving it using the Alias button. Alternatively an alias may be typed straight in, provided that the user is clear about the correct use of brackets under such circumstances. Figure 6.18 shows the interface mechanism used.

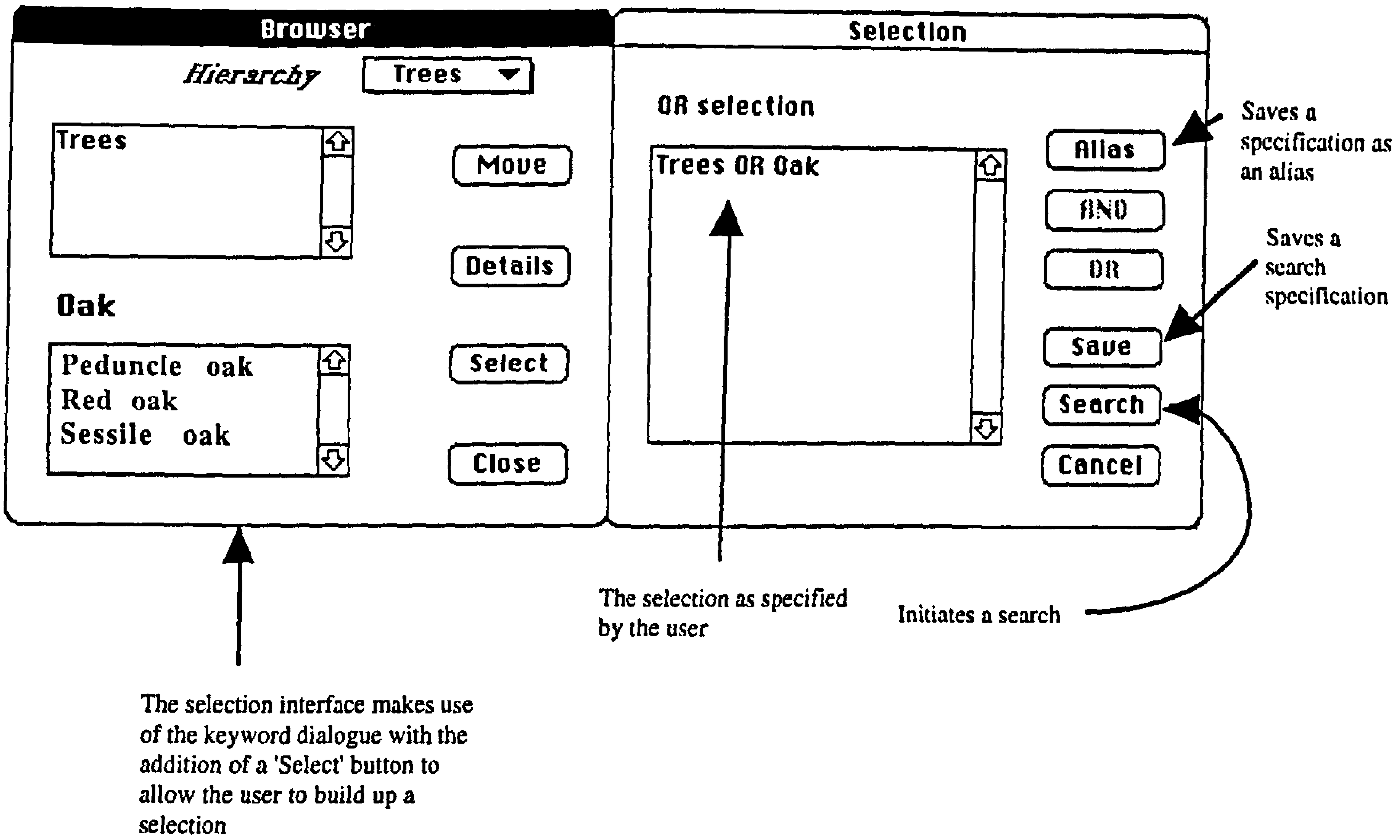


Figure 6.18 The selection interface.

Because the source information in the knowledge base automatically generates a keyword hierarchy, search criteria may specify the source to be considered.

Because particular combinations of search criteria may be frequently used, combinations can be saved using the Save button. The saved combinations become aliases. Saved combinations are appended to the bottom of the Search menu and can be called up by selection. Even where they are not saved, aliases may created using the Alias button; aliases created in this way may be combined with other keywords or aliases using the AND or OR buttons.

Once search criteria have been specified, the search can be instigated by clicking on the Search button. A choice of three search modes is available - 'node only', 'node and descendants', 'node and family'. 'Node only' recognises only statements of fact to which the keywords selected are actually tagged. 'Node and descendants' abstracts statements of fact in which either the specified keyword(s) or the daughters of those keywords in the hierarchy, the daughters of those daughters etc. are tagged. The 'Node and family' option additionally includes the parents of the selected keyword, their parents etc.

The results of a search are displayed as a collection of statements as in Figure 6.15. These sets will provide an input for reasoning tools for exploring, evaluating and using knowledge. Each set of sentences may then be given a suitable name and saved. The name is appended to the Selection menu (Figure 6.19) and may be recalled by selection from that menu.

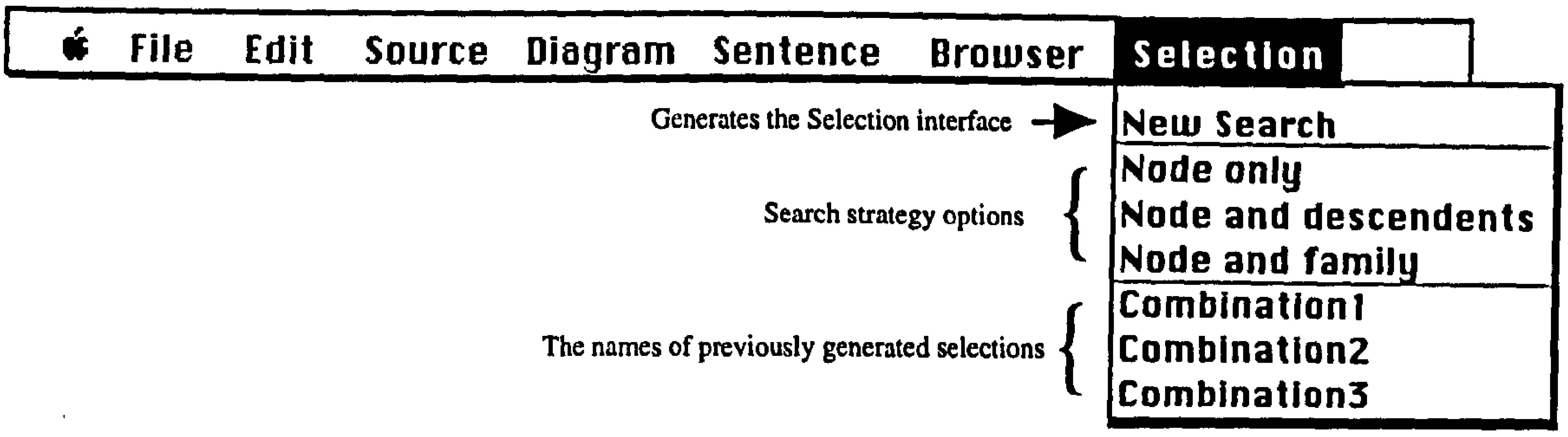


Figure 6.19 The selection menu.

6.6 MEMORY CONSTRAINTS

The most significant implementational limitation to the use of the toolkit relates to computer memory requirements. The AKT1 software is designed for Apple Macintosh machines operating System 7.0 or later and with at least 4 Mb of RAM. However, memory requirements have meant that the use of AKT1 on machines with only 4 Mb of RAM is significantly limited.

System 7 requires a minimum of 1.7 Mb to run. This means that with a 4 Mb machine, memory allocation to the AKT1 software can be set at up to 2.3 Mb. This available memory is divided by MacProlog between evaluation space and application heap space (Johns, 1992).

The evaluation space is the memory used in running a query. The application heap space is used for application code, terms, properties, text, window records and so on. As a result an appropriate ratio between evaluation and application heap space is required. Saving and, particularly, loading knowledge bases that contain diagram sets requires considerable evaluation space. The optimum ratio of evaluation to application heap space for the 2 Mb available on a 4 Mb machine will vary according to the knowledge base.

On Macintosh machines with more RAM or with a 68020 or later processor (in which hard disk space can be converted to RAM using the virtual memory facility) diagram size limitations are concomitantly decreased.

The introduction of a mechanism to enable the user to vary the evaluation space according to RAM available at any given moment (accessed through the 'About AKT1' option under the Apple menu) and traps to warn when diagrams are becoming too large to reload given current available evaluation space has alleviated, but not removed, these problems.

6.7 SUMMARY

The AKT1 software described in this chapter provides a software environment for the creation of knowledge bases conforming to the specifications for the intermediate and formal representation of ecological knowledge about agroforestry developed in Chapters 3 and 4 respectively. Intermediate statements of fact may be entered into the knowledge base textually or through the development of node-and-link diagrams. Each statement of fact may be stated as both an intermediate and a formal representation and is tagged with:

- associated conditional information;
- the source(s) of the statement;
- the keywords within the statement; and
- topic keywords about the statement.

The keywords associated with the statements of fact are separately identified within glossaries which may be hierarchically structured. Definitions may be provided for each keyword. Information about the sources of the statements of fact in the knowledge are separately stored in the knowledge.

In September 1993 functionality was added to AKT1, by Gill Kendon of the Department of Artificial Intelligence, University of Edinburgh, through the incorporation into the software of :

- a parser for checking the syntax of formal representations, enabling subsequent use of those formal statements in automatic reasoning; and,
- two reasoning tools.

The resulting toolkit is referred to as AKT2. The additional functionality found in AKT2 is detailed in Appendix 3.

Three parts of the specifications developed in Chapters 3 to 5 have not been implemented in the AKT software: storing customised secondary source information (§3.2.6), generating keyword definitions using a prescribed set of statements of definition (§3.4), and creating node-and-link diagrams as a means of viewing knowledge already entered into the knowledge base (§5.6).

The functionality of the secondary source information features was available within TEAK (Appendix 2) and has been demonstrated as having a clear utility (Thapa and Walker, 1992). It has not been implemented in AKT because it had already been tested whereas other specifications had a higher priority within the limited programming time available.

The implementation of mechanisms for the creation of definitions using a restricted set of template statements of definition and the creation of diagrams as an interface for viewing the changing content of the knowledge base are both implementationally demanding tasks and have not been possible within the limited programming time available. However, the utility of both approaches can be evaluated without full software implementation. This is undertaken along with the evaluation of the AKT software in Chapter 7.

The implementation of AKT1 provided a means of creating knowledge bases conforming to the specifications developed in Chapters 3, 4 and 5, thereby enabling an evaluation of the utility of these specifications (Chapter 7). Furthermore, experience in the iterative creation of the toolkit highlighted the sensitivity of the application of the approach to knowledge base creation developed in this thesis to effective software implementation.

CHAPTER 7 AN EVALUATION OF KNOWLEDGE BASE CREATION

7.1 INTRODUCTION

This chapter describes and assesses the creation of knowledge bases with the AKT software and associated approach.

Knowledge bases have been created by nine researchers to date. Thapa (in preparation) has created a fully implemented knowledge base, described in §7.7.1. Preechapanya (in preparation), Kilahama (in preparation), Jinadasa (in preparation) and Hitanayake (in preparation) are developing knowledge bases with the AKT software. The knowledge base created by da Costa (1993) is an intermediate representation in the AKT software, making use of both text and diagrams. The knowledge base created by Holmes (1993) is similiar in structure but paper based. The knowledge base created by Southern (in preparation) is an intermediate representation in AKT, with previous versions created in TEAK (Appendix A) and using a word processing package. The knowledge base created by Garde (1992) is entirely diagram based and was developed on paper.

The procedure for knowledge base creation that was developed for use by these researchers through iterative assessment of their knowledge acquisition activities is described and evaluated in this Chapter. For the purposes of evaluation, three discrete stages are distinguished in the procedure for creating a knowledge base; preparation (§7.2), knowledge base creation (§7.3) and evaluation of the representativeness of the knowledge base (§7.6).

Evaluation is achieved through :

- the reimplementation, within the AKT software, of the knowledge base on tree fodder management and evaluation in the hill region of Uttar Pradesh developed by Garde (1992);
- the combination of this knowledge base with the knowledge base on tree fodder in Nepal developed by Thapa (in preparation); and
- the addition to this combined knowledge base of further knowledge from key texts.

7.2 PREPARATION

Preparation for the creation of a knowledge base involves the specification of the objectives for the use of the resultant knowledge base (§7.2.1), definition of the boundaries of the proposed knowledge base (§7.2.2) and design of a knowledge acquisition strategy appropriate to meeting these objectives (§7.2.3).

7.2.1 Specification of objectives

Experience in supporting the practical application of the AKT approach has shown that specificity and clarity of objectives have a significant impact on knowledge base creation. The creation of a knowledge base involves identifying the knowledge to be included in the knowledge base and decisions on the form of representation. These decisions are most effectively and consistently made if taken in relation to an explicit set of objectives. For this reason, the first step in preparation for the development of a knowledge base is to clearly define the objectives for the use of that knowledge base.

Application of the knowledge-based systems approach has demonstrated that where knowledge acquisition has been undertaken without specific and tractable objectives, it will tend to result in knowledge bases of significantly lower utility than where objectives are clearly specified (compare Southern, in preparation, with the later research with more exacting objectives undertaken by Thapa, in preparation). In specifying objectives, it has been found necessary to strike a balance between the need for detailed and exact objectives to drive the knowledge acquisition process (Wilson, 1989) and the need to avoid compromising knowledge acquisition through inappropriate preconception (Johnson, 1989) particularly when working with indigenous knowledge (Knight, 1980; Werner and Schoepfle, 1987a). An appropriate balance may best be achieved by framing objectives in terms of intentions in the creation of the knowledge base, rather than as a specification for the structure and content of the knowledge base.

For example, the research in Nepal (Thapa, in preparation) was intended to :

```
Document explanatory ecological knowledge used in decision
making by farmers in managing their farmland tree fodder
resources in order to better inform national and regional
research efforts.
```

Subsidiary objectives for the same research included:

```
Consideration of the distribution of knowledge between and
within communities.

Comparison of the local knowledge with scientific knowledge.
```

These objectives for using the contents of the knowledge base(s) provided a framework for its creation.

The iterative nature of the process of knowledge base creation (§7.3), provides opportunities for the reassessment and modification of objectives.

7.2.2 Boundaries of the knowledge base

The knowledge base is an arbitrary unit. Two knowledge bases may be merged into a single knowledge base or a single knowledge base may be split into two. Furthermore, different sets of knowledge may frequently be needed for different tasks. The mechanisms for selecting a subset of an intermediate knowledge base (§6.5.6) or developing an extract of a formal knowledge base (Appendix B) provide means of partitioning a subset of the knowledge base, effectively creating temporary knowledge bases for particular purposes.

It is desirable to create fewer knowledge bases covering a broader range of knowledge rather than many small knowledge bases. This reduces repetition of core knowledge on the domain between knowledge bases. However, ensuring consistent and unambiguous use of terms, coherence and completeness becomes an increasingly demanding task as the knowledge base grows in its breadth of coverage. Furthermore, implementational constraints have an increasing impact as the knowledge base grows in size. In practice, it has proved necessary to define clear boundaries for a knowledge base and to represent unrelated knowledge in different knowledge bases in order to ensure the creation of tractable results.

On this basis, the second stage in the process of preparation is the definition of the boundaries and contents of the proposed knowledge base. As with all other stages in the process, boundaries have been defined with reference to the stated objectives.

For example, on the basis of the objectives given in §7.2.1 and consideration of the domain Thapa (in preparation) identified three discrete topics that had to be investigated in order to generate a comprehensive record of the 'explanatory ecological knowledge used by farmers in managing their farmland tree fodder resources': the propagation of tree fodder resources, tree-crop interactions and the selection and evaluation of fodders for livestock. As a result, three knowledge bases were created.

As with the specification of objectives, the specification of boundaries may be iterative. Initial knowledge acquisition by Thapa demonstrated that while these three topics merited consideration, much basic understanding was common to all of them. The knowledge base boundaries were, therefore, modified by merging the three knowledge bases.

Identification of the boundaries of the knowledge base provides a specification upon which the knowledge acquisition strategy (§7.2.3) can be based.

7.2.3 Designing a knowledge acquisition strategy

It will not usually be possible to elicit knowledge from all appropriate sources when creating a knowledge base about agroforestry. As a result, a sampling strategy must be designed that facilitates the efficient development of a knowledge base that is representative of the knowledge of a defined community, or set of communities, and enables the evaluation of the resulting knowledge base against the knowledge held by those communities.

To facilitate this task, a framework for designing a knowledge acquisition strategy appropriate for a particular instance was developed. This framework was based on Knight's (1980) methodology for investigating indigenous knowledge in order to create models of the cognitive systems associated with agricultural practices for particular societies.

Knight's methodology is divided into three stages:

- **specification** of terminology and cultural domain;
- **formalization** of queries for discovery of taxonomic relationships, specification of constructs or components of cognitive structure, and dictionary writing; and
- **generalisation** of results across a large reference group.

Adaptation of this approach for use in the current context has included the addition of an 'Introduction' stage.

The four stages, as applied in knowledge acquisition in the creation of an AKT knowledge base, are outlined below.

(i) Introduction

The detailed design of a knowledge acquisition strategy is best undertaken during a period of introduction and establishment with the source community. Attempts to design a detailed knowledge acquisition strategy prior to fieldwork have been found to be inappropriate. The introductory period serves to :

- familiarise the researcher with the source community (where applicable), and vice versa;
- allow adjustment of the basic objectives in knowledge base creation through refinement of problem specification;
- provide a preliminary assessment of the basic and universal information held by the community on the topic in question;

- help to identify suitable informants for later stages (where applicable); and
- identify parameters within the community that might account for differences in knowledge (where applicable).

The introductory stage may involve a questionnaire survey of heads of households, or all members of the community, or a random sample of these groups (see Thapa, in preparation; Kilahama, in preparation; Jinadasa, in preparation).

(ii) Specification

The specification stage is used to develop an overall understanding of the domain in question, defining boundaries and identifying terminologies.

Sources are purposely non-randomly selected from the source community. These 'key informants' have been selected from local communities on the basis of interest, articulateness, depth of knowledge and willingness to participate (see e.g. Southern, in preparation). Key informants known to be in some way significantly unrepresentative of the community as a whole (for example school teachers in local farming communities) have tended to be avoided.

Because specification has a significant impact on the 'shape' and content of the final knowledge base, an adequate spread of key informants is desirable to maximise the chances of developing a representative framework within the time available for this stage of knowledge acquisition.

(iii) Formalisation

The formalisation stage (not to be confused with formal representation) of the knowledge acquisition strategy is used to record detailed knowledge within the framework developed in the specification stage and to indicate the variability of knowledge over the community as a whole.

For local communities, Knight (1980) suggests a random, but small, sample from the source community. In the current context, a stratified random sample has been found to be more appropriate. Key determinants of variability in knowledge relative to research objectives (e.g. gender, altitude and ethnic group; Thapa and Walker, 1992) within the source community are identified. Appropriate groups combining these factors to allow assessment of the influence of these factors on the distribution of knowledge were identified. Within each group a random selection of informants was made.

A similar approach can be applied to professional communities (by discipline, for example). However, the relatively small size of professional communities or bibliographies mean that in practice the distinctions between specification and formalisation may be blurred.

(iv) Generalisation

Generalisation is intended to assess the representativeness of the abstracted knowledge base of the distribution and coherence of knowledge about the domain within the community and to facilitate appropriate modification.

Generalisation is returned to in §7.6.

The number of informants selected for each stage depends on the nature of the source communities, the size and quality demanded of the knowledge base and the time available. The number of informants grows at each stage. So, six key informants might be adequate for specification, 20 for formalisation and 100 for generalisation in a relatively homogeneous local community. In general it has proved more productive to speak to fewer people on more occasions than to cover a larger number.

The following strategy (from Thapa, in preparation) is typical of knowledge acquisition strategies undertaken to date :

> **Introduction** - Questionnaire survey of 10% of the heads of households of each of the nine wards in the study village, totalling 60 heads of household and detailed tree inventory on the farmland of 30 of these households.
>
> **Specification and Formalisation** - A stratified random sample of farmers divided according to gender, ethnic group and altitudinal zone, totalling 40 informants.
>
> **Generalisation** - A questionnaire derived from the content of the knowledge base created during specification and formalisation on a random sample of 100 farmers.

Given that a knowledge base may be created that contains knowledge of a set of identifiable topics it may often be desirable to go through the procedure of identifying informants for each topic separately.

7.3 KNOWLEDGE BASE CREATION

Having identified objectives, boundaries and an appropriate knowledge acquisition strategy, creation of the knowledge base can start. The procedure for knowledge base creation that was developed is diagrammatically summarised in Figure 7.1.

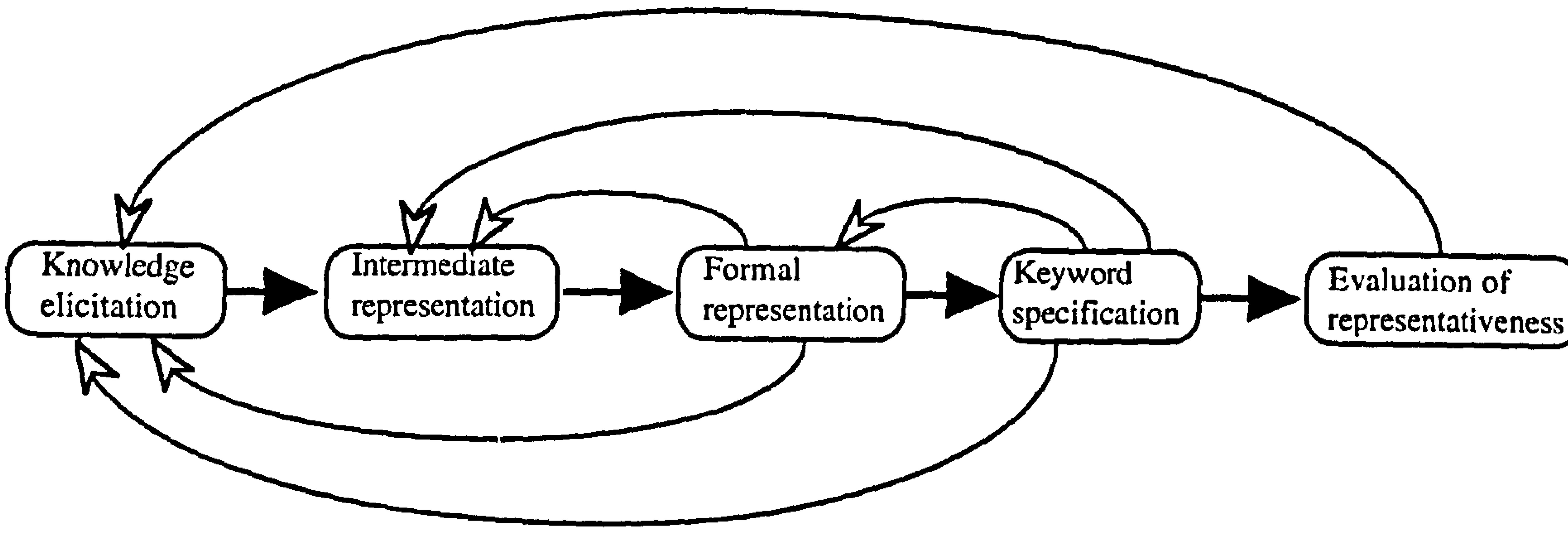


Figure 7.1 Activities in the creation of a knowledge base. Creation of the knowledge base is a linear activity (bold arrows), but, iterative evaluation during the creation of the knowledge base and consequent return to previous activities (fine arrows) means that all the stages in the process can occur concomitantly throughout the creation of the knowledge base.

While, in principle, the process of creating a knowledge base is linear, in practice its key feature is its iterative nature. The importance of iterative knowledge acquisition has been clear from the outset of this research (Walker *et al.*, 1991). The iterative evaluation of the knowledge base has also been shown, through practical experience, to have a central role in all stages of knowledge base creation.

In particular, intermediate and formal representation have proved to have a significant utility in iterative evaluation of elicited and represented knowledge, driving the continuous improvement of the knowledge base in relation to objectives. Knowledge elicitation and representation were initially envisaged as representing discrete processes. Appreciation of the value of intermediate representation in driving field work led to the development of TEAK (Toolkit for Eliciting Agroforestry Knowledge), (Haggith *et al.*, 1992 and Appendix A). At this stage, intermediate representation during knowledge acquisition, and formal representation after knowledge acquisition, were proposed as discrete activities. Subsequently, the relative ease of formal representation (§7.7.3) and the power of the process of formal representation in clarifying the content of the knowledge base (§7.7.3) resulted in a further concatenation of activities, such that knowledge elicitation, intermediate representation, formal representation and planning further elicitation have become a daily cycle.

This change has engendered a realisation that the knowledge base not only represents an output of the process but also has a central role throughout as a dynamic reference point for progress. The discipline of creating a useful knowledge base has, in fact, proved more important in understanding the domain in question than the output itself for all applications to date. The development of an iterative methodological cycle from an originally linear activity has influenced the development of the approach as a whole and has greatly facilitated the creation of useful knowledge bases.

In the following section, the processes of knowledge elicitation, intermediate representation, formal representation and keyword specification are outlined. Each of these stages involves iterative evaluation of a growing knowledge base.

7.3.1 Knowledge elicitation

The creation of knowledge bases in association with the research described here has relied heavily on semi-structured interview (see e.g. Garde, 1992; Southern, in preparation; Thapa, in preparation). Typically, the knowledge base creator identifies, on the basis of consideration of the current state of the knowledge base and, particularly, previous interviews with the same informant, a set of three to five issues to be covered in an interview. These provide a checklist to help the interviewer to structure the conversation. Interviews are typically between 45 minutes and one hour. Interviews have usually been tape recorded.

More specialised techniques may be appropriate under particular circumstances. Ethnographic fieldwork (Gorden, 1969; Spradley, 1979; Dobbert, 1982; Ellen, 1984; Kumar, 1987; Werner and Schoepfle, 1987a, b) and knowledge elicitation techniques (Hart, 1986; Jackson, 1986; Waterman, 1986b; Cordingley, 1989; Diaper, 1989; McGraw, 1989) are well documented.

Particular techniques are appropriate for particular types of knowledge, domains, cultural circumstances, research objectives and researchers. Southern (in preparation) and Jinadasa (in preparation) have made use of interactive diagramming in knowledge elicitation. Southern (in preparation) and Thapa (in preparation) have made use of mapping and slip sorting techniques. Questionnaires have been used to provide some basic information (§7.2.3). Other techniques are reviewed in Walker *et al.*, (1991).

7.3.2 Intermediate representation

Techniques for intermediate representation have been discussed in Chapters 3 and 4.

To date, intermediate representation has been primarily achieved through the textual representation of intermediate statements of fact. Some use has been made of diagramming (Garde, 1992; Walker and Southern, 1992; Holmes, 1993; da Costa, 1993; Thapa and Walker, 1993). While more demanding, experience demonstrates that this approach results in more tractable and useful intermediate knowledge bases.

Knowledge acquisition for all of the knowledge bases currently developed was undertaken, at least in part, prior to the development of the formal grammar and techniques for its application. As a result, techniques for manipulation and iterative evaluation and modification of the intermediate knowledge base played an important role in the development of coherent and complete knowledge bases. However, the utility of immediate formal representation (§7.3.3) means that evaluation and modification of the intermediate representation is now a less significant activity in the creation of the knowledge base.

7.3.3 Formal representation

The process of formal representation (§4.6) has proved to be the most powerful means of facilitating the explicit and unambiguous statement of knowledge.

The knowledge base developers who have applied the approach to date have rapidly become sufficiently proficient at formal representation to formalise the majority of intermediate statements of fact collected. For this reason, formal representation can be undertaken throughout the process of creation of the knowledge base.

7.3.4 Keyword specification

Structural manipulation of the knowledge base involves the development of consistent and useful glossaries of terms and ensuring the consistent use of those terms within statements of fact.

The development of glossaries of attributes, processes and link types is automated in the current version of the AKT software (AKT2 v.2.7). Objects are automatically identified but the user is required to identify 'type of' relationships and 'part of' relationships through the creation of keyword hierarchies and application of the formal grammar respectively.

However, while the mechanisms for the creation of glossaries and hierarchies are straight forward, the creation and management of hierarchies and glossaries has proved to be a demanding task. The consistency of use of terms has had a profound impact on the utility of the resulting knowledge base, particularly in relation to the use of automated reasoning tools. It was facilitated by:

- minimising the number of object, attribute, process and link terms used;
- ensuring the consistent use of values for attributes; and
- providing definitions for each term, such that their use is transparent and can be assessed by the knowledge base developer or another user.

The creation of object hierarchies makes further demands. Sets of objects can be hierarchically classified in many different ways for different purposes. Indeed it may not be possible to develop a single classification of a set of objects that is appropriate for all the intended uses of a knowledge base.

Figure 3.4 illustrated three different classifications of the same set of objects included in a single knowledge base. Because the identification and creation of appropriate hierarchies is a demanding process, it has been found to be expedient to develop a single flat hierarchy of objects until such time as the knowledge base is sufficiently 'mature' to merit the development of structured hierarchies.

7.4 ITERATIVE EVALUATION

Having outlined stages in the process of creation of a knowledge base, this section discusses iterative evaluation of the knowledge base during its development. The iterative evaluation of the growing knowledge base to drive further elicitation is distinct from the evaluation of the representativeness of the knowledge base considered to be complete in relation to the objectives for its creation (§7.6) and the correctness and utility of its content. While these processes have tasks in common, their motivations and outputs are different.

Iterative evaluation of the knowledge base occurs throughout intermediate representation, formal representation and the development of glossaries and hierarchies. It is an inherently informal process and demands consideration of both individual statements of fact and sets of statement of fact as well as the specification of keywords and the relationships between keywords.

So, for example, the intermediate statement :

```
Orange caterpillars can cause sickness in livestock
```

(taken from knowledge collected by Garde, 1992) is incomplete and probably contains implicit meaning. Reference to the knowledge source shows that it can be more completely represented as :

```
Feeding orange caterpillars to livestock can cause sickness.
```

This still does not really capture the meaning of the source knowledge which is that :

```
Feeding leaves to livestock can cause sickness to those
livestock if there are orange caterpillars on the leaves.
```

While this is now an explicit representation of a piece of knowledge, the word 'can' along with a lack of any information about the circumstances under which the statement is held to be correct makes it of limited utility. This suggests a need for further knowledge elicitation, possibly revealing that this always occurs, i.e.

```
Feeding leaves to livestock causes sickness to those livestock
if there are orange caterpillars on the leaves.
```

This statement can now be meaningfully formalised. Leaves, livestock and orange caterpillars are objects. Feeding to is a process. Sickness might be viewed as a process or to be a change (decrease) in the value of an attribute (health). This statement might be formalised in many different ways, all capturing essentially the same meaning but with different emphases, for example :

```
process(leaves, 'fed to', livestock) causes att_value(livestock,
health, decrease) if att_value('orange caterpillars', location,
'on leaves').
```

or

```
process(leaves, 'fed to', livestock) causes att_value(livestock,
health, decrease) if att_value(caterpillars,location,'on
leaves') and att_value(caterpillars, colour, orange).
```

In other contexts a more precise definition of the results of consuming orange caterpillars may be necessary - sickness may be considered to be a particular type of decline in health that has particular consequences, the difference between the sickness caused by orange caterpillars and other sicknesses might be significant. All these factors will affect both the intermediate and formal representation of the knowledge and depend on the consideration of this statement of fact with other statements of fact in the knowledge base.

Consideration in the context of the other statements of fact in the knowledge base may reveal repetition, contradiction, incomplete sets of statements of fact, or inconsistent use of terms. In this example, there are statements that :

```
Green caterpillars cause sickness in livestock

Black caterpillars cause sickness in livestock
```

There is also information in the object hierarchies that there are three types of caterpillars; orange, black and green. If it is assumed that this is the finite set of caterpillars in this context then the three statements can be replaced with the single statement that :

```
Feeding leaves to livestock causes sickness if there are
caterpillars on the leaves.
```

If it is suspected or found that there are other types of caterpillars that may not, or do not, cause sickness if fed to livestock, the three statements may be replaced by the single statement :

```
Feeding leaves to livestock causes sickness if (there are
caterpillars on the leaves) and (those caterpillars are green,
or those caterpillars are black or those caterpillars are
orange).
```

Alternatively, green, orange and black caterpillars might be classified as being poisonous caterpillars such that :

```
Feeding leaves to livestock causes sickness if there are
poisonous caterpillars on the leaves.
```

Consideration of this statement of fact in relation to other statements may also demand consideration of linkages - for example are there statements which describe the consequences of a decline in the health of livestock? Apparently contradictory statements may be identified and resolved. For example, does the statement that :

```
Leaves of fodder trees that are attacked by caterpillars do not
cause sickness in livestock
```

mean that caterpillars do not cause sickness or does it mean that the leaves of those trees that are attacked by caterpillars do not cause sickness, provided that there are no caterpillars on them?

At this juncture, consistency of use of terms can be ensured. The term livestock is, in the current context, defined in the object hierarchy as cows, goats and buffalo. Do all the statements that use the term livestock refer to all three of these, or should some be replaced, for example, with 'cattle' or 'milking livestock' (i.e., cows, goats or buffalo that are currently lactating)? Precise use must be considered in terms of precise definition. The term 'fed to' is used in this statement. It may be that this term is defined as meaning the process of providing fodder for livestock that are stalled. This might be distinguished from the process of 'browsing' where livestock actively seek out and select fodder on trees. This process of selective browsing may mean that fodder effects relating to stall-fed livestock do not occur for free-range livestock.

This example illustrates something of the range of activities that can be involved in iterative evaluation and improvement of the knowledge base. The following sections anatomise some of the key processes occurring in iterative evaluation.

7.4.1 Evaluation of individual statements of fact

Individual statements of fact may be evaluated in terms of validity of representation, relevance, utility and ambiguity.

(i) Validity of representation

The restricted structure of the formal grammar, use of a parser[7] and the stylised English sense check in formal representation in AKT2 (Appendix B) and the restricted structure of the diagramming interface and English sense check (§6.5.3) provide some evaluation of the validity of representation of formal statements of fact and intermediate statements of fact created through diagrammatic representation. By contrast, no such checks are applied to intermediate statements of fact entered through the text

[7] A 'parser' is code that disassembles a statement into its component parts and checks the relationships between those parts, thereby ascertaining whether that statement conforms to a particular grammar.

interface (§6.5.4). Such statements frequently do not represent valid representations, not because they do not make sense but because they do not conform to the definition of a statement of fact (§3.2.1). This is frequently because a proportion of the statement of fact is itself a legitimate statement of fact. For example, the statement :

```
Nutrient uptake by fodder trees planted on crop land causes a
decrease in soil fertility.
```

can be broken down into two statements :

```
Planting fodder trees causes an increase in soil nutrient uptake

An increase in soil nutrient uptake causes a decrease in soil
fertility.
```

Statements which do conform to the fundamental definition of a statement of fact may often be broken down through the use of condition. For example :

```
Thorny leaves are unpalatable.
```

may be stated as :

```
Leaves are unpalatable if they are thorny.
```

This is important because of its subsequent impact on formal representation. Formal representation of the former would demand that 'thorny leaves' be treated as an object. This is less flexible than a formal statement in which leaves are the object and their palatability and thorniness are two attributes of that object.

Whether or not the intermediate statement is actually edited in the knowledge base to correspond to the intermediate statement that is effectively formalised is a matter of judgement.

(ii) Relevance and utility

Evaluation of articulated knowledge against objectives provides an important means of maintaining the quality (fitness for purpose) of the knowledge base. Effective filtering of articulated knowledge irrelevant to the knowledge base during knowledge elicitation and intermediate representation greatly reduces the need for subsequent rationalisation. Statements of fact may be relevant, in as far as they are concerned with the domain under consideration, but irrelevant in that they are not useful (in the context of the knowledge base as a whole). So, for example, the statements :

```
Manure production influences crop production
```

and

```
Some species of tree fodder have a beneficial effect on milk
production
```

were deleted during the reimplementation of the knowledge base created by Garde (§7.7.1) because both relationships were expanded in much greater detail in other parts of the knowledge base.

(iii) Ambiguity

Unambiguous (precise and complete) articulation and representation of a statement of fact is required if a useful knowledge base is to be created. Experience has shown that the majority of ambiguity encountered in a knowledge base is resolvable, being an artefact of articulation and representation. However, knowledge may be inherently ambiguous, so some intrinsic ambiguity will remain in any knowledge base.

The level of resolvable ambiguity in the knowledge base has been found to be influenced by :

- the nature of the knowledge source;
- the nature of the knowledge elicitation process;
- the type of interface used for entering the knowledge;
- the experience of the knowledge base developer; and
- the clarity of objectives for creating the knowledge base.

Resolvable ambiguity is most frequently exposed during the processes of formal representation and the creation of glossaries and hierarchies, both demanding a precise use of terms. Where ambiguity is the result of inadequate intermediate representation it may be resolved through reference to interview material. However, where ambiguity is a result of incomplete or muddled articulation by the knowledge source, further knowledge elicitation is demanded.

Unambiguous representation depends on the complete representation of the meaning of a statement of fact, complete representation of the context of application of that statement of fact and precise use of terms within that statement of fact.

Complete specification of meaning

Statements of fact may frequently contain an implied meaning. Even where it can be assumed that this is understood by potential users of the knowledge base, it may significantly constrain automated reasoning.

For instance, while it might be valid to assume that users understand that the statement, in a knowledge base about tree fodder,

```
Celtis australis leaves can cause sickness
```

means that feeding the leaves to livestock may cause sickness in livestock, automated reasoning will not pick up this implication. Even simple reasoning such as answering the question :

```
What is the consequence of feeding Celtis australis leaves to
goats?
```

can not, therefore, be automated. As a result, formal representation must explicitly state implied meaning, in this case maybe as :

```
process(part('Celtis australis', leaves), 'fed to', livestock)
causes att_value(livestock, health, decrease)
```

Context

Although the approach to knowledge base creation developed in this thesis is based on the disaggregation of sets of statements of fact, knowledge is inescapably contextual at some level. Recording conditional information associated with a statement of fact provides contextual information about the application of that statement of fact. However, it is impractical to exhaustively record the context of application: some level of understanding from common knowledge or the evaluation of the knowledge base as whole must be assumed. As a result, judgement of the completeness of conditional information is subjective and dependent upon context.

It is important that where a statement is considered to be unconditional this is explicitly recorded.

Terminology

A final source of ambiguity is the imprecise use of terminology. The definition of terms and the consistent use of terms are discussed latter (§7.3.4). Here the concern is with the inaccurate use of terms within a statement. A widespread example is the use of the term 'shade', which occurs in all the knowledge bases created to date. It has frequently become apparent that a statement of fact refers to one aspect of shade (for example a change in light levels, light quality or temperature) rather than a composite of all shade effects.

Intrinsic ambiguity

In a fundamental sense, virtually all knowledge is ambiguous at some level. In the current context, however, intrinsic ambiguity is of interest where it has an impact on the practical utility of knowledge.

Ambiguity is intrinsic where the source of an ambiguous statement cannot resolve the ambiguity in that statement. This is generally because the ambiguity is not clear to the informant, frequently because the issue in question is at the margins of his or her comprehension. Frequent examples have been encountered in relation to the meaning of terms within a statement. Clearly, distinguishing intrinsic ambiguity from the inability of the knowledge base developer to understand the concepts under consideration is a matter of judgement. For example, the precise meaning of the term 'tapkan' in relation to tree-crop interactions in Nepal remains unclear (Thapa, in preparation). Whether it is an intrinsically ambiguous (if very useful) term or a more precise definition exists and is applied but has yet to be elicited has yet to be resolved.

7.4.2 Sets of statements of fact

The content of a knowledge base is greater than the sum of the content of the individual statements of fact. As a result, iterative evaluation must include consideration of sets of statements of fact as well as individual statements.

Sets of statement of fact may be evaluated in terms of repetition, contradiction, completeness and consistency in use of terms.

(i) Repetition

A compact knowledge base is significantly more tractable, and, therefore, more useful, than a less compact version. Compaction is achieved by identifying and removing repetition in the knowledge base.

Two types of repetition can be identified, strict repetition and deducible repetition. Strict repetition is where a piece of information is stated more than once. Deducible repetition is where a statement in the knowledge base can be deduced from other statements in the same knowledge base and is therefore superfluous.

Strict repetition

Where repetition is identified, repeated statements of fact can be merged into one and tagged according to all the sources and occasions from whom and during which the statement was elicited. While some repetition may be self evident, the identification of repetition will frequently only be achieved by reference to the objectives of the use of the knowledge base. So, two different statements of fact using different terms but capturing very similar information may or may not be defined as repetition depending on the potential impact of the different formulations on the use of the knowledge base.

Alphabetical listings of the statements of fact in a knowledge base have revealed a surprising proportion of the repetition. The use of selection facilities to identify all statements concerned with a particular keyword or combination of keywords has revealed further repetition. Comparisons of sets of formal statements have revealed further repetition as the formal representation of statements often causes apparently distinct statements to converge, particularly where care has been taken to minimise the use of equivalent terms.

Deducible repetition

Mechanisms for compacting the knowledge base through the identification of deducible information in relation to inheritance down an object hierarchy have already been discussed (§3.5.1). Ensuring that knowledge is recorded at the 'highest' level can greatly reduce the size of a knowledge base.

Deducible repetition can also occur where the implications of linked sets of statements of fact are explicitly stated in the knowledge base. Where a statement contains the statements that :

```
Fertiliser application increases soil fertility if ...

An increase in soil fertility causes an increase in crop yield
if ...
```

The statement that :

```
Fertiliser application increases crop yield.
```

is a deducible repetition.

Similarly it can be logically deduced from the second of the above statements that :

```
A decrease in soil fertility causes a decrease in crop yield if
...
```

Systematic identification and removal of repetition in a knowledge base may often halve the number of statements of fact in the knowledge base. Balaram Thapa (in preparation) was able to reduce a set of over 2000 statements of fact to 700. The majority of this compaction was the result of identifying and removing repetition, particularly deducible repetition.

(ii) Contradiction

Contradictory statements of fact may be of two types - conflicting and inconsistent. Conflicting knowledge is knowledge from two different sources which is contradictory. Inconsistent knowledge is knowledge from a single source which is contradictory. Until contradiction is resolved, either by rejecting one statement of fact in favour of another or by demonstrating that apparent contradiction does not represent actual contradiction, the two contradictory statements of fact can be viewed as being competitive[8].

Where contradictory knowledge in the knowledge base is identified, it may be resolved through further knowledge elicitation or flagged within the knowledge base. Contradictory statements of fact may become apparent at any stage during the creation of the knowledge base. Equally, contradictions may only become apparent once the implications of the knowledge have been illuminated through reasoning with the knowledge.

Once contradiction has been identified it may be resolved by one of two means.

- Apparent contradiction may be resolved through clarification of the meaning of contradictory statements; in particular, two apparently contradictory statements may be distinguished by specification of the conditions under which they are held to be true,

- Contradiction may be resolved by demonstrating that there is a significant difference between the reliability of two statements such that the less reliable statement can safely be rejected in favour of the more reliable.

Inconsistent knowledge should, in principle, be resolved by one of the above mechanisms. Conflicting statements may frequently not be resolved: under such circumstances the two statements of fact may be flagged as being alternative views.

(iii) Completeness

The incompleteness of a set of statements of fact (as opposed to an individual statement, §7.4.1) is assessed through the identification of apparent gaps in the knowledge in relation to objectives. Gaps may occur in a knowledge base because the knowledge needed to fill those gaps is unknown or because the relevant knowledge has not been elicited. Therefore, the identification of gaps demands further knowledge elicitation and, if the necessary knowledge is found to be available, addition to the knowledge base. Completeness can only be defined in relation to objectives in the creation of the knowledge base. Even then, iterative evaluation of completeness tends to be subjective.

Generating diagrams (Chapter 5) during intermediate representation provides a powerful means of facilitating the identification of gaps in sets of knowledge. Figure 7.2 shows a diagram constructed as the result of one of a series of interviews with farmers in Nepal (Thapa and Walker, 1992). Consideration of this diagram provided a set of questions for further interviews (for example, 'what are

---

[8]Knowledge may be competitive without being contradictory. Two predictive theories may predict a phenomenon on the basis of different variables. For example, two explanatory theories may explain a phenomenon by reference to different, but not mutually exclusive, causal mechanisms.

the stem strength properties of the different crop species grown and the different varieties of each species?'; 'what influence does crop head size have on crop yield?'; 'what are the consequences of an increase in straw height?'; and 'does an increase in shade always result in an increase in pest incidence?') and shows how this diagram helped to identify topics for further discussion with the informant.

Similarly the use of automatic reasoning techniques such as the navigation tool (Appendix A) can facilitate the identification of gaps in deducible chains of relationships.

(iv) Consistency and precision of the use of terms

Comparison of the use of the same term in different statements to ensure that this term has a consistent meaning and the comparison of apparently similar terms to identify overlaps of terminology help the knowledge base developer to move towards the consistent and precise use of terminology required for effective use of the knowledge base.

Terminology in knowledge systems about agroforestry is frequently not consistently or precisely used. A demand for an exacting consistency in use of terms in creating a knowledge base may lead to a proliferation of terms and become increasingly unrepresentative of the actual use of those terms by source communities. Here again the strategy to be taken depends on objectives in the creation of the knowledge base, whether, for example, they are primarily to study the current state of knowledge of a target community or to develop a knowledge base for use in providing decision support.

7.4.3 Evaluating keyword structures

As well as evaluating and modifying statements of fact, iterative evaluation of the knowledge base demands evaluation of the sets of keywords specified, the relationships between those keywords and the definitions of those keywords.

(i) Sets of keywords

As with identifying and removing repetitious statements of fact, identifying and removing repetitious terminology can significantly increase the utility of the knowledge base. The glossary features available in AKT help to encourage consistent use of terminology, particularly through formal representation. Nevertheless, regular comparison of the various keyword terms may frequently reveal overlapping terminology. The ability to identify equivalent terms for the same object provides a useful means of allowing a natural articulation of knowledge while still identifying equivalent meaning and could be usefully extended to processes, attributes and values.

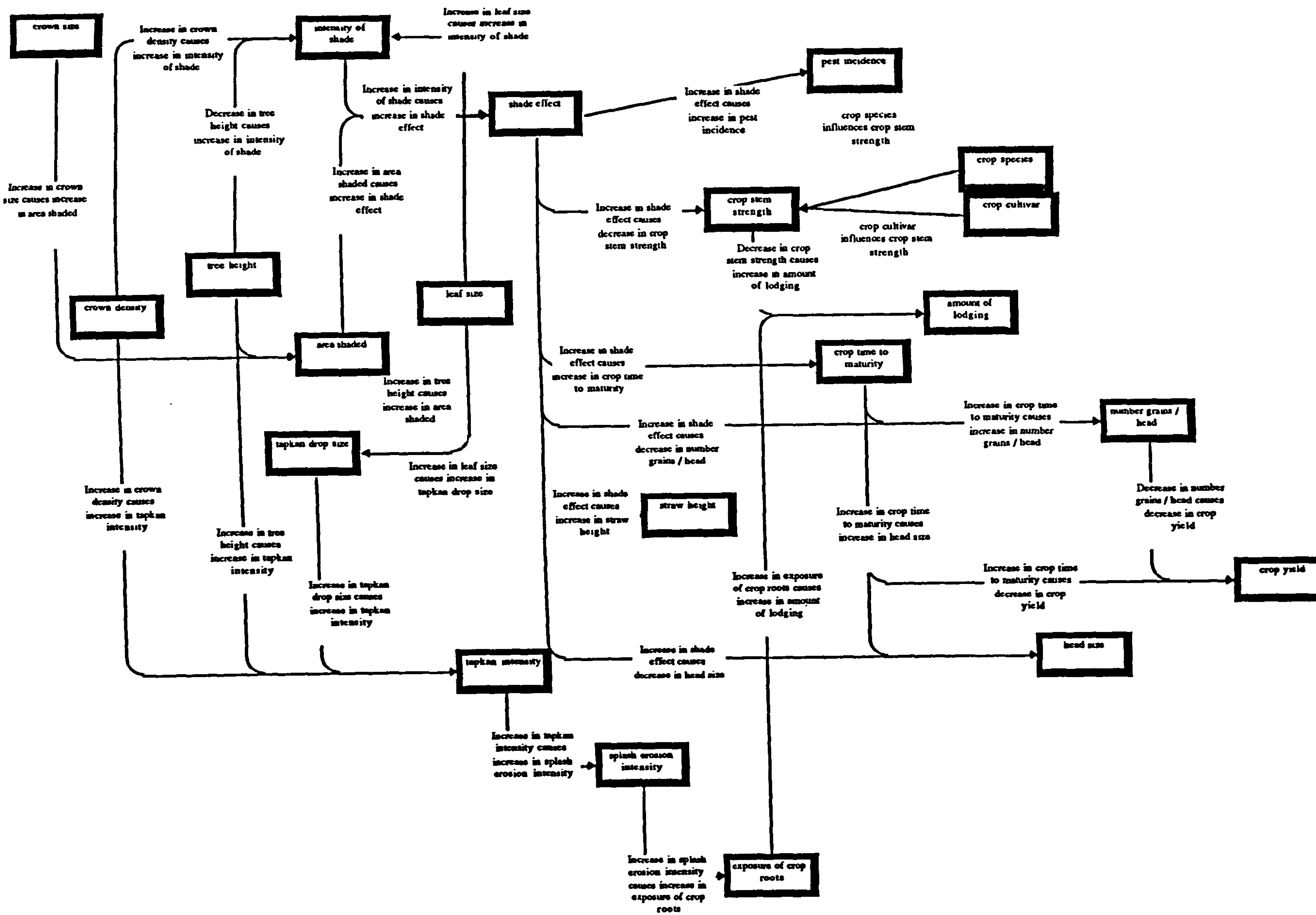


Figure 7.2 A diagram constructed from knowledge elicited through interview as a means of identifying questions for further interviews.

(ii) Relationships between keywords

The development of object hierarchies has already been identified as one of the most demanding tasks in the creation of the knowledge base (§7.3.4). Regular evaluation of keyword hierarchies, including testing those hierarchies on source communities has proved to be valuable. Southern (in preparation) has demonstrated the utility of iteratively testing keyword hierarchies on individual informants and groups of informants in creating an unambiguous and representative soils taxonomy.

Equally, assessment of the set of attributes that are considered in relation to a particular object or process and the set of values that are linked to an attribute is important in ensuring consistent terminology.

In the Indian tree fodder knowledge base (§7.7.1), for example, two separate objects, 'leaves' and 'manure' have an attribute 'texture'. However, the set of values that this attribute can take differs for the two objects ('soft' and 'hard' for leaves, 'loose' and 'firm' for manure). As a result one or both of the attributes must be renamed to avoid this apparent repetition. By contrast, both 'leaves' and 'manure' also have the attribute 'water content', but the values that this attribute can take ('high' or 'low') are the same for both.

(iii) Definitions of keywords

The definition of a keyword will frequently be iteratively improved over the course of the creation of a knowledge base. After each refinement of the definition of a term, all existing uses of that term in statements of fact must be checked to ensure a valid use of terminology.

7.5 CREATING KNOWLEDGE BASES THROUGH THE COMBINATION OF EXISTING KNOWLEDGE BASES

In §7.2 to §7.4, the creation of knowledge bases from scratch has been described. The AKT software was designed and has been used for this task. However, it is also possible, and may frequently be desirable, to create a knowledge base through the combination of two or more existing knowledge bases (on the same domain but from different sets of sources for example).

In principle the combination of two knowledge bases is a straightforward task. First the object hierarchies, attribute, process, link and part glossaries for each of the parent knowledge bases are printed. Then the sets of statements of fact are brought together into a single file. The hierarchies are now rationalised by identified keywords in the two different sets that are equivalent and either identifying this equivalence in the hierarchies or glossaries or changing one of the terms to the other wherever it occurs. The process of merging knowledge bases is, therefore, essentially one of ensuring consistency of terminology.

7.6 EVALUATION OF THE REPRESENTATIVENESS OF A KNOWLEDGE BASE

Most uses that can be envisaged for the content of a knowledge base will be predicated on the assumption that it is representative of the knowledge of a defined community, or that definable parts of the knowledge base are representative of the knowledge of definable communities.

The four-stage knowledge acquisition strategy outlined in §7.2.3 provides a means for eliciting knowledge by relying, in the first instance on particularly knowledgeable, experienced and co-operative informants (or key texts). The resulting knowledge bases will normally be intended to be representative of the source community(s). Any use of the knowledge base must be informed by an evaluation of the representativeness of the knowledge base. The detailed design of mechanisms for evaluating representativeness is research specific (see e.g. Thapa, in preparation; Jinadasa, in preparation). This section highlights some of the issues in this process.

7.6.1 Definitions of representativeness

Definitions of, and requirements for, 'representativeness' depend on the objectives for the creation and use of a knowledge base. A knowledge base may be evaluated in terms of, for example :

- the extent to which the knowledge in the knowledge base represents a valid abstraction of the knowledge of the sources from which that knowledge was elicited; or
- the extent to which the knowledge elicited from a sample of a community and represented in a knowledge base is representative of the knowledge held by the community.

7.6.2 Validity of abstraction

The extent to which a knowledge base represents a valid abstraction of the knowledge held by the source(s) from which knowledge was elicited is a standard problem in the creation of expert systems (Hart, 1986; MacGraw, 1989). However, in these circumstances the primary concern is to identify where the knowledge in the knowledge base falls short of the knowledge held by the source on the domain in some significant way and to address this shortfall. A 'measure' of the representativeness of the knowledge base of the knowledge held by the source is irrelevant, as the functioning of the knowledge base in an expert systems context is evaluated in terms of its accuracy in decision support or expert replacement.

By contrast, in the current context, an objective measure of the extent to which the knowledge base represents a valid abstraction of the knowledge held by the sources used in its creation may be required if it is to be used, for example, in planning extension activities targeted at the group of informants. This measure is generated through the creation and use of questionnaires designed to test the validity of abstraction of 'key' (or randomly selected) statements of fact from the knowledge base.

7.6.3 The representativeness of the knowledge base of the knowledge held by the community

Applications of the AKT approach to date have all developed knowledge bases that are intended to be representative of the knowledge of a defined community. These knowledge bases have been created through knowledge elicitation from a sample of the members of the community. As a result, the representativeness of the knowledge base of the broader community demands further knowledge elicitation comparing the content of the knowledge base with the knowledge held by members of the community not originally interviewed.

The extent to which a knowledge base that is assumed or shown to be a valid abstraction of the knowledge of the sources of that knowledge is also representative of the knowledge of a wider community will depend on :

- the heterogeneity of the knowledge held by members of the community on the domain in question; and
- the impact of the sampling bias of purposive selection of key informants.

(i) The heterogeneity of knowledge

The heterogeneity of knowledge within a community is clearly domain specific. There is some reason to believe that ecological knowledge about agroforestry may be fairly evenly distributed within farming communities. Knight (1980) states that in a relatively close-knit farming community, such as an African village, basic agricultural-type knowledge will be universal and uniform. This is corroborated by experience in knowledge acquisition in association with the current research to date. Most applications to date that have undertaken preliminary investigations of the distribution of knowledge have found little evidence for significant variability between sources (Southern, Kilahama, Garde, Thapa, pers. comm.). Work in Thailand has suggested some variation in relation to generation (Preechapanya, in preparation).

Other work, however, casts doubt on the uniformity of distribution of ecological knowledge. Johnson's work in the north-east of Brazil (Johnson, 1972) found considerable variation between smallholders' knowledge, perception and circumstance related to ecological variability, variability in the capability of individuals and disagreement between individuals over facts and their interpretation, such that he contended that each family unit had to be treated as unique.

This apparent contradiction may, in fact, reflect different degrees of detail being sought in the elicited knowledge. Werner and Schoepfle (1987a) assert that, for homogeneous communities :

> (1) shared knowledge rarely exceeds 60%, (2) unique knowledge is rarely less than 30%, and (3) knowledge shared between any two members of the group (beyond what is shared by all members) rarely exceeds 5% (the remaining percentages consist of items of knowledge that are difficult to assign).

It seems reasonable to suppose that shared knowledge represents basic terminologies and theories while other knowledge is more detailed and less generic. On this basis, it is proposed that the knowledge acquisition strategy outlined here provides a sound basis for efficiently eliciting the basic, orthodox and fundamental knowledge held by a community about particular topics relating to agroforestry. Alternative strategies might provide more efficient means of eliciting and evaluating knowledge that is unevenly distributed, conflicting or unorthodox.

Clearly it may not be immediately apparent which knowledge belongs to which category. However, on the basis of the figures cited above it has been suggested that if knowledge is shared between three or more people, it is probably shared by all the members of a community (d'Andrade, 1970). Corroboration and extension of measures of this sort could provide an effective means of identifying the most appropriate knowledge elicitation strategy for a particular topic in relation to a particular target community.

(ii) Sampling bias

The specification stage of the knowledge acquisition strategy outlined (§7.2.3) has a profound impact on the content and 'shape' of the knowledge base. This process is based on the purposive selection of key informants. There are clear advantages to working with articulate, knowledgeable and willing key informants during specification. It has commonly been found that key informants particularly enjoy attention and welcome an opportunity to think about, and possibility improve, their understanding of the domain. This makes the process of knowledge acquisition significantly more productive.

If, however, these key informants are experts in relation to the domain in question and, therefore, their knowledge is in some fundamental sense not representative of the source community as a whole, relying heavily on key informants during the early stages of knowledge creation may significantly reduce the representativeness of the resulting knowledge base.

Much has been made in expert systems literature of the specific nature of the knowledge of a domain expert (someone who has skills that have been developed and built as a result of complex experience and are not based on mere repetition). Experts are often characterised as having idiosyncratic ways of working and a reputation as an authority amongst contemporaries (Bell and Hardiman, 1989). Furthermore, it is noted that "the more competent domain experts become, the less able they are to describe the knowledge they use to solve problems " (Waterman, 1986b). It remains open to question whether key informants in local communities, are, in this sense, experts on the particular aspects of agroforestry under consideration.

It is accepted that there is specialist knowledge in almost any community, folk-medicinal knowledge being a commonly cited example. However, there is no evidence to suggest that explanatory ecological knowledge about production practices falls into this type of category. Experience in knowledge elicitation to date has not suggested any significant difference in the knowledge articulated by key informants as opposed to other members of a source community. Until such time as evidence to suggest otherwise is encountered, it seems reasonable to assume that the knowledge held by all members of the community is fundamentally the same in nature.

7.7 CASE STUDY ASSESSMENT

7.7.1 Introduction

The procedure for knowledge base creation described in §7.2 and §7.4 is a synthesis of experience in trial knowledge base creation and in providing support for the application of the approach. Garde (1992), da Costa(1993) and Holmes (1993) have evaluated the content of the knowledge bases that they created, and indicated a clear utility resulting from the explicit representation of knowledge. These knowledge bases did not, however, include formal representations. Thapa (in preparation) has completed a formal knowledge base resulting from extensive fieldwork and is undertaking an assessment of the representativeness of the knowledge base before evaluating its content and utility. Jinadasa (in preparation), Preechapanya (in preparation), Kilahama (in preparation) and Hitanyake (in preparation) are creating and evaluating formal knowledge bases.

In this section a case study assessment of the utility of the AKT approach, with a particular emphasis on the role of the software as an environment for creating and evaluating knowledge bases, is described. This assessment involved the reimplementation of a paper knowledge base created by Garde (1992) (the Indian tree fodder knowledge base), merging this knowledge base with another (the Nepali tree fodder knowledge base, created by Thapa, in preparation) and enhancing the resulting knowledge base (the combined tree fodder knowledge base) with further information from the literature.

(i) The Indian tree fodder knowledge base

The Indian tree fodder knowledge base was developed by reimplementation of a knowledge base developed by Garde (1992). The original knowledge base was created, under the supervision of the author, after six months fieldwork by Garde in Uttar Pradesh. It was developed without the use of a computer. Instead a set of diagrams conforming to the specifications outlined in Chapter 5 and notes associated with each link in the diagrams were recorded on paper. This knowledge base consisted of 41 hierarchically structured diagrams showing a total of 224 unique links (i.e. ignoring repetitions in different diagrams). Some of these links represented more than one relationship between two nodes, hence text information on a total of 261 relationships was recorded.

Reimplementation of this knowledge base in AKT provided a means of evaluating the AKT2 v. 2.11 software. Diagrammatic representation had enabled Garde to produce a succinct and explicit record of knowledge on the domain. However, representation on paper allowed the creation of diagrams that did not strictly conform to the diagramming convention implemented in the diagramming interface in AKT (§5.4). In particular, the text associated with relationships in the diagrams often greatly exceeded a single statement of fact. As a result, the diagrams could not be directly entered into the AKT software. Intermediate representation of the content of Garde's knowledge base in AKT was, therefore, achieved by identification and textual representation of all the statements of fact represented in the diagrams. The resulting knowledge base (the Indian tree fodder knowledge base) consisted of 306 intermediate statements of fact of which 280 were formalised. It contained seven object hierarchies containing a total of 73 object keywords. There were 75 attribute terms, five link terms, three part terms and 18 process terms.

(ii) Combination with the Nepali tree fodder knowledge base

The Nepali tree fodder knowledge base was created by Balaram Thapa, with support from the author. Development was undertaken first in TEAK (Appendix A) and then in AKT. An evaluation of the utility of AKT in the creation of this knowledge base is being undertaken by Thapa (in preparation).

The combination of this knowledge base with the Indian tree fodder knowledge base provided a further evaluation of the AKT software and approach. Both knowledge bases were concerned with the selection and use of tree fodders in the Himalayas. They were developed independently by different researchers in different localities but showed a considerable overlap in content. For this reason, it was proposed that their combination into a single knowledge base would produce a more useful resource than the two knowledge bases individually. It is important to note, however, that the Nepali knowledge base was concerned with on-farm tree fodder while the Indian knowledge base was primarily concerned with common-property forest resources.

At the time of combination the Nepali tree fodder knowledge base contained 698 statements of fact of which 626 were formalised. These were derived from knowledge elicited from 40 informants, each interviewed four times. As a result the majority of statements are ascribed to a very large number of interviews, more than 120 in many instances. This significantly increased the size of the knowledge base, concomitantly slowing the manipulation of that knowledge base by the software. While important in the interpretation and use of the original knowledge base, this information was of limited relevance once incorporated into a combined knowledge base. For this reason, the source information associated with statements from the Nepali knowledge base was changed prior to combination such that each was referenced back to the original knowledge base. The knowledge base contained 16 object hierarchies containing a total of 460 nodes. Many nodes appeared several times in different hierarchies; there were 231 unique object terms, 155 attribute terms, 30 link terms, seven part terms, and 27 process terms are specified in glossaries.

Combination of the two knowledge bases(§7.7.5) resulted in a knowledge base of 1028 statements of fact with 200 attribute terms, 52 link terms, eight part terms and 43 process terms. The 209 object terms were divided over seven hierarchies

(iii) Additional material from the literature

After the creation of the combined tree fodder knowledge base, further knowledge was added from Panday (1982) and Singh (1982), two key texts on tree fodder resources in the Himalayas. Knowledge about the six predominant fodder species at each of the two sites at which the knowledge bases were created was abstracted from both texts.

The three predominant fodder species from higher altitudes (*Ficus neriifolia, Leucospectrum canum* and *Ficus auriculata*) and the three predominant species from lower altitudes (*Artocarpus lakoocha, Bauhinia purpurea* and *Rhus parviflora*) reported by Thapa (in preparation) for his study site were selected to complement the knowledge abstracted from the Nepal case study. The three predominant forest fodder species (*Quercus leucotricophora, Quercus semecarpifolia* and *Viburnum coriaceum*) and three predominant farm land fodder species (*Ficus nemoralis, Grewia optiva* and *Ficus roxburghii*) as reported by Garde (1992) were selected to complement the knowledge abstracted from the Indian case study.

A total of 116 statements of fact were added to the knowledge bases as a result. The information in Singh (1982) is presented on a species-by-species basis and is largely descriptive. The statements of fact abstracted reflect the nature of the source knowledge and are species specific. As a result, formal representation was primarily as attribute-value statements describing, for example, the altitudinal range, frost tolerance and light requirements of different species and the crude protein content of fodder derived from these species.

In the following sections, this case study is used as a basis for assessing the practicability of intermediate representation, formal representation, keyword specification, iterative evaluation and combination of knowledge bases with the AKT software.

Assessment of knowledge acquisition strategies and knowledge elicitation techniques is being undertaken elsewhere (Jinadasa, in preparation; Kilahama, in preparation; Thapa in preparation). However, it is proposed that the approach to the creation of knowledge bases about agroforestry developed in this thesis is not significantly sensitive to elicitation technique. The procedures for the design of a knowledge acquisition strategy and the knowledge elicitation techniques described in this thesis have been successfully applied in a range of contexts. However, it is contended that alternative acquisition strategies and elicitation techniques could be as effectively employed and that the approach as a whole is tolerant of different approaches to elicitation.

7.7.2 Intermediate representation

Intermediate representation using the text interface during reimplementation of the Indian tree fodder knowledge base was rapid and straightforward. The source knowledge was coherent and explicit, only requiring minor re-interpretation. In particular, previous diagrammatic representation meant that the connections between statements of fact were easily and effectively captured.

Intermediate representation of knowledge from the literature was rapid and straightforward. Because the knowledge represented was primarily species-specific information represented as attribute-value statements of fact, there was no demand to ensure the connectedness between statements of fact. However, two particular points merit further attention.

Firstly, resolving ambiguity in the interpretation of knowledge demanded a different approach to knowledge gained through interview. This is because it was impossible to gain clarification from the knowledge source (i.e. the author of the text). So, informed decisions had to be made as to whether, for example, the statement that *Quercus leucotricophora*

    ".. can grow on a wide variety of soils except very dry and
    shallow soils"

should be interpreted as meaning that the species will not grow on soil if it is very dry or if it is shallow or growth is only inhibited if the soil is both very dry and shallow.

Secondly, many statements, particularly in Singh (1982) were (even if not presented as such) statements of particular instance. For example :

    The leaves [of *Quercus semecarpifolia*] contain (on a dry matter
    basis) 10.20 - 11.42 per cent crude protein ....".

Clearly, what is meant is that values between, and including, these figures have been reported. This is difficult to represent in the formal knowledge base, which is designed for precise statements of belief. This statement was represented, unsatisfactorily, as :

```
Dry Quercus semecarpifolia leaves contain between 10.20 - 11.42
% crude protein.
```

If effective use is to be made of knowledge bases containing scientific knowledge, an ability to represent quantitive data in formal statements and make full use of those data (rather than, for example, treating them as being qualitative) are merited. In principle this is feasible, but would demand extension and modification of both the formal grammar and reasoning tools.

In other cases the interpretation of data was less problematic. For example the following data (extracted from a table of data about *Quercus leucotricphora* in Singh, 1982) :

Type of leaves	Moisture (%)
Tender	64.40
Medium	55.60
Mature	48.80

are ambiguous (tender, medium and mature leaves are not defined precisely enough to merit the accuracy in moisture content). However, for the purposes of this knowledge base, the relevant content can be summarised as :

```
"The moisture content of Quercus leucotricophora leaves declines
with age".
```

In summary, the process of intermediate representation both through the text and the diagramming interfaces was found to be intuitive and straightforward. The AKT software provided a robust and supportive environment for intermediate representation.

These observations are consistent with the experience of the researchers who have used the AKT software to create knowledge bases. Software development from the use of a word processing environment, through the combination of TEAK and HyperNet, on to AKT1 and then AKT2 (Appendix A) and the concomitant conceptual development and clarification, have made the process of intermediate representation increasingly straight forward. Development of, and familiarisation with, the formal grammar has made the process of intermediate representation less critical while also making the requirements for intermediate representation increasingly clear. The comparative merits of the two approaches to intermediate representation are also reflected in these applications. While Thapa (in preparation), Southern (in preparation) and Preechapanya (in preparation) have, to date, relied primarily on the text interface, Jinadasa (in preparation), Kilahama (in preparation) and da Costa (1993) have invested effort in relying, in the first instance, on creating diagrams, only using the text interface for knowledge that is not readily included in diagrams. The consensus encourages future development of the diagramming approach.

7.7.3 Formal representation

This section describes the problems encountered in the formal representation of the Indian tree fodder knowledge base in some detail. These examples are representative of the problems encountered in the formal representation of additional material from the literature and from interview.

Formal representation of two-thirds of the content of the intermediate Indian tree fodder knowledge base involved rapid and routine application of the guidelines for formal representation (Sinclair *et al.*, 1993). Many of these statements were simple attribute-value or causal statements. Formal representation of the remaining third demanded greater analysis and interpretation but was achieved without undue difficulty. As a result, 280 of the 306 intermediate statements of fact were formally represented in a matter of hours. However, successful formal representation of the residue of sentences could only be achieved as a result of the development of new techniques for applying the formal grammar or extension of the formal grammar or could not be formally represented without an unacceptable degree of reinterpretation.

The following are some examples of problems encountered with these problematic statements and, where applicable, the solutions applied.

(i) Comparisons of objects of the same type

Of the original 306 intermediate statements of fact 45 were statements of comparison. Of these, 38 involved comparison between the same object where an attribute of that object held a different value. The formal grammar and parser only allow the comparison of the value of an attribute for two different objects. In some instances, this problem could be overcome by creating different objects. So, for example, the statement:

```
Soil on cultivated land is moister than soil on forest land
```

could be formalised as

```
comparison('moisture', 'cultivated soil', greater_than, 'forest
soil').
```

Because forest soil and cultivated soil are identified as types of soil in the object hierarchy, no information is lost.

However, this approach can result in a proliferation of terms. For example, the formal representation of :

```
Younger leaves have a softer texture than older leaves
```

as :

```
comparison (softness, 'young leaves', greater_than, 'older
leaves')
```

is clumsy. One means of getting around this problem is to use variables and state the sorts of those variables :

```
(comparison (softness, X, greater_than, Y) if att_value(X, age,
young) and att_value(X, age, young) and att_value(Y,age,>young),
[X:leaves, Y: leaves]).
```

This is legitimate within the grammar as it stands, although it would require parser and interface modification. It is suggested, however, that the majority of comparisons of objects of the same sort can more effectively be treated as causal statements, in this case :

```
att_value(leaves, age, decrease) causes att_value (leaves,
softness, increase).
```

(ii) Referring to a change in an attribute of an object in a process

Many causal relationships involve a change in the value of an attribute of an object that is an integral part in a process. For example :

```
An increase in the greenness of leaves fed to livestock causes
an increase in the nutritiousness of the resulting milk
```

Here the attributes that are changing (greenness, nutritiousness) are of objects in processes rather than processes themselves. This statement can be formalised :

```
att_value(leaves, greenness, increase) causes
att_value(milk,'nutritional content', increase) if
att_value(process(leaves, 'fed to', livestock) occurrence,
occurs) and link(produce,livestock,milk)
```

It is, however, doubtful that reasoning tools will be developed that will cope with statements with this structure. This statement might be more effectively represented if the consequence of the relationship could be represented as a process but the attribute in that process statement could be an attribute of one of the objects rather than the process itself.

It is worth noting that the nature of the relationship between an object or a pair of objects and the named process with which they may be associated in a process statement is unclear.

(iii) Causal events as conditions

The current grammar does not allow the use of causal statements as conditions (on the basis of the risk of excessive stacking of statements). This can constrain formal representation. For example :

```
Caterpillars are not found on leaves that cause sickness
```

is most intuitively formally represented as :

```
not(link('found on',caterpillars,leaves)) if process(leaves,'fed
to',livestock) causes att_value(livestock,health,decrease)
```

This is not, however, accepted in the current implementation.

(iv) Causing a process

Similarly, the result of a causal statement is currently restricted to a change in the attribute of a value. In practice, processes are often caused by causal linkages. For example :

```
Pollination of a flower by a bee causes production of viable
seed by a flower
```

can only be currently represented as :

```
process(bee,pollinate,flower) causes att_value(process(flower,
produce,seed) occurence,change)
```

But might be better represented as :

```
process(bee,pollinate,flower) causes process(flower,
produce,seed)
```

or

```
process(bee,pollinate,flower) causes
att_value(process(flower,produce,seed) occurence,occur)
```

(v) Linking attributes

The statement :

```
Seeding rate varies with species
```

can be represented as :

```
link(influence,species,'seeding rate').
```

However, species and seeding rates are both attributes, but influence links are only permitted between objects or processes. In fact it seems logical that it will always be the attribute of an object or the attribute of a process that will influence an attribute of an other object or another process

(vi) Potential events

The statement :

```
Eating karsu leaves can cause sickness in livestock.
```

is an example of a statement of a potential event. However, only unequivocal statements can be represented with the formal grammar. Clearly it is unacceptable to interpret this statement as meaning that eating karsu leaves does cause sickness.

It is clear that the statement is of limited value. The effect is likely to be conditional, but without information on the conditions under which it applies its utility is constrained Further knowledge elicitation to establish conditional details (if available) is desirable.

If conditional information is not available it may still be useful to record the fact that the consumption of karsu leaves can cause sickness. This might be represented as :

```
Eating karsu leaves increases the danger of sickness in
livestock.
```

(vii) Unrepresentable statements

A few of the statements proved unrepresentable. These fell broadly into two categories, temporal statements and management type statements.

Temporal statements

Automated temporal reasoning is very demanding, so a temporal element has been specifically excluded from the grammar. Statements of the time of an event can be represented :

```
soil temperature is low during the dry season

att_value(system, temperature, low) if att_value (system,
season, 'dry season').
```

However, statements in which an event occurs over a period of time are more difficult. In some cases representation is possible :

```
the water content of tree fodders increases until the leaves are
mature
```

```
att_value(leaves, age, increase) causes att_value(leaves, 'water
content', increase) if att_value(leaves, age, immature)
```

but in others it is effectively impossible :

```
feeding fodder to livestock over a long period of time makes
them become accustomed to it.
```

Management statements

A conscious decision has been made to develop a structure for representing ecological rather than management-type statements about agroforestry. As a result statements such as :

```
Manure and bedding are mixed together to produce compost
```

cannot be formally represented.

In the context of the case studies undertaken this restriction has not been burdensome. However, the problems encountered in representing management statements have imposed significant constraints in other contexts. Jinadasa (in preparation) has developed knowledge bases on plant ideotype specification and mother tree selection by the farmers of the Kandy forest garden system. These knowledge bases contain a significant proportion of knowledge on the desirability of characteristics (e.g. large clove seeds are desirable) and management actions (e.g. mother plant selection is actively undertaken for clove). Straightforward extensions of the grammar would accommodate the representation of such statements. In spite of concerns about the impact of mixing ecological and management type-statements in the same knowledge base on the development of a coherent structure such an extension to the grammar would appear to be justified.

(viii) Stylised English and the parser

All the above examples of problems encountered in the formal representation of the Indian tree fodder knowledge base relate to the grammar and its application. Indeed, the software implementation proved robust. The parser provided effective support in generating syntactically correct statements, identifying and creating hierarchies and glossaries. Similarly, the stylised English statement generated through parsing helped in checking the sense of the formal representation, although this was not always helpful. For example the statement :

```
Dudhilla fodder does not fill the stomach rapidly
```

can be legitimately, if clumsily, formally represented as :

```
att_value(dudhilla,'ability to satisfy appetite',low)
```

but, the stylised English that results is, misleadingly,

```
the ability to satisfy appetite of dudhilla is low.
```

This section has illustrated some of the difficulties encountered in the formal representation of intermediate statements of fact. However, the emphasis should be on the satisfactory expressiveness of the formal grammar and the ease with which it can be applied. Modifications to the grammar and the development of techniques for the formal representation of problematic statements are to be expected, nevertheless the approach has been shown to be a practicable means of developing a formal knowledge base. The research students making use of the formal grammar have all been able to create formal representations of most of the statements of fact that they encountered within a few hours of introduction. As with any skill, the efficiency and efficacy of formal representation does, however, improve over a period of weeks as the implications of different decisions become clear.

Development of a parser for translating diagrammatic intermediate statements of fact into formal statements can be envisaged (§5.5). This would further reduce the time involved in formal representation. Greater reliance on the diagramming interface may provide a means of completely automating formal representation, thereby reducing the training needed by the knowledge base developer and the time involved in knowledge base development. However, this must be balanced with the reduced expressiveness of diagrammatic as compared with textual intermediate representation (§5.5.2).

While it is suggested that formal representation is a practicable process and provides a valuable discipline in synthesising knowledge and ensuring the explicit representation of knowledge, it has been stressed that the primary justification for formal representation is the use of the resulting knowledge base with reasoning tools. Throughout the process of formal representation the knowledge base developer is making decisions between alternative approaches in the light of ultimate objectives. It is also clear that where those ultimate objectives involve the use of reasoning tools, an understanding of the functioning and limitations of those tools will have an impact on decision making during formal representation.

7.7.4 Keyword specification

Keyword specification and management has been identified as a key process in the development of a coherent and tractable knowledge base (§7.3.4). With the increasing reliance on formal representation, and the resultant decline in the role of intermediate statements, keywords have become a facet of the formal rather than the intermediate knowledge base. The identification of attribute, part, link and process terms from formal statements and the compilation of the glossaries is automatic. Objects are also identified by the parser but the construction of the object sort hierarchies is left to the knowledge base developer. These facilities are straightforward and help to encourage the consistent use of terminology. However, while the mechanisms of keyword specification are straightforward, the process itself is demanding in at least four respects :

- the hierarchical structuring of object sort relationships;
- achieving an appropriate balance between a proliferation of terms and inconsistent use of terminology;
- providing a definition for keywords; and
- the consistent use of values for particular attributes.

The development of keyword sort hierarchies

Iterative development of a set of object hierarchies during the re-implementation of the Indian tree fodder knowledge base was not problematic. However, three observations were made as a result of the more demanding process of combining the object hierarchies of the Indian and Nepali knowledge bases (§7.7.5).

- While many objects, species names or soil types for instance, are naturally and intuitively classified into a hierarchical structure, many of the objects in the knowledge base bear no significant sort relationship to other objects.
- Information may be arbitrarily divided between hierarchies and statements of fact. For example, in the Nepali knowledge base, fodder tree species occur within a species hierarchy, a hierarchy identifying the livestock group to which each species is primarily fed, a hierarchy to classify species according to whether they are chiso or obano, a hierarchy to classify species according to whether they are rukho or malilo and a hierarchy to classify species according to whether they are posilo or kam posilo (Thapa, in preparation). The information in all except the first of these hierarchies could have been represented as statements of fact (see below) while other information about the species (crown size for example) could have been used as a basis for further object hierarchies.
- The information captured in the sort hierarchies is not easily explored. This is partly as a result of the use of a text- as opposed to diagram-based interface for representing object hierarchies. Additionally, because it is not represented as a statement of fact, information in the object hierarchy will not be recalled as a part of a keyword search. For example, the fact that *Quercus leucotricophora* is a type of oak and that Banj is another name for *Quercus leucotricopohora* will not be revealed by a keyword search on *Quercus leucotricophora* even though this information is stored in the sort hierarchy.

The last of these issues demands software modification. Statements of fact identifying 'type of' and equivalence relationships between terms could be automatically generated for inclusion in lists of statements created during a search, making this information of the same status as other knowledge in the knowledge base. This would further require knowledge in hierachies to be tagged according to source. Additionally, the upgrading of the hierarchy interface would be desirable.

Avoiding the time-consuming generation of an over-complex set of hierarchies demands careful management of the process of hierarchy creation. On the basis of experience in the creation of sort hierarchies for the combined knowledge base, the following guidelines are proposed.

- In the first instance all object terms should be appended to a flat object glossary. This could be automated, as in the generation of attribute, process, link and part glossaries.

- Hierarchies should be created only for those object terms where the hierarchies are important in representing source knowledge, for example in capturing a local soil taxonomy or in helping the management of the structure of the knowledge base in relation to reasoning (§3.5).

- Hierarchical 'type of' relationships should not be created where the knowledge could equally be represented as an attribute-value statement.

- Hierarchical 'type of ' relationships should not be created where the knowledge can be more completely represented as a link statement.

The last two of these guidelines merit further explanation.

There may be circumstances under which a case could be made for using some attribute-value relationships as a basis for an object hierarchy while leaving other attribute-value statements as statements of fact. It could be argued that the rukho-malilo, chiso-obano and posilo-kam posilo distinctions provide three important and practically applied classifications of trees by farmers while crown shape, for instance is not used as such. On this basis, representing the former in hierarchies but the latter as attribute - value statements for each species results in a more faithful representation of source knowledge than using either approach exclusively. However, inconsistent application of approaches to representation is likely to constrain the use of automatic reasoning tools. It is argued that consistency of representation is, in this instance, more important than a marginal improvement in the expressiveness of representation of the source knowledge. For this reason it is proposed that attribute-value-type statements (e.g. that 'bans is a type of rukho[9] fodder tree') should not be

[9]Ruhko is a Nepali term describing trees that depress soil fertility.

represented in hierarchies but should be captured as attribute-value statements, each species having an attribute (e.g. rukhoness) for which it has a particular value(e.g. rukho).

Much the same argument can be made where information in the hierarchies could be represented as link statements. For example, in the Nepali knowledge base, fodder species are classified in one hierarchy as being 'cattle buffalo fodder' or 'goat fodder', i.e. the species is fed to cattle and buffalo or to goats. This information could be more explicitly represented in a set of link statements such as :

```
Arkholo is fed to goats

link('fed to', arkholo, goats).
```

The rationale behind the creation of the original hierarchies was to enable statements such as :

```
Goat fodder is less posilo[10] than cattle/buffalo fodder
```

to be recorded and to ensure that, through reference to the sort hierarchy, it could be deduced that :

```
Arkholo is less posilo than badahar[11].
```

while maintaining the recognition of the fact (through reference to the sort hierarchy) that this applies to arkholo. However, where the information is recorded as a set of link statements exactly the same result can be achieved by including the rule that :

```
comparison(posiloness,X,less_than,Y) if link('fed to', X, goats)
and (link('fed to', Y, cattle) or link('fed to', Y, buffalo)).
```

It is argued that this is a more explicit representation of the set of relationships, reduces the complexity of the set of keyword hierarchies and reduces the proliferation of terms ('goat fodder, 'cattle/buffalo fodder').

These guidelines represent a rationalisation of the function of the hierarchies. The hierarchies were originally developed as a means of identifying relationship between keywords (irrespective of their status as object, attribute, process or value terms) in the intermediate knowledge base (§3.5) to facilitate flexible searches of that knowledge (§3.9), but now also capture sort hierarchy relationships between objects. Given that searches can now be conducted on the formal knowledge base using keywords and their relationships, changing the search mechanism for the intermediate knowledge base such that searches were not linked to specified keywords but any text in intermediate statements would be acceptable. This would simplify the role of object hierarchies, thereby facilitating rational development.

---

[10] Posilo is a Nepali term describing fodder quality, posilo fodders are highly nutritious.

[11] Identified in the sort hierarchy as a cattle/buffalo fodder

By applying these guidelines, the combination of a total of 23 hierarchies containing 533 object terms from the India and Nepal knowledge bases was reduced to seven hierarchies containing 209 object nodes in the combined knowledge base.

Consistent but restricted terminology

The efficacy of exploration of the content of both the intermediate knowledge base (i.e. key word searches) and the formal knowledge base (i.e. automated reasoning) depends to a large extent on the consistent and, therefore, restricted, use of terminology.

A proliferation of terms will mean that potential linkages in reasoning may be lost while keyword searches must be carefully managed to ensure that all pertinent statements are retrieved.

The knowledge base must, therefore, contain as few object, process, attribute and value terms as possible. These issues do not, however, apply to the natural articulation and interpretation of ecological knowledge. As a result, a tractable knowledge base is a simplified model of the understanding of the ecology of the system. The appropriate balance between the increased tractability and reduced representativeness resulting from simplification will depend on objectives. Even with clear objectives, achieving this balance proves demanding.

Providing definitions for keywords

It has been suggested that explicit definition of the meaning of keywords is desirable in producing a transparent representation of knowledge. Mechanisms to achieve this have been discussed (§3.4).

The original intention was that this process should capture the meaning of the words as defined by the knowledge source. It has become apparent, however, that the definitions provided by a source may not provide a useful definition for the user of the knowledge base[12] and, furthermore, that the process of synthesis and simplification in the creation of the knowledge base may mean that the terminology used in the knowledge base does not precisely equate to the terminology used by any individual informant. Nevertheless, retrospective specification of definitions for terms in the combined knowledge base highlighted the need for knowledge elicitation undertaken with the express intention of deriving definitions.

Creation of definitions by the knowledge base developer often reveals ambiguity and inconsistency in the use of a term but also often fails to result in a useful definition. This is particularly because the current interface for providing a definition (§6.5.5) provides no cues for structuring that definition.

---

[12] Definition of species, has, for example, often only been given by example. i.e. *What is a banana tree? That tree over there is a banana tree.*

The templates for statements of definition developed in §3.4 were applied on a trial basis for terms in the combined knowledge base by using these template structures to fill in the definition dialogue box in the current implementation. This was found to facilitate the process of generating explicit definitions. On this basis, it is suggested that replacing the current 'empty box' interface with an interface providing greater support in the construction of the definition (Figure 7.3 illustrates an appropriate implementation) would facilitate the specification of explicit definitions of keywords. Given that the statements of definition making up the definition have a known syntax and may, individually, repeat statements of fact in the knowledge base, the status of these statements and their potential role in automated reasoning require further consideration.

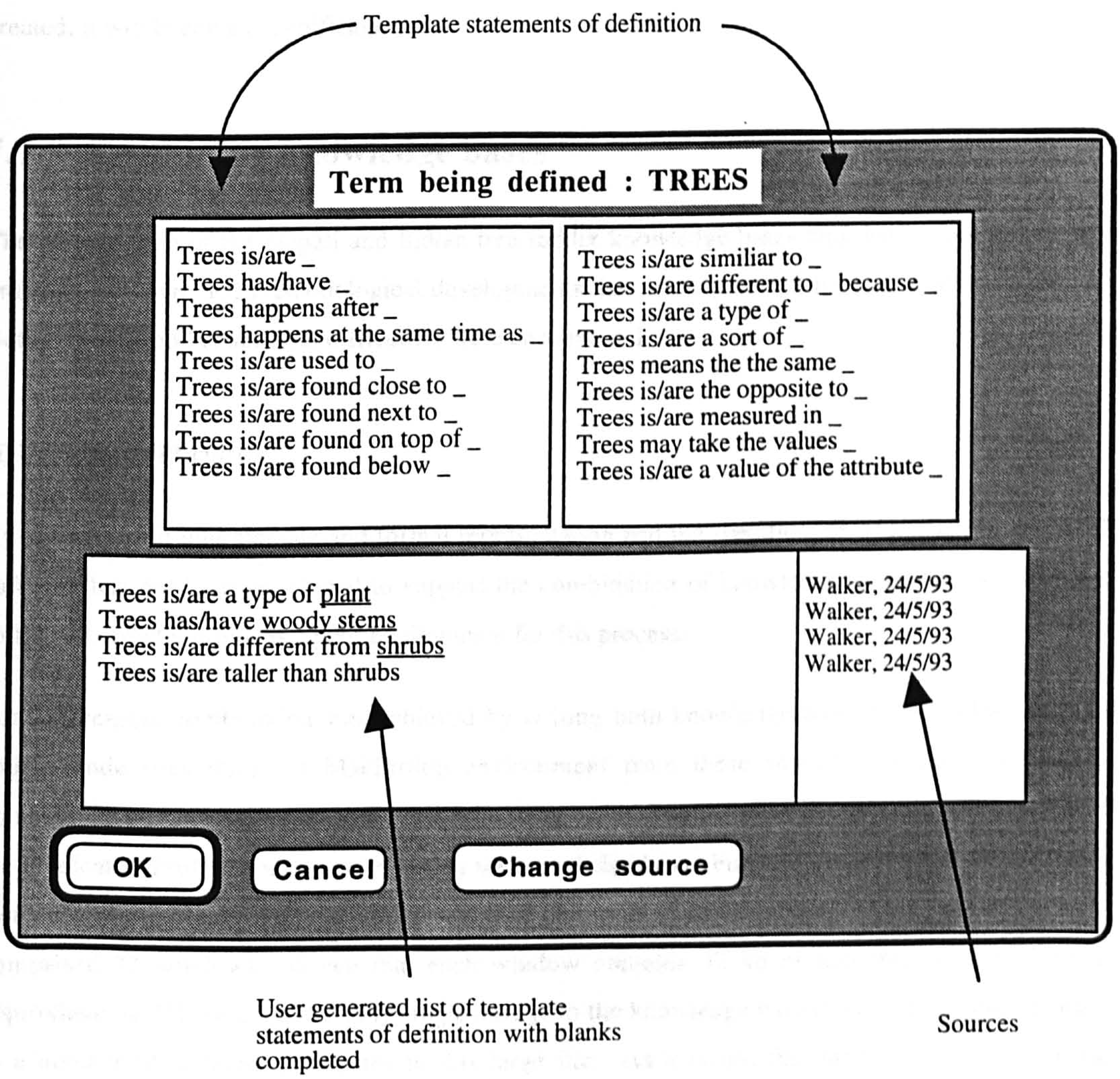


Figure 7.3 An interface for keyword definition. The user selects and completes appropriate template statements of definition to produce an aggregate definition. Each statement of definition may be ascribed to a different source or set of sources.

Consistent use of values for attributes

Consistent use of values for an attribute is not facilitated by the AKT software. There is no means of identifying the set of values that have been previously applied to an attribute or of specifying the relationship between those values (for example, whether they are cardinal, ordinal, represent a complete or incomplete set or range etc.).

Because the navigation tool (Appendix B) is designed to make use of causal statements in which values are restricted to a special set (increase, decrease, change, nochange) the consistent use of values currently has limited impact on automated reasoning. If, however, more sophisticated tools are created, it will become a significant issue.

7.7.5 Combining knowledge bases

The combination of the Nepali and Indian tree fodder knowledge bases was demanding and clearly indicated software and methodological developments that are required if this task is to be supported. Both technical and conceptual difficulties were encountered.

(i) Technical difficulties

By contrast with intermediate and formal representation and the specification of keywords, the AKT software has not been developed to support the combination of knowledge bases. As a result, the software currently provides a poor environment for this process.

For this reason, combination was achieved by writing both knowledge bases to text files, creating source code files within a MacProlog environment from these source files and undertaking combination within MacProlog. This demanded detailed understanding of the structure and function of the predicates (Prolog terms) used in saving the knowledge base. Furthermore, even with the facilities available within MacProlog, the editing was time consuming and demanding. At the outset the code comprised 72 windows. Given that each window contains 32 kb of text this is equivalent to approximately 360, 000 words. Any single change to the knowledge base usually demanded changes to a number of different predicates in this large file. As a result, the initial combination of the knowledge bases, in principle a straightforward task (§7.5) was several days' work. Once editing was completed, the source code was saved as text, the text file consulted from within MacProlog and then saved as object code. This object code could then be opened as an AKT knowledge base.

It is likely that any practical use of the AKT software would involve the occasional combination of knowledge bases. Software development to support this process is desirable. This software would :

- prompt the user to identify the two knowledge bases to be combined;
- change the sentence identification numbers[13] for all the sentences from one of the knowledge bases to ensure that each sentence retains a unique identification number;
- reset the current sentence identification number marker;
- change the source information on each statement such that each is sourced from its parent knowledge base (a facility to override this procedure would be desirable);
- merge the part, attribute, process and link glossaries, identifying any terms appearing in both knowledge bases; and
- disaggregate all object hierarchies into a single flat hierarchy with duplication of terms identified.

On completion of this procedure the knowledge base developer would have to sort through the glossaries to identify equivalent terms. A 'find and replace' editing facility allowing one of the terms to be deleted from intermediate statements, formal statements and glossaries would greatly facilitate subsequent rationalisation of the knowledge base.

(ii) Conceptual difficulties

Although effective software support for merging knowledge bases can be envisaged, it remained to be ascertained whether merging two separately created knowledge bases was valid. Two considerations had to be addressed.

In the first place, it was clear from consideration of the Nepali and Indian tree fodder case studies that two knowledge bases on the same domain are unlikely to use precisely the same terminology, especially where created by different knowledge base developers. In this instance, only 14 terms in the object hierarchies and glossaries were the same. Many equivalent terms (for example, alternative names for the same species) could be identified without difficulty, however, equivalence for other terms was much more problematic and, even where definitions of terms had been provided, could only be resolved through reference to the original knowledge base developer. As a result, it is to be doubted that consistent use of terminology across knowledge from different knowledge bases is practically possible unless strict protocols for the use of terminology in the creation of the knowledge bases in the first instance are developed and applied.

[13]Each statement of fact entered into an AKT knowledge base is tagged with an identification number, starting with one for the first statement.

A further problem was encountered because the content of the knowledge bases to be merged has been previously formally represented. It has been demonstrated that the process of formal representation involves many decisions about the most appropriate approach to representation. When this process has been undertaken by different individuals for different knowledge bases these decisions will not have been consistently applied across both sets of knowledge. The 'match' between formal statements from the Indian and Nepali knowledge bases was found to be poor. The implications of inconsistency in formal representation for automated reasoning are likely to be serious.

Upon completion, the combined knowledge base comprised 1394 statements of fact. At this point a major review of the content was deemed necessary in order to :

- identify and rationalise repetition, ambiguity and other inadequacies in the representation of the knowledge already elicited;
- provide a basis for planning further knowledge elicitation; and
- provide a basis for evaluating the utility of the knowledge base in its current form.

The next three sections describe some of the difficulties encountered in this process.

7.7.6 Evaluating attribute-value statements

Of the 1394 statements of fact in the knowledge base, 828 were attribute-value-type statements. These primarily related to the 128 different species about which knowledge is represented. In order to evaluate these species-specific data, the first requirement was to standardise species names. Knowledge from Singh(1982) and Panday(1982) was recorded according to binomials but knowledge derived from fieldwork by Garde(1992) and Thapa (in preparation) made use of a mixture of binomial and vernacular names. It was decided to standardise all names to binomials. This was achieved for the formal knowledge base by changing all species names in the sort hierarchy to binomials, while recording all vernacular terms as equivalents. The binomial name is, as a result, automatically substituted for any vernacular name previously used in a formal statement of fact. It was necessary to make the same changes for the intermediate knowledge base as this is primarily what the user sees in exploring the knowledge base. This was achieved by doing a search on each species name and manually changing any vernacular names within intermediate statements of fact to the appropriate binomial. The problems encountered in this rationalisation related to the knowledge rather than the software function. There was some disagreement between sources on the correspondence of vernacular and binomial names, in some cases because different species were, or were believed to be, associated with a vernacular name and in others because of the use of alternative binomial nomenclature. Species identification was not available for some vernacular names and, more particularly, appropriate binomial nomenclature was not available for the majority of locally identified sub-species or varieties.

Once names had been standardised, repetitions could be identified and removed by conducting searches on species names. Furthermore, apparently irrelevant statements of fact relating to particular species could be identified and deleted.

By using the query tool (Appendix B), the consistency with which attributes were recorded for different species was assessed. For example, the query :

```
att_value(O,'growth rate',V)
(i.e. What objects or processes are there in the knowledge base
for which a growth rate is recorded?)
```

reveals that information about tree growth rate is recorded in the knowledge base for 29 species (Table 7.1) and incidently shows an inconsistent use of values, with 'rapid', 'high', 'slow' and 'low' all being used. Information on growth rate is, therefore, not available for 99 out of the 128 tree fodder species in the knowledge base. It may be appropriate to elicit this information for these species.

However, eliciting further information will only be appropriate where that information is of relevance. It has been suggested that an AKT knowledge base is more than simply a database because the content of the knowledge base can be reasoned with. However, any attribute-value statements that do not occur elsewhere in the knowledge base and cannot, therefore, be used in reasoning, are simply data exactly equivalent to the content of a database. Given the present structure of the grammar, attribute-value statements can only be linked to other statements where they occur elsewhere as conditions. A search in the knowledge base reveals that there are no statements in the knowledge whose use is dependent on information about the growth rate of a species.

In summary, iterative evaluation of the species-specific attribute-value statements in the knowledge base proved to be both straightforward within the current software and effective in supporting rationalisation of the content of the knowledge base and the identification of gaps in the knowledge base that might merit further knowledge elicitation.

Table 7.1 Output from a search of the combined knowledge base using the Query tool

```
Query :  att_value(O, growth rate, V)

att_value('Quercus leucotricophora', 'growth rate', low)
att_value(seedling, 'growth rate', rapid)
att_value(process(hadchur, regrowth), 'growth rate', slow)
att_value(process(kabro, regrowth), 'growth rate', slow)
att_value(process(khayer, regrowth), 'growth rate', rapid)
att_value(process(bakaino, regrowth), 'growth rate', rapid)
att_value(process(chamlayo, regrowth), 'growth rate', rapid)
att_value(process(padori, regrowth), 'growth rate', rapid)
att_value(process(latikath, regrowth), 'growth rate', slow)
att_value(seedling, 'growth rate', moderate)
att_value(timla, 'growth rate', high)
att_value('Grewia opptiva', 'growth rate', high)
att_value('Celtis australis', 'growth rate', high)
att_value('Quercus glauca', 'growth rate', low)
att_value('Quercus semicarpifolia', 'growth rate', slow)
att_value('Quercus leucotricophora', 'growth rate', slow)
att_value(baiku, 'growth rate', high)
att_value('Pinus roxburghii', 'growth rate', 'very high')
att_value('Ficus nemoralis', 'growth rate', high)
att_value('Ficus nemoralis', 'growth rate', high)
att_value(process('Prunus cerasoides', regrowth), 'growth rate', rapid)
att_value(process('Ficus auriculata', regrowth), 'growth rate', rapid)
att_value(process(tanki, regrowth), 'growth rate', rapid)
att_value(process('Artocarpus lakoocha', regrowth), 'growth rate', rapid)
att_value(process('Bambusa sp.', regrowth), 'growth rate', slow)
att_value(process('Litsea polyantha', regrowth), 'growth rate', slow)
att_value(process('Saurauia napaulensis', regrowth), 'growth rate', slow)
att_value('Quercus semecarpifolia', 'growth rate', low)
att_value(process('Grewia tiliaefolia', regrowth), 'growth rate', rapid)
att_value(process('Brassaiopsis hainla', regrowth), 'growth rate', slow)
att_value(process(hadchur, regrowth), 'growth rate', slow)
att_value(process(kabro, regrowth), 'growth rate', slow)
att_value(process(khayer, regrowth), 'growth rate', rapid)
att_value(process(bakaino, regrowth), 'growth rate', rapid)
att_value(process(chamlayo, regrowth), 'growth rate', rapid)
att_value(process(padori, regrowth), 'growth rate', rapid)
att_value(process(latikath, regrowth), 'growth rate', slow)
att_value(process('Prunus cerasoides', regrowth), 'growth rate', rapid)
att_value(process('Ficus auriculata', regrowth), 'growth rate', rapid)
att_value(process(tanki, regrowth), 'growth rate', rapid)
att_value(process('Artocarpus lakoocha', regrowth), 'growth rate', rapid)
att_value(process('Bambusa sp.', regrowth), 'growth rate', slow)
att_value(process('Litsea polyantha', regrowth), 'growth rate', slow)
att_value(process('Saurauia napaulensis', regrowth), 'growth rate', slow)
att_value(process('Grewia tiliaefolia', regrowth), 'growth rate', rapid)
att_value(process('Brassaiopsis hainla', regrowth), 'growth rate', slow)
```

7.7.7 Causal statements

The majority of the remaining 566 statements of fact in the knowledge base were causal. This reflects, in part, the nature of the knowledge but also the precedence of causal statements in the representation of that knowledge. As a result, the causal content of the knowledge base represents the primary description of the domain. Means of evaluating the causal content of the knowledge base are necessary both to facilitate the process of moving towards an increasingly coherent representation and to allow a user to access and explore that description. Causal knowledge is particularly important in AKT knowledge bases created for use with the currently available reasoning tools. The query tool (Appendix B) represents a flexible means of retrieving knowledge (which may involve some reasoning), but the navigation tool (Appendix B) is the primary vehicle for the dynamic exploration of the content of the knowledge base.

While diagrammatic representation had been used by both Thapa (in preparation) and Garde(1992) in creating components of the combined knowledge base, none of these diagrams were recorded in the combined knowledge base; instead the knowledge was represented as a list of statements of fact (§7.7.1). As a result, the navigation tool provides the only means of exploring the linkages between causal statements. Using this tool the user may move forwards or backwards along a chain of linkages, exploring the consequences or causes of a particular event.

However, the navigation tool was found to be seriously constrained in at least three ways :

- limitations in the functionality of the tool meant that many legitimate linkages between statements were not recognised;
- while the tool allowed the user to move along chains of causality it did not provide an effective overview of the web of causal relationships; and,
- the interface for the navigation tool demanded that the user both understand the structure of the formal grammar and know sufficient about the content of the knowledge base to know what questions to ask.

These problems are further discussed below.

(i) Inferencing inadequacies

In the current implementation of the navigation tool, causal linkages are only recognised if they are 'in the same direction'. So, for example, the linkage between the statements :

```
An increase in crop moisture stress causes a decrease in crop
vigour
```

and

```
A decrease in crop vigour causes decrease in crop stem thickness
```

would be recognised. However, the linkage would not be recognised if the second statement were represented instead as :

```
An increase in crop vigour causes an increase in crop stem
thickness.
```

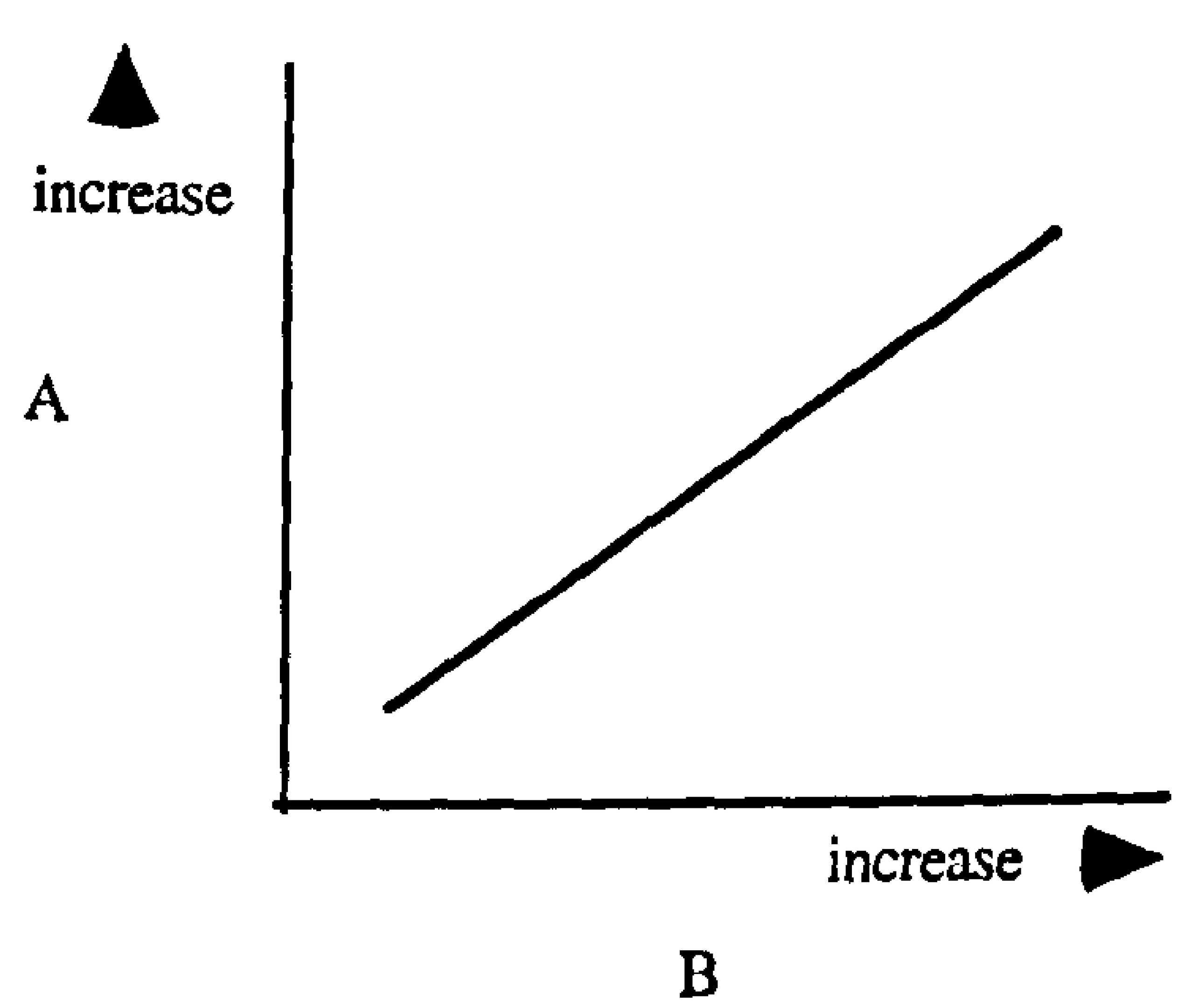
One potential solution to this problem is to recognise that :

```
an increase in A causes an increase in B
```

is exactly equivalent to :

`a decrease in A causes a decrease in B`[14].

This can be visualised as moving either up or down the relationship when graphically represented :



[14]By the same token:
'increase in A causes a decrease in B'
is the same as
'decrease in A causes an increase in B'.

This correspondance may not always be meaningful. For example, an individual plant cannot get younger (decrease in age). However, given that statements will normally refer to groups rather than specific instances this may rarely be an issue, so, for example, it is quite reasonable to state that :

```
Younger trees are less prone to wind throw than older trees
```

While conditions will generally apply for the inverse of a statement as well as the relationship as stated, there may be instances where this does not occur.

Modification of the navigation tool is demanded such that the equivalents outlined above are recognised (except in the exceptional circumstances in which they do not apply). During formal representation, the user should be asked to confirm that the inverse applies. Where it does, the corresponding formal representation should then be automatically generated such that the intermediate statement is paired to two formal statements. These would both then be accessible for formal reasoning, allowing all deducible causal linkages of this type to be used.

(ii) Overview

The navigation tool can provide an effective means of exploring the causes or consequences of an event. However, it has already been demonstrated that sets of chains of causal association (e.g. Table 5.3) give a less useful impression of relations than the equivalent diagram (e.g. Figure 5.1). In the same, way, the navigation tool provides a means of moving along such chains but does not provide an effective overview. As a result the tool may provide an effective means of answering particular questions but does not provide effective support for the user wishing to synthesise his or her understanding. In particular, the navigation tool only illuminates causal linkages where these are completely identified. In the creation of an increasingly coherent representation it is necessary to identify points where linkages are implied (i.e. can be deduced or seem likely) but are not captured, either because the necessary knowledge is inadequately expressed or because it is missing from the knowledge base.

(iii) Interface inadequacies

The interface for the current navigation tool is shown in Appendix B. This is all that is presented to a user in setting out to explore the knowledge base. As a result, anyone wishing to explore the causal content must first run through enough of the statements of fact in the knowledge base to identify a suitable staring point. They must then be able to frame a suitable question through use of the formal grammar. While these are not fundamental constraints they limit the ease with which the causal content of the knowledge base can be evaluated.

In summary, while, in principle, the navigation tool provides an effective means of exploring the causal content of the knowledge base, implementational constraints, solutions for which are given above, currently make it of limited utility.

In an attempt to overcome these constraints in the present context, alternative means of exploring the causal content of the combined knowledge base were explored. It was decided that this might best be achieved by creating a diagram or set of diagrams of the causal content of the knowledge base. While a diagramming interface for the intermediate representation of knowledge (§5.2.2) has been implemented (§6.5.3), an interface for automatic creation of diagrams from formal statements of fact (§5.6) has not yet been developed. However, a diagrammatic representation remains the best means of evaluating a set of causal statements of fact. For this reason an attempt was made to delete all the causal statements of fact from the combined knowledge base and re enter them as diagrams. The procedure used is summarised in Table 7.2.

Table 7.2 Creating diagrams from a knowledge base

1. Prolog source code file created from AKT object code file.
2. An arbitrary causal statement identified.
3. The statement represented diagrammatically on paper.
4. Statement deleted from the source file along with all associated information.
5. Either the caused or the causing event selected and another causal statement in which that event occurs identified through use of the MacProlog searching facilities.
6. Back to 3. until all causal statements containing the selected event are deleted from the knowledge base.
7. Another event on the growing paper based diagram selected then back to 3, until all the causal statements removed from the knowledge base.
8. MacProlog source code converted back to AKT object code file.
9. Paper-based diagrams entered back into the knowledge base through the diagramming interface and then reformalised.

Contrary to expectation, the knowledge from different sources (principally Garde, 1992 and Thapa, in preparation) became well mixed in the resulting diagram. However, significant problems were encountered in the management of the diagram. It had been anticipated that a number of discrete causal diagrams could be created. However, sensible boundaries between such diagrams could not be identified. Any attempt to break the growing causal network, containing hundreds of statements of fact, meant leaving a significant number of links out. Attempts to reduce the size of the causal diagram by imposing a hierarchical structure (§5.4) were equally frustrated. While this experience does not cast doubt on the utility of creating hierarchically structured sets of diagrams, it clearly demonstrated that imposing a hierarchical structure on causal knowledge may not, in practice, be possible.

The causal content of the combined knowledge base could only be represented as a single causal diagram containing several hundred links. The size of this diagram (over 500 links) is well in excess of the size of an individual diagram that can be accommodated by current implementations of the AKT software (about 50 links) and, indeed, of what can be usefully represented on a single diagram (up to 100 links).

In summary, it was found that evaluation of the causal content of the knowledge was extremely constrained. It proved impossible to rationalise and synthesise causal knowledge from different sources except by consideration and manual manipulation of the list of causal statements. It is proposed that this problem would most effectively be alleviated by improving the inference mechanism of the navigation tool and by improving the interface such that :

- the user is able to construct queries in natural language using menus to identify the events which are recorded within causal statements in the knowledge base; and
- the users exploration of the content of the knowledge base is recorded on screen as a growing and manipulatable diagrammatic representation of the network of causal associations explored.

7.7.8 Evaluation of link and comparison statements

In addition to the attribute value and causal statements, the knowledge base contains some link and comparison statements. At present, these can be retrieved through use of the query tool but play no role in any other reasoning. While individual statements can be accessed by scrolling through the knowledge base or, in their formal form, by the use of the query tool in the reasoning mode there is effectively no support given in the software for the evaluation of their significance. The intention is that the navigation tool should be extended to allow the exploration of link statements. By contrast, inference mechanisms making use of comparison statements have not yet been designed.

7.8 DISCUSSION

(i) Knowledge acquisition

The case study described in this chapter has resulted in a knowledge base integrating knowledge from four different sets of sources on tree fodder resources and their place in farming systems in the mid hills of Nepal and northern India.

The knowledge base can be conceptualised as containing information on 129 different species of fodder trees and the fodder derived from those trees and a complementary set of knowledge about the interactions of these tree species and fodders with other components of the hill farming systems in which they are found.

The case study has demonstrated that the AKT approach and software provide a practical environment for representing ecological knowledge as intermediate statements of fact and representing a significant proportion of ecological knowledge as formal statements of fact.

Furthermore, the case study suggests that the AKT approach provides an appropriate means of representing ecological knowledge about agroforestry. The resulting knowledge base combines some scientific knowledge from key texts with knowledge derived from a range of local informants in two different communities. While there is some conflict between the knowledge from these four different sources, they are integrated into an effective common resource.

One of the fundamental premises of the research undertaken in this thesis was that it would be useful to develop formal mechanisms for synthesising local and scientific knowledge about the ecology of agroforestry. While this case study does not, *per se*, demonstrate that this is useful, it has clearly shown that the AKT approach and software provide a suitable environment in which to undertake this synthesis.

In Chapter 1, a comparison between scientific and local knowledge about agroforestry was undertaken. It was suggested that no fundamental difference in the motivation or the process of generation existed between these different sources of knowledge. In Chapter 3 statements of fact were proposed as a structure for capturing any ecological knowledge about agroforestry from any source. This paradigm has been tested through implementation (Chapter 6 and Appendix 2), demonstrating that this knowledge can be represented. Furthermore, application has not suggested any reason to believe that there is ecological knowledge about agroforestry that cannot be expressed as a statement of fact or a set of statements of fact.

In Chapter 4, a formal grammar for the representation of ecological knowledge about agroforestry was proposed. Formal representation using this grammar restricts the structure of represented knowledge.

Nevertheless, it has been demonstrated that most ecological knowledge about agroforestry can be represented using this grammar.

As a result, local and scientific knowledge have been treated in exactly the same way in the creation of knowledge bases, making them comparable and, for the purposes of synthesis, compatible in structure. The processes of comparison and combination are therefore concerned with the rationalisation of differing terminologies and concepts rather than structure of articulation.

The question remains, however, as to whether undue distortion occurs in application of the knowledge representation developed. Distortion occurs only where the sense of knowledge, rather than the idiom of its articulation, are altered. It has to be accepted that some cross-cultural translation (and, indeed, linguistic translation) must be undertaken if knowledge from different sources is to be used in conjunction.

So, two questions are important :

- Are the basic structures used for knowledge representation here (causality, comparison, other linkages, negation, conditionality, taxonomic or other hierarchical relationship) valid for use in representing scientific and local knowledge?
- Is there significant scientific or local knowledge about the ecology of agroforestry whose sense cannot be captured through representation combining these structures?

It is suggested, on the basis of experience, that all the structures proposed are widely used in the natural articulation of scientific knowledge about the ecology of agroforestry. Clearly, the same assertion cannot be made about all the different cultures in which local or indigenous knowledge about agroforestry may be articulated. Explicit tests to assess the use of the structures proposed amongst communities for whom local knowledge has been represented would be merited. Nevertheless, no one applying the formal grammar to date has identified any reason to believe that the structures proposed do not capture the sense of natural articulation by their source communities.

The second potential source of distortion is the exclusion of knowledge that cannot be represented from the knowledge base. It is clearly impossible to demonstrate with absolute certainty that a formal grammar captures all knowledge of a defined type. Furthermore, it is difficult to envisage valid means of undertaking such a test to a stated degree of confidence because of the inherent difficulties in controlling the interpretation involved in representation. It seems sufficient to point out that, to date, no significant types of predictive or explanatory knowledge about agroforestry have been found that cannot be represented using the formal grammar, always with the caveat that the legitimacy of interpretation in the process of formal representation is not known.

In summary, it has been shown that the approach to knowledge representation developed does enable the comparison and combination of local and scientific knowledge about agroforestry and that there is no reason to believe that either is unacceptably distorted during this process.

(ii) Accessing and exploring the content of the knowledge base

The contents of this knowledge base may be accessed by the user in a number of ways. The user can scroll through a list of statements of fact, accessing further details on individual statements of fact. Those sets of statements of fact recorded diagrammatically may also be explored through the diagramming interface, again accessing further details for particular statements of fact. Alternatively the knowledge base, or parts of the knowledge base (intermediate statements of fact, object hierarchies, glossaries or source information) may be written to a separate text file or printed.

The use of selection facilities (§6.5.6) provides a more dynamic means of exploring the content of the knowledge base, analogous to searching a data base. The intermediate statements of fact resulting from a keyword search (using the node-only search option) the term *Bauhinea purpurea* are shown in Table 7.3.

The combination of the knowledge base and the AKT software transcends a database function because the formal representation of the content of the knowledge base enables automatic reasoning with that content. In the current version of the AKT software this is through the use of the navigation and query tools (Appendix B).

The case study described in this chapter has demonstrated that while these existing approaches to accessing and exploring the knowledge base are useful in creating a coherent and comprehensive knowledge base, they do not provide adequate support. Further development is demanded.

Table 7.3 Statements in the combined knowledge base identified by keyword (node only) search on the term *Bauhinea purpurea*

'Tanki' is the Nepali name for *Bauhinea purpurea*. Lahara tanki and buthure tanki are varieties of tanki. Statements containing these terms have been identified here in spite of this being a 'node only' search becaue they contain the word 'tanki'.

lahare tanki has bigger leaf than buthure tanki (Thapa, Balaram*1993)
lahare tanki is more palatable to animals than buthure tanki (Thapa, Balaram*1993)
lahare tanki is more posilo to animals than buthure tanki (Thapa, Balaram*1993)
lahare tanki produce more foliage biomass than buthure tanki (Thapa, Balaram*1993)
lahare tanki produce more number of pods than buthure tanki (Thapa, Balaram*1993)
Bauhinia purpurea is found up to 1500 m altitude (Singh, R.V.*1982)
Bauhinia purpurea is found at between 1000 and 2150 mm of rainfall (Singh, R.V.*1982)
Bauhinia purpurea flowers from september to december (Singh, R.V.*1982)
Bauhinia purpurea flowers are pollinated by bees (Singh, R.V.*1982)
Bauhinia purpurea seedlings have a strong tap root (Singh, R.V.*1982)
Bauhinia purpurea seedliongs stopgrwoing in winter montrhs (Singh, R.V.*1982)
Bauhinia purpurea is a moderate light demander (Singh, R.V.*1982)
Bauhinia purpurea is frost hardy (Singh, R.V.*1982)
Bauhinia purpurea leaves are a medium quality fodder (Singh, R.V.*1982)
Bauhinia purpurea is found at below 800 m (Panday, K.K.*1982)
Bauhinia purpurea may be deciduous or may be evergreen (Panday, K.K.*1982)
Bauhinia purpurea has big leaf (Thapa, Balaram*1993)
Bauhinia purpurea has bilopped leaf (Thapa, Balaram*1993)
Bauhinia purpurea has droopy leaf (Thapa, Balaram*1993)
Bauhinia purpurea has large crown (Thapa, Balaram*1993)
Bauhinia purpurea has round shaped crown (Thapa, Balaram*1993)
Bauhinia purpurea has thick crown (Thapa, Balaram*1993)
Bauhinia purpurea is a malilo fodder tree (Thapa, Balaram*1993)
Bauhinia purpurea is a medium size tree (Thapa, Balaram*1993)
Bauhinia purpurea is a posilo tree (Thapa, Balaram*1993)
Bauhinia purpurea is fed to buffalo (Thapa, Balaram*1993)
Bauhinia purpurea is fed to cattle (Thapa, Balaram*1993)
Bauhinia purpurea produce edible shoots (Thapa, Balaram*1993)
Bauhinia purpurea put on regrowth rapidly after lopping (Thapa, Balaram*1993)
Bauhinia purpurea regenerate abundantly in the farmland (Thapa, Balaram*1993)
Bauhinia purpurea shed leaf late in the dry season (Thapa, Balaram*1993)

7.9 CONCLUSIONS

The current implementation of the AKT software is a powerful environment for knowledge acquisition, resulting in an explicit record of ecological knowledge.

It also provides an environment for accessing and exploring the contents of the knowledge base. However, a number of desirable improvements have been identified. These include:

- upgrading of the diagramming interface;
- development of means for automating the formal representation of statements entered through the diagramming interface;

- provision of mechanisms for recording secondary source information; and
- provision of mechanisms supporting the development of explicit definition of keywords.

CHAPTER 8 APPLICATION OF THE AKT APPROACH

8.1 INTRODUCTION

In Chapter 7, the functionality of the AKT software and approach in creating concise and explicit knowledge bases was considered. This chapter presents an assessment of the utility of such knowledge bases in a development context.

Assessment was achieved by case study investigation of:

- constraints to effective knowledge management in research and extension activities within a development-orientated research institution, and
- the potential application of the AKT software and approach in addressing these constraints.

This case study approach enabled detailed and rigorous analysis in a specific context of application. It is likely that many of the conclusions resulting from this case study are applicable to other development organisations.

8.2 CASE STUDY BACKGROUND

Following a four week visit to Pakhribas Agricultural Centre (PAC) in Nepal in 1992, an eight month study entitled 'Use of a knowledge-based systems approach in the improvement of tree fodder resources on farmland in the eastern hills of Nepal - Pilot phase'[15] was inititated. The evaluation of the application of the AKT approach within existing development activities described in this chapter was undertaken in the context of this project.

The project undertook an analysis of constraints to research and extension at PAC. This was achieved between May and July 1993 through analysis and synthesis of the information collected by Bianca Ambrose (see Sinclair *et al.*, 1993, for methodology employed in collecting the information). This analysis provided a basis for the evaluation of the potential contribution of a knowledge-based systems

[15]The research was undertaken by the author in collaboration with Balaram Thapa and Laxman Joshi of PAC and Fergus Sinclair of the School of Agricultural and Forest Sciences of the University of Wales Bangor and funded by the UK Overseas Development Administration. Bianca Ambrose was employed as a research assistant on the project and spent three months at PAC in 1993. The study has been reported in Sinclair *et al.* (1993).

approach in addressing these constraints and the development of proposals for the modification of the AKT software and its implementation at PAC.

8.3 INSTITUTIONAL BACKGROUND

PAC is located 25 km north-west of Dhankuta, 87°30'E- 27°10'N, in the eastern hills of Nepal. It is responsible for research and development activities in natural resource management in 11 districts ranging from 300 m asl. to over 4000 m asl., with a climate that varies from sub-tropical to alpine and arctic, and an ethnically and economically heterogeneous farming community (Gibbon and Schultz, 1989). Farming systems are, as a result, diverse.

PAC is divided into ten technical sections, involved in strategic and adaptive research, leading to the production of improved materials and technologies, and the development of extension materials (PAC, 1992).

Extension is undertaken by relevant government agencies. However, PAC runs an Outreach Research programme which forms the closest link between farmers, PAC researchers and extension workers (Gibbon, 1993). Outreach Research maintains direct contact with the farming community and provides a mechanism for monitoring and evaluating the results of PAC research outputs (PAC, 1990).

The analysis of constraints to research and extension was based on the premise that there were opportunities for improving the effectiveness of research at PAC through the development of AKT knowledge bases in two main ways:

- research could be made more relevant to the priorities and existing knowledge of client farmers; and
- more effective use could be made of previous research findings.

Both depend on the nature and extent of existing local and scientific knowledge and its accessibility to researchers. On this basis, the analysis of constraints involved :

- description of the research cycle in operation at PAC;
- description of the interactions occurring amongst PAC staff and between them and other external agencies involved in research;
- analysis of the use and flows of knowledge in planning, implementation, and reporting of research and thereby

- identification of points in the research cycle where the effectiveness of research was constrained, characterisation of these constraints and identification of opportunities for interventions that would improve the effectiveness of the research process.

A summary characterisation of the research cycle at PAC is given in §8.4. The constraints identified through analysis of the research cycle are summarised in §8.5, in §8.6 the potential contribution of creating and using AKT knowledge bases in alleviating the constraints is examined.

8.4 SOURCES, FLOWS AND STORES OF KNOWLEDGE AT PAC

The flows of knowledge between the five main sinks and sources of knowledge at PAC are summarised in Figure 8.1. This provides an overview of the research cycle at PAC. Figures 8.2 and 8.3 expand on the cycle shown in Figure 8.1.

8.4.1 Overview

The PAC research cycle began and ended, as far as possible, with the farmer. PAC technical sections collected and used knowledge from farmers either through specific collection activities, or as a result of on-farm outreach research. There was also a limited amount of informal exchange of knowledge from researchers to the farming community through informal researcher contact with farmers during the execution of formal research activities .

Extension workers were an important link between the researcher and farmer. Extension workers brought technologies, recommendations and other knowledge generated by research to the farming community. Research outputs were sent either directly to extension agents from PAC technical sections, or through PAC outreach activities. As a corollary, extension workers and line agency staff gained an understanding of farmer reaction to research recommendations and technological interventions. Some of this feedback from farmers, accumulated by extension workers, passed back to researchers through formal contacts, for example during research planning consultations, or through less formal contacts with researchers, for example via their participation in outreach activities.

Scientific literature represented another knowledge input to PAC. There were many exchanges of knowledge within the research community itself. During the research process, previously documented scientific knowledge, and local knowledge was evaluated and used in planning and executing research and included in research outputs.

Outputs from research included information for dissemination to extension workers and farmers and research results for use in further experimentation.

8.4.2 Flows of knowledge within PAC

Flows and uses of knowledge within the research community are represented in Figure 8.2. Some of the knowledge already existing at PAC was documented, some remained undocumented. Local, professional and scientific knowledge all played an important role.

Documented knowledge was collected, collated and stored for future access. The stores themselves were situated in a range of locations and were of a range of types. Central stores were open-access areas such as the central technical library. Section stores were subject to restricted access. Personal stores represent the private collections of material belonging to individual researchers and other staff members.

Documented knowledge was stored in a number of forms, including reports and PAC technical papers, scientific papers and researchers' notes. A significant proportion of the knowledge used during the research cycle was not documented.

Researchers identified perceived farmer priorities through formal institutional mechanisms of prioritisation as well as informal personal interpretation of information. These provided a basis for prioritising research. Planning processes, at centre and section level, acted to produce research protocols. Protocols were produced at project level and at centre level. After being subjected to centre and section approval, research protocols were then implemented.

Research activities fell into two main categories: core research, driven by institutional activity at the centre, and outreach research, which was essentially adaptive in nature. Core research was basic, strategic or applied. Outreach research provided further information for, and feedback on, core research. It was adaptive and linked to core research. Both types of research activity generated new knowledge which formed the basis of research output, or was fed back immediately into ongoing research activities. The new knowledge generated through research formed the primary research outputs, produced as documents or technology packages.

In documented form, the outputs were disseminated to farmers and extension workers, or placed within the existing body of knowledge within the centre. Placing these outputs into PAC stores added to the available body of knowledge. Research results also had an impact on the professional knowledge held and used by researchers. Both processes of (i) adding documents to the central body of documented knowledge, and (ii) changing the professional knowledge of research staff, can be considered as updating the knowledge within the centre. This updating added to the breadth and depth of the knowledge available as a basis for further research.

8.4.3 Interactions with farmers and extension workers

The use and flows of farmers' local knowledge in relation to PAC activities is outlined in Figure 8.3.

Farmers' local knowledge was passed on to the PAC research community in a number of different ways. These included contact and participation in PAC information and knowledge collection activities; participation in experimental research activities; and informal interactions with researchers who came into contact with the farming community as part of their day-to-day activities either on or off-station.

When knowledge was synthesised into extension materials, it was sent directly to farmers, or passed to farmers through the efforts of extension workers. PAC research activity findings also reached farmers through informal personal contacts. New knowledge, in both documented and implicit form, was interpreted and assessed by farmers. Some was accepted, added and integrated into existing local knowledge. Extension workers also passed knowledge that they had received from the farming community to the research staff through informal contacts or through participation in PAC activities such as the Samuhik Bhramans[16]. Research outputs reached extension agents directly. These could be passed directly to the farming community. Alternatively, having read, absorbed and interpreted the knowledge contained within these documents, extension agents added it to their professional knowledge. This could then be exchanged with the farming community in the course of professional or informal activities.

8.4.4 Characterisation of the sources of knowledge at PAC

The knowledge available to PAC researchers had important impacts throughout the research cycle. The knowledge that existed at the beginning of the cycle was used to prioritise, target, design and implement research work. The efficacy of the capture and analysis of local, scientific, and professional knowledge, at the outset of the cycle determined the extent to which research was related to farmer circumstances and priorities, and the extent to which it capitalised on scientific and professional insights that were already available. Failure to make effective use of existing knowledge restricted the quality and outcome of research activities, resulting in research outputs which were, for example, relevant only to certain small groups within the target community, or duplicated knowledge that was already held by farmers or available through previous research at PAC or elsewhere. These issues had a direct bearing on the effectiveness of research outputs and the efficiency with which resources were used to produce them.

[16]An interdisciplinary team exercise where researchers spend some weeks in the field involved in intensive formal and informal collection of knowledge and data (Chand and Gibbon, 1990).

The type, breadth, and depth of undocumented and documented knowledge that existed and was available to PAC from farmers, PAC staff, the extension community, and other sources is characterised below.

(i) Farmers' needs and knowledge

PAC has a 'near market' research remit (SEADD, 1989). Its research activities are intended to benefit the farming communities found within its research command area. On this basis alone, a clear understanding of the current state of knowledge of the farming communities and the needs of those communities is critical to effective research.

(ii) Farmers' local knowledge

PAC has made extensive efforts to utilise local knowledge, through, for example, the Outreach Research programme, and the documentation and survey of farmer practices and understanding carried out as a precursor to research investigations. Research by the technical sections has demonstrated the nature of the knowledge held by the farming community to be diverse and fundamental. It includes a detailed knowledge of plant taxonomy, soil taxonomy, soil fertility processes, ecological relationships, and farming systems management (Chand *et al.*, 1990).

Farmers in the eastern hills have been practising agroforestry techniques for many years, as trees and fodder resources form an integral part of local farming systems. An increasing body of evidence points to agroforestry practices becoming more widespread and significant as the number of trees on farmland increases across the eastern hills (Carter and Gilmour, 1989; Subba, 1992). This research has led to the suggestion that farmers in the region may have developed a substantial body of knowledge about, for example, tree-crop interactions, fodder tree management techniques, and inter species and intra species variation in fodder quality and productivity. Preliminary research by Thapa (in preparation) has demonstrated that farmers' knowledge can be explanatory and predictive as well as of technical relevance.

Farmers' local knowledge can be expected to be a valuable source of knowledge about land management practices. The opportunities open to research at PAC through a more effective utilisation of the knowledge that exists within the farming community may be substantial.

Furthermore, investigation of farmers' local knowledge may enhance understanding of the needs of the people in communities forming PAC's overall client group. This understanding is critical in formulating effective research. The eleven hill districts in the PAC research command area cover an area of 2.2 million hectares with an approximate total of three hundred thousand farming households (PAC, 1992). Differences in caste, ethnicity, socio-economic status and the bio-physical environment often occur over short distances, and are substantial in their extent and nature. The needs and priorities of farmers in PAC's research command area are particularly complex and diverse.

PAC has, particularly through the efforts of the Outreach Research programme, made strenuous attempts to address this problem. Outreach activities collect detailed information through documenting farmer reactions to experimental interventions, conducting household acceptance tests, profiling the ethnic composition of outreach sites, and categorising households according to food security criteria, which PAC uses to facilitate an understanding of farmers' needs and perceived priorities. The low rate of adoption of agricultural technologies introduced into the PAC research command area (Kunwar and K.C., 1993), and reports suggesting that previous PAC research has been useful to a minority of better-resourced, progressive farmers (World Bank, 1990), indicate that incomplete understanding of the needs and priorities of the farming community in PAC's research command area is still limiting the effectiveness of research.

Some research is, however, clearly linked to farmers' needs and priorities. Two factors are of critical importance to the establishment and success of such research:

- the informal flows of knowledge and communication between PAC staff and farmers; and
- a lack of knowledge about particular concerns common to the majority of farmers.

(iii) Knowledge held by PAC staff and extension workers

PAC staff at all levels, from technicians to research officers, hold a substantial body of scientific, professional, and local knowledge. This may have been acquired during professional training, through participation in research activities or through local experience.

Extension agents form an important point of contact between researchers and farmers. PAC's previous experience with extension work has shown that the stimulus for improved research can often come from extension staff (Thapa *et al.*, 1988). Their general understanding of farming communities and their knowledge concerning the adoption or rejection of research interventions is of particular interest to research at PAC.

(iv) Other scientific and professional knowledge

In addition to PAC publications and research outputs, a wealth of documented scientific knowledge is available to PAC research staff from the external research community. The depth and breadth of this knowledge is significant. Most is available in books, journals and other papers or articles. Not all of this documented scientific knowledge is of immediate relevance to the PAC research command area.

Nevertheless, fundamental knowledge and concepts important to promoting efficient and effective research is available.

8.5 CONSTRAINTS TO EFFECTIVE RESEARCH AND EXTENSION AT PAC

The constraints presented here are categorised in relation to : knowledge acquisition; the storage of, and access, to knowledge; the use of knowledge in planning research; and producing research outputs.

A perceived need for improved means of sharing the results of research activities (be they experimentation, knowledge acquisition or synthesis) between staff at PAC ran through all these categories. The applied and adaptive research undertaken at PAC typifies development-orientated research activities in rural resource management in general, and agroforestry in particular, in demanding an interdisciplinary approach. However, the analysis clearly identified problems encountered in creating an interdisciplinary structure at an institutional level. PAC remains, therefore, divided into sections on the basis of discipline. Furthermore, research fora that promote and provide for interdisciplinary exchanges of knowledge were limited. There were no formal mechanisms to allow for the formulation and implementation of joint research protocols, or joint research strategies. There were no structures to facilitate and maintain co-ordination between researchers from different sections. As a result, knowledge was confined to individual sections, and not accessible to other sections.

8.5.1 Knowledge acquisition

The analysis of activities in knowledge acquisition demonstrated that not all the knowledge available to PAC that was useful was articulated. Furthermore, not all the knowledge that was articulated was recorded. The recorded knowledge was often neither coherent, nor comprehensive in relation to target research topics and often contained unresolved contradictions. Furthermore, it tended to be recorded in a form suitable for the immediate purpose at hand but of little value for future reference or for use by other researchers with different objectives.

Detailed constraints in knowledge acquisition were identified in relation to both formal and informal knowledge acquisition from farmers, PAC staff, extension workers and non-PAC scientific and professional sources.

(i) Formal acquisition of farmers' knowledge

Objectives for the collection of farmers' knowledge and statements of need by PAC staff were not explicitly stated. There were no generally accepted guidelines at PAC on the methods to be used for knowledge elicitation and recording. There was no agreed, standardised and complete scheme for the classification of the biophysical nature of farmland, and the socio-economic character of farming communities. This caused difficulties in sampling of farmers, and recording of standard information about informants. As a result, knowledge elicitation activities carried out by individual sections provided knowledge of limited utility to other sections within PAC.

(ii) Informal acquisition of farmers' knowledge

While considerable informal knowledge acquisition from farmers occurred at PAC, there were no mechanisms for recording the resulting knowledge and evaluating it in relation to its implications for planning research and extension activities. This contraint particularly applied to:

- the substantial body of farmers' knowledge acquired by outreach staff;
- feedback from farmers about on-farm experimental work; and
- feedback from field-level technical staff.

Furthermore, farmer/researcher and researcher/line agent interactions and exchanges of knowledge on-station were not recognised as useful sources of farmers' knowledge.

iii) Acquisition of knowledge from PAC staff

Despite substantial informal interaction between research staff on a 'need to know' basis and the recognition of the value of more formal recording of the knowledge held by PAC staff, there were no mechanisms for achieving this.

(iv) Acquisition of knowledge from extension workers

Although there were regular interactions between PAC and extension staff, and the knowledge held by extension staff had been shown to be useful in informing the research process, there were no formal mechanisms for eliciting and recording such knowledge. Where extension agents took part in PAC activities, few explicit statements of extension workers' knowledge pertinent to research at PAC were recorded.

(v) Acquisition of other scientific and professional knowledge

Although there was a recognition that knowledge held by research staff at other development institutions was useful to research at PAC, there were no mechanisms for recording this knowledge once elicited.

(vi) Recording knowledge during knowledge acquisition

Knowledge that was articulated during knowledge acquisition exercises was rarely explicitly recorded. Instead the implications of that knowledge in relation to the particular objectives of the exercise were recorded. As a result, knowledge was not recorded in a form appropriate to future consideration in relation to other research activities for which it may have been relevant.

8.5.2 The storage of, and access to, knowledge

The analysis showed that knowledge that had been acquired from farmers, previous research, PAC staff, extension workers, and line agency staff by researchers was rarely stored as explicit statements. The knowledge that was recorded was not easily accessible to researchers because:

- it was difficult to locate the knowledge relevant to a particular research topic; and
- it was frequently recorded in a form from which it was difficult to abstract information relevant to research.

Furthermore, knowledge was not stored in collections relating to targeted research needs. It was found that there was no central mechanism organising access to knowledge in relation to particular research topics because (a) information was dispersed through diverse stores at PAC and not centrally catalogued, (b) catalogues allowing access to references according to a problem focus did not exist.

8.5.3 Use of knowledge in planning research

Research can be divided into three stages : planning, implementing and evaluating. Planning research activities in a development context can be conceptualised as comprising four steps :

- identifying potential research topics;
- prioritising potential research topics in relation to their predicted impact on decision making;

- prioritising research topics in relation to client priorities, resource allocation, institutional and researcher motivations and objectives; and

- designing research procedure.

The emphasis here is on identifying and prioritising research objectives.

The analysis demonstrated that it was not possible, at PAC, to base research planning on a rigorous evaluation of the existing knowledge of farmers, PAC staff and extension staff, and from previous research. In prioritising research, formal, explicit statements of farmer needs were limited in number and depth, and there was confusion between farmer needs, as articulated by farmers, and researchers' perceptions of farmer needs.

In particular:

- available methods for evaluating recorded knowledge were not powerful enough to enable the knowledge to be used to fulfil research objectives;

- farmer priorities were recorded and research recommendations were generated from them, but there were no formal mechanisms for using these in the prioritisation of research; and

- there was no guidance facilitating the prioritisation of research on the basis of the overall and sectional objectives of PAC.

8.5.4 Production of research outputs

Outputs produced from research undertaken at PAC can be divided into three categories: physical inputs, technologies and knowledge.

Physical inputs - Physical inputs into the target farming system may include improved crop or livestock strains, chemical inputs and machinery or other equipment.

Technology packages - Prescriptive technology packages, for example the combination of an improved rice variety seed and detailed instructions on seeding rates passed on for extension to farmers in specified agroecological conditions, require limited understanding for application.

Knowledge - Research generates knowledge about the components and interactions in farming systems and their management. This knowledge may be applied by farmers in their decision making processes.

These categories are not mutually exclusive, and may frequently be complementary.

Extending *knowledge* to the farmer in order to provide a more robust and accurate basis for decision making is a more challenging process than extending physical inputs or technologies. However, knowledge, rather than inputs or technologies, may be much more flexibly adopted and adapted by the farmer to suit specific requirements. Physical inputs or technologies may, in the absence of accompanying knowledge, be difficult for the farmer to effectively integrate into current farming practices, particularly where only certain elements of the technology are relevant.

Analysis of activities at PAC showed that research was not formulated in terms of the new knowledge that it would generate. For this reason, explicit statements of the knowledge gained by research activities were not routinely produced. As a result, more knowledge was created during research than appeared in outputs. Furthermore, the needs of target clients, in terms of knowledge and of knowledge presentation, were not fully understood and procedures for updating central stores of knowledge were restricted to adding new knowledge, as opposed to re-evaluating existing knowledge.

As a result, documents did not necessarily meet the demands of clients, and there were no formal mechanisms for clients to express their needs. Research outputs from various activities and sections relevant to clients' particular problems were rarely synthesised. Documented outputs fulfilled different, and sometimes conflicting, roles including (a) recording the current state of knowledge, (b) recording the outcome of research activities, and (c) synthesising knowledge to fulfil the defined needs of defined users. There were no mechanisms for deciding whether knowledge generated by research should be disseminated, and, if so, how this should be done.

Furthermore, it was found to be difficult and time consuming to abstract key research findings for dissemination: as a result, time lags between research and the dissemination of outputs were often significant.

8.6 THE APPLICATION OF THE AKT APPROACH

In §8.5, constraints to improving the relevance of research at PAC to client farmers were described in relation to:

- knowledge acquisition;
- storing and accessing recorded knowledge;
- using existing knowledge in planning extension activities; and
- using existing knowledge in planning research.

In this section, the utility of the AKT approach in a development context is evaluated through consideration of the role of the creation and use of AKT knowledge bases in addressing these constraints at PAC.

8.6.1 Knowledge acquisition

The PAC case study demonstrated a role for effective means of knowledge acquisition in evaluating the current state of knowledge. Trial application of the AKT approach has demonstrated that creating knowledge bases can facilitate rigorous knowledge acquisition[17] because this process :

- provides a framework for eliciting and recording knowledge;
- enables iterative evaluation of the knowledge acquired, thereby driving further knowledge elicitation;
- provides an environment in which knowledge acquired from different sources (e.g. local communities and researchers) in different forms (e.g. through interview and from literature review) may be combined and synthesised; and thereby
- provides a framework for the explicit, unambiguous and comprehensive statement of available knowledge on a defined domain.

The approach provides, therefore, a framework for more methodological and rigorous knowledge acquisition than is currently undertaken at PAC, resulting in a more explicit and coherent record of the knowledge articulated. The standardised structure of this record means that it will be :

- more amenable to synthesis with knowledge elicited from other sources or collected on different occasions for different objectives, and
- more appropriate for future evaluation against criteria other than those for which it was orginally developed,

than the results of current knowledge acquisition activities at PAC, which were often informal, *ad hoc* and rarely resulted in comprehensive records of results.

For example, Thapa(in preparation) and Garde(1992) were able to create coherent and succinct knowledge bases (§7.7.1) using the AKT approach, that recorded detailed information about the ecology and management of tree fodder resources. These enabled detailed evaluation of local

---

[17]Garde(1992), da Costa(1993), Holmes(1993), Jinadasa (in preparation); Kilihama (in preparation); Preechapanya (in preparation), Southern (in preparation), Thapa (in preparation).

knowledge on this domain. Furthermore, the resulting knowledge bases were subsequently merged without reference to these two researchers (§7.7.5). It is doubtful whether these activities would have been as effectively achieved by relying on the approaches to knowledge acquisition currently employed at PAC.

Adopting more formal approaches to knowledge acquisition can enhance rather than displace existing activities. So, the AKT approach can build on existing rapid rural appraisal methodologies, such as the Samuhik Bhraman at PAC. Encouraging the acquisition of explanatory ecological knowledge rather than the analysis of farmer problems that predominates in rapid rural appraisal methodologies (Hildebrand, 1990), may result in a more flexible and robust resource. It may be, however, that these two approaches will be found to be complementary, enabling more effective evaluation of farmer needs. An evaluation of the efficacy and efficiency of the AKT approach to knowledge acquisition in comparison with traditional rapid rural appraisal techniques, as well as the complementarity of these two approaches is under way (Hitanyake, in preparation).

One of the key findings of the analysis of activities at PAC was the need for effective interdisciplinary interaction. This is particularly important during the analysis of farmer problems and the evaluation of the current state of knowledge in planning research and extension activities, hence the emphasis on the interdisciplinary Samuhik Bhraman at PAC. The creation of an integrated knowledge base on a defined problem or domain with consistent use of terminology by a group of researchers requires effective interdisciplinary interaction.

In summary, implementation of the AKT approach at PAC would provide a framework for building on and integrating existing mechanisms, faciliating more effective knowledge acquisition.

8.6.2 Storing and accessing knowledge

In the PAC case study, it was found that existing knowledge was not easily accessible to researchers because (a) it was frequently not recorded, (b) where recorded it was difficult to locate knowledge relevant to a particular research topic, (c) this knowledge was frequently recorded in a form from which it was difficult to abstract information relevant to research, and (d) knowledge was not stored in collections that related to targeted research needs.

The creation of an AKT knowledge base can result in an explicit statement of knowledge. Furthermore, evaluation (Chapter 7) of the approaches to intermediate and formal representation developed in this research (Chapters 3 to 5) has demonstrated that they provide an expressive and flexible means of representing ecological knowledge about agroforestry. The knowledge base is, therefore, a suitable form in which to store that knowledge. The use of computers for storing knowledge bases allows appropriate information technology techniques to be used to make the content of these knowledge bases easily accessible.

AKT knowledge bases can be divided and combined as necessary(§7.5); as a result knowledge from many different sources can be synthesised into knowledge baes with an appropriate topic focus. In the PAC context, creating and maintaining a set of centralised knowledge bases integrating local and scientific knowledge on farmer-led problem or topic foci, thereby creating a corporate and dynamic record of the current state of knowledge, would result in a more easily accessed and robust store of knowledge for use in research and extension activities.

Furthermore, the contents of these knowledge bases may be made more widely available than current stores, for example, by providing research and Outreach staff on-line access to the contents of centralised knowledge bases, while providing extension workers and farmers access to the contents of the knowledge base through the development of appropriate printed material on request (§8.7). This means that the knowledge base is a resource that can be accessed in its own right and also provides a basis for the generation of extension material.

In summary, implementation of the AKT approach at PAC could result in the creation and maintenance of accessible, centralised, integrated and problem orientated stores of knowledge.

8.6.3 Planning extension activities

The development of knowledge bases that are representative of the knowledge of a target community may provide a powerful means of planning extension activities. It has been argued (§1.5) that extension activities will be more effective where based on an understanding of the current knowledge of the target community. The development of an AKT knowledge base may foster this understanding, as may subsequent exploration of that knowledge base. For example, consideration of the content of the knowledge base created by Thapa (§7.7.1) reveals a considerable understanding amongst local farmers about tree-crop interactions. Local perception of the above-ground interactions is based largely on understanding of the relative shade and tapkan effects caused by different tree species and the implications for crop yield. A combination of several tree attributes (including tree height, crown thickness, crown size, crown shape and leaf size) is used in predicting the impact of a species on associated crops. This combination is based on a clear perception of the way in which each of these attributes influences interactive processes such as tapkan. Furthermore, explanations of the mechanisms of impact on crop plant height, crop stem strength, crop lodging, head size and crop yield are well developed. Differences of impact on different crop species and cultivars are known. The impact of a range of spatial and temporal factors, for example aspect and differences in time to maturity of different crop cultivars can be explained.

This information alone demonstrates, for example, that extension activities extolling the virtues of planting tree fodder resources on private farmland will be of little relevance to the community studied.

Rather, the extension of detailed information about the ecology of tree-crop interactions that may be relevant to farmers in formulating their management strategies is merited.

Comparison of the knowledge held by the community with scientific knowledge, or the knowledge of other communities, may be achieved by rigorous comparison of the content of two knowledge bases. This could provide a systematic means of identifying available knowledge from other sources that may be relevant to the target farming community.

The knowledge base also provides a basis for the investigation of the distribution of knowledge within the community. Such investigation may, in conjunction with other socio-economic information, help to identify particular target groups within the community.

Thapa (in preparation), Kilahama (in preparation), Jinadasa (in preparation) and Preechapanya (in preparation) are exploring the use of the AKT approach in characterising the knowledge of target communities and the use of these characterisations in planning extension activities. There is a particular emphasis in current activities on the statistical evaluation of the representativeness of a knowledge base.

In summary, the creation of AKT knowledge bases at PAC that are demonstrated to provide a representative abstraction of the knowledge of target communities could provide a sound basis for planning extension activities.

8.6.4 Planning research

The importance of basing research planning on an evaluation of the current state of knowledge has been highlighted by the case study described in this chapter. In Chapter 1, it was argued that large volumes of complex knowledge may be applicable for use in decision making in agroforestry. Furthermore, knowledge from a range of sources may be relevant because "the greater the universe of perspectives, the greater the opportunity for productive solutions to be found" (Knight, 1980). The case study has demonstrated that this knowledge is dispersed, may be implicit and ambiguous and may be inconsistent in terminology and therefore not immediately comparable. A carefully developed, managed and updated knowledge base provides a comprehensive and reliable resource for use in the evaluation of the current state of knowledge.

For example, the understanding that can be derived from consideration of the contents of the knowledge base developed by Thapa (§7.7.1) may have important implications for tree fodder research at PAC. Systematic research on tree fodder remains in its infancy at PAC. This is partly because the basic information required to formulate research plans is scarce, which means that a clear research strategy is lacking. In this context the knowledge base itself becomes a useful resource in providing a basis for developing an appropriate research strategy. The knowledge base provides insights into farmers' current knowledge about tree-crop interactions, showing that farmers' understanding is based on intimate knowledge of tree attributes and the effect of tree attributes on crops. For example, it suggests that shade and tapkan are the primary means by which trees reduce crop yields, and that tree attributes such as crown and leaf dimensions and height influence shade and tapkan. This indicates that tree canopy architecture is an important area of consideration in planning tree fodder research in the future. In the absence of such information in the past, PAC has tended to follow a more conventional approach to tree fodder research. Species selection programmes have tended to emphasise survival rate, growth rate and total foliage production as important parameters.

8.6.5 Summary

The analysis in §8.5 has suggested that the linkage between the end of one set of research and the beginning of the next at PAC was frequently compromised. The success of individual research was frequently not reflected in advances in 'corporate' understanding.

Causes of this weak linkage included :

- inadequate documentation of knowledge gained through experimental research, interaction with farmers, extension and research communities; and

- methodological problems associated with assessing existing knowledge on a topic in devising research, because of the diversity and dispersed nature of that knowledge.

It is proposed that the implementation of the AKT approach at PAC would address these problems by :

- allowing the multiple and diverse stores and flows of knowledge currently found within the PAC research cycle to be synthesised more effectively into a corporate research cycle;

- encouraging the synthesis of existing knowledge from different sources - resolving inconsistencies, ambiguity, contradiction, conflict and irrelevance;

- enabling the development of a centralised corporate record of the current state of knowledge with a problem-related structure that is more coherent and accessible (and, therefore, more useful) than the sum of knowledge currently held in various places;

- improving the accessibility of pertinent knowledge, thereby improving the knowledge upon which decisions are based; and

- providing a decision-support environment encouraging more rigorous evaluation of existing knowledge.

These improvements would enable :

- a more thorough understanding of target farming communities through evaluation of the knowledge that they hold;
- more effective identification and prioritisation of research objectives, leading to more appropriate research; and
- an improved dissemination of knowledge to target farming communities.

8.7 ENHANCING SOFTWARE FUNCTIONALITY

In §8.6, the utility, in principle, of the application of the AKT approach at PAC (and, by implication, other development-orientated organisations) was demonstrated. However, this utility cannot, in practice, be fully realised through use of the current software. This section proposes further development of the AKT software that could be achieved without modification of the current functionality of the software or the formal grammar but would enhance its utility in a development context.

'Read back' (§3.9), searches on the intermediate or formal knowledge base(§6.5.6), diagrammatic representation of linked sets of knowledge(§5.3) and the use of reasoning tools(Appendix B) have been developed to allow users to access knowledge. Although these approaches are in use, they have been shown to be constrained in the current implementation(7.7). Means of improving the accessibility of the content of the knowledge base have been suggested (7.7.8). These include :

- the development of a mechanism allowing the user to view the contents of the knowledge base by creating diagrams from that content, as opposed to viewing diagrams entered by the knowledge base developer (5.6); and
- adaptation of the current reasoning tools (the query and navigation tools, Appendix B), particularly through the creation of improved interfaces.

These developments would provide the knowledge base developer, or other computer-literate users, with flexible means of accessing and exploring the contents of the knowledge base.

However, not all potential users will be computer literate. For example, it has been proposed in the PAC context that Outreach staff should have on-line access to the knowledge base (Sinclair *et al.*, 1993). Because this group are not computer literate, on-line access would demand the development of simplified interfaces tailored to their needs. The expense of such developments may not be justifiable. An alternative to on-line access would be the development of paper-based interfaces customised for the needs of particular groups. This could provide a broad range of users with access to the contents of the knowledge base. Such a system would demand institutional mechanisms whereby a farmer, extension worker or other user could indicate the information they were seeking and a knowledge base system operator could identify and synthesise relevent knowledge from the knowledge base and produce a printed output. This type of approach has been applied by the Idaho Department of Lands in generating customised documentation of the outputs of an expert system on pest damage in trees (Gast *et al.*, 1988).

Facilities within the current implementation of AKT already provide rudimentary means of achieving this. Diagrams may be appropriate extension material (§5.3.2); these can be printed once entered into AKT. Figure 8.4 shows a diagrammatic representation of relationships in the amelioration of tropical soils by trees. Although the content of this diagram does not provide answers to questions it is proposed that it might provide a suitable prop for discussion between extension workers and farmers.

The creation of mechanisms for generating diagrams from the content of the knowledge base (§5.6) would mean that diagrams customised to particular questions could be developed and printed, possibly through the use of the navigation tool (Appendix B). Such diagrams might simply represent the reltionships between the set of statements of fact in the knowledge base conforming to a particular selection criteria or identified through application of one of the reasoning tools to the content of the knowledge base.

Improvements to the functionality of inference mechanisms such as the navigation tool, to include, for example, approaches to propagating qualitative values through a set of ecological interactions (Guerrin, 1991) and the development of comprehensive knowledge bases, capturing for example, detailed knowledge about causal relationships in the use of trees to ameliorate soils for a range of trees over a range of soils in a range of climates under a range of management regimes would enable the development of diagrams detailing, for example, the predicted consequences of a particular management action. This qualitative approach to modelling to provide a basis for decision support may be more appropriate than more familiar quantitative approaches given that the broad nature of relationships will normally be known before rigorous quantification is possible. Nevertheless, experience suggests that implementing appropriate inference mechanisms and developing knowledge bases that are both sufficiently coherent and comprehensive is unlikely to be a practicable proposition.

Sets of intermediate statements conforming to search criteria and the records of the use of the reasoning tools can be printed, or recorded as text, editted in a word processing package and printed.

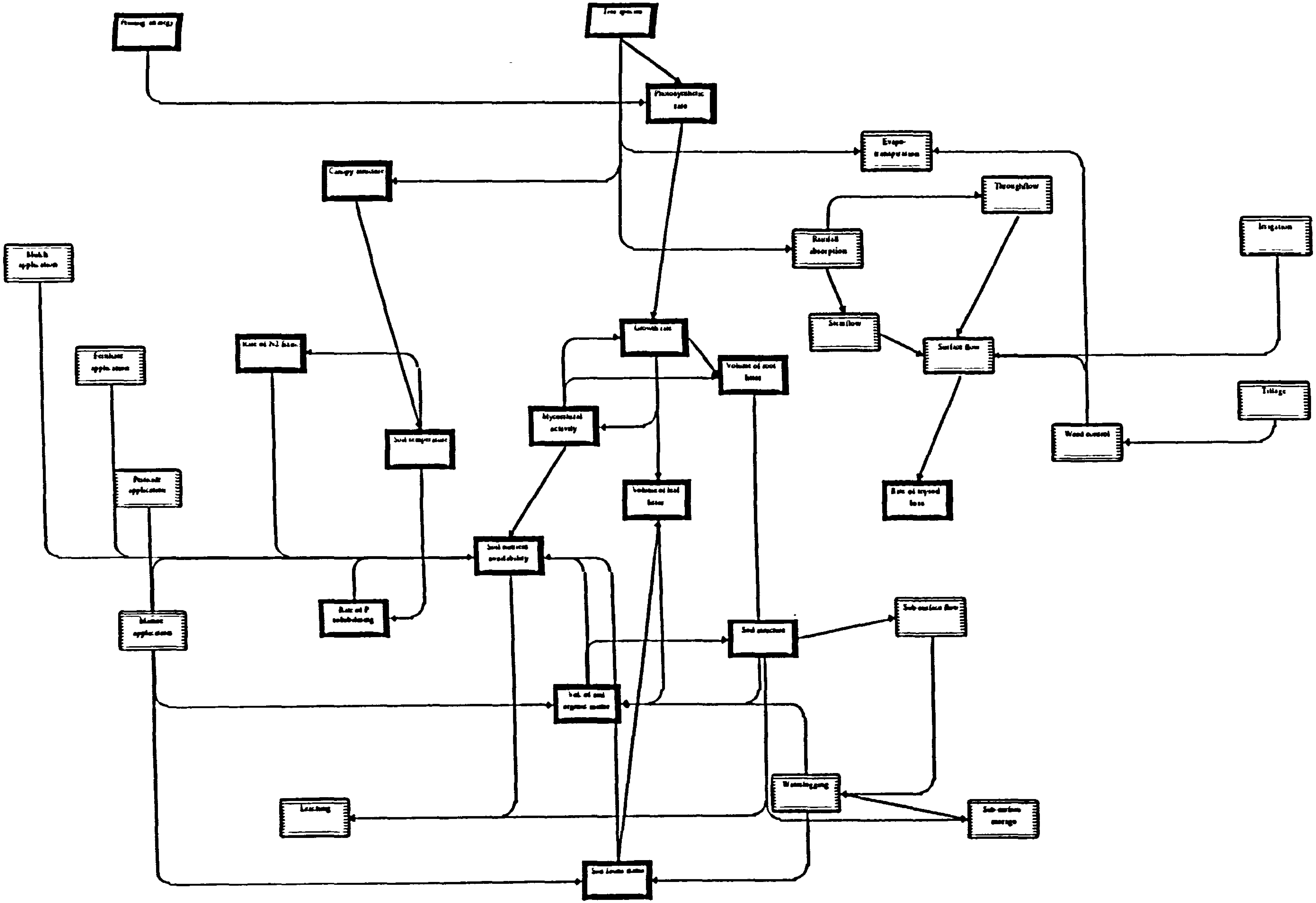


Figure 8.4 A diagrammatic summary of some relationships in the amelioration of tropical soils by trees. Adapted from Furtado *et al.*, 1986.

Mechanisms to automate the development of printed extension material, using standard proforma, could be easily implemented. Tables 8.1, 8.2 and 8.3 illustrate such an application. Table 8.1 shows a proforma for collating all the information on a specified object in the knowledge base into a single report. All the information in this proforma can be derived from formal statements of fact, the hierarchies and statements of definition in the current knowledge base structure. No use is made of intermediate statements of fact. Instead, stylised English statements are generated by fitting the content of formal statements into the proforma. Tables 8.2 and 8.3 illustrate the use of this proforma. In Table 8.2 all the information on a particular tree species recorded in the combined knowledge base (§7.7) is shown (compare with Table 7.3). In Table 8.3 the same proforma is used to collate information about a particular soil type from a knowledge base created by Southern (in preparation). These examples illustrate the flexibility of this simple approach. Furthermore, while the knowledge reported is clearly incomplete and contains ambiguity and inconsistency, it is suggested that output in this form still represents a useful resource.

Table 8.1. A simple proforma for collating information on an object from the content of a formal knowledge base. Terms shown in italics are filled in by refrence to the knowledge base. Information about the name, equivalent terms, definition and hierarchical status are derived from the object hierarchy and the statements of definition. The body of the text is made up of stylised English statements derived from formal statements. The simplest stylised structure for each of the of the four statement tyopes is shown.

OBJECT INFORMATION SHEET

Time Date

Source *:source knowledge base*

Name : *Object*

Equivalents : *Equivalent terms*

Definition : *Statements of definition*

Hierarchical relationships :

Object is a type of Parent, Parent which is a type of Parent 2

Daugther is a type of Object Daughther 2 is a type of Daughter

The *attribute* of *object* is *value* IF *conditions.*

The *atributre* of *object* is *comparison type* than the *attribute* of *object2* IF *conditions.*

object link object 2 IF *conditions.*

A *change* in the *attribute* of *object 1* causes a *change* in the *attribute2* of *object 2* IF *conditions.*

Table 8.2. An object information proforma completed for *Bauhinea purpurea*, a fodder tree species. Knowledge derived from fieldwork by Balaram Thapa (in preparation)

BAUHINIA PURPUREA INFORMATION SHEET

11 February '94 1.17pm

Source : Nepal tree fodder knowledge base

Name : Bauhinia purpurea

Equivalents : Tanki

Definition : ***Statements of definition***

Hierarchical relationships :

Bauhinea purpurea is a type of fodder tree, fodder tree is a type of plant
Buthure tanki is a type of Bauhinea purpurea
Lahare tanki is a type of Bauhinea purpurea

The altitudinal range of Bauhinia purpurea is 0 to 1500 m.
The rainfall range of Bauhinia purpurea is 1000 - 2150 mm
The flowering period of Bauhinia purpurea flowers is Sepember - December
The vigour of the tap roots of Bauhinia purpurea is strong IF
they are seedlings
The light requirement of Bauhinia purpurea is moderate
The frost tolerance of Bauhinia purpurea is frost hardy
The fodder quality of the leaves of Bauhinia purpurea is medium
The altitudinal range of Bauhinia purpurea is up to 800 m
The annual cycle of Bauhinia purpurea is deciduous
The annual cycle of Bauhinia purpurea is evergreen
The leaf size of Bauhinia purpurea is large
The leaf shape of Bauhinia purpurea is bilopped
The orientation of the leaves of Bauhinia purpurea is droopy
The crown size of Bauhinia purpurea is large
The crown shape of Bauhinia purpurea is round
The crown density of Bauhinia purpurea is thick
The size of Bauhinia purpurea is medium
The posiloness of Bauhinia purpurea is posilo
The growth rate of Bauhinia purpurea is rapid IF timing of lopping of Bauhinea purpurea recent
The natural regeneration of Bauhinia purpurea is abundantly IF The location of Bauhinea purpurea is farmland
The timing of leaf drop of Bauhinea purpurea is dry season
The rukhoness of Bauhinia purpurea is malilo

The leaf size of lahare tanki is greater than the leaf size of buthure tanki
The palatibility of lahare tanki is greater than the palatability of buthure tanki
The posiloness of lahare tanki is greater than the posiloness of buthure tanki
The foliage biomass of lahare tanki is greater than the foligae biomass of buthure tanki
The rate of pod production of lahare tanki is greater than the rate of pod production of buthure tanki

Bauhinia purpurea flowers pollinated by bees
Bauhinia purpurea fed to buffalo
Bauhinia purpurea fed to cattle

Table 8.3. The object information proforma completed for kirimati, a soil type identified by farmers in the Kandy district of Sri Lanka. Knowledge derived from fieldwork by Alison Southern (in preparation)

KIRIMATI INFORMATION SHEET

11 February '94

Source *:* Sri Lankan homegarden Knowledge base

Name : Kirimati

Equivalents : White clay

Definition : *Statements of definition*

Hierarchical relationships :

Kirimati is a type of clay

Clay is a type of soil

The colour of Kirimati is greyish
The particle size of Kirimati is very fine
The value of Kirimati is low
The soil location of Kirimati is 15 feet below the surface
The soil location of Kirimati is a few inches below the surface
The soil location of Kirimati is 5-6 feet below the surface
The location of Kirimati is in valleys IF Kirimati deposited by stream
The hardness of Kirimati is hard IF Kirimati dried by sun
The density of Kirimati is dense
The yield of crops is low if crops grown in Kirimati.
The location of Kirimati is not paddy fields
The location of Kirimati is paddy fields
The growth rate of paddy is slow IF paddy grown on kirimati
The colour of Kirimati is black
The colour of Kirimati is white
The colour of Kirimati is reddish
The pohora content of Kirimati is low
The location of Kirimati is moist places
The location of Kirimati is uplands
The stickiness of Kirimati is sticky

The stickness of Kirimati is greater than the stickiness of Mati pasa
The particle size of Mati pasa is greater than the particle size of Kirimati

Kirimati depoisted by springs IF the soil location of Kirimati is 15 feet below the surface AND the location of spring is underground
Kirimati becomes gurugala. IF the time of Kirimati presurrised is long
Kirimati found underneath top soils
Kirimati found underneath Rathu pasa

The examples of proforma in this section could all be automatically generated. Reports in which individual statements of fact are synthesised into a more cohesive statement could be generated interactively by a report editor. While this approach would require a trained operator, it would significantly improve the accessibility of the content of the knowledge base.

8.8. FURTHER DEVELOPMENTS

In the analysis of activities at PAC, identification and prioritisation of research objectives and planning of extension activities were shown to be based on the evaluation of existing knowledge. While these activities may be facilitated by the development of AKT knowledge bases, they are not directly supported by the current software. Software development to explore the provision of a decision-support framework for these activities is merited.

This would demand investigation of the planning processes in relation to research and extension activities to identify detailed decision support tasks and, thereby, develop appropriate inference mechanisms. Tasks involved would include :

- detailed investigation of planning procedure in a sample of development organisations;
- the identification of generic procedural inadequacies;
- the development of a model for optimised planning procedures in relation to an evaluation of current knowledge;
- identification of activities within these procedures for which decision support mechanisms would be appropriate and technically possible; and
- evaluation of the utility of the approaches to knowledge representation currently implemented in AKT and modification accordingly.

So, for example, it may be found that the range of knowledge types that can be represented in the AKT knowledge base need to be expanded if effective decision support for planning activities is to be developed. The AKT knowledge base structure does not currently provide means of storing information about management strategies or records of farmer's problems. This information is widely used in planning research and extension activities, and, is therefore, a central focus in many rapid rural appraisal methodologies. Research into the relationship between statements of problems or management actions and ecological knowledge in order to facilitate the inclusion of this information into a knowledge base, or develop means by which it may be derived from a knowledge base, would be necessary.

8.9 CONCLUSIONS

The research described in this chapter suggests that application of the AKT approach within PAC and similar organisations would support :

- the improvement of institutional mechanisms for identifying and recording the knowledge on defined domains available from the range of sources which might hold relevant information; and
- the improvement of the accessibility of existing knowledge.

This in turn can enable more effective use of existing knowledge, particularly in :

- producing synthesised material for dissemination, including extension;
- providing a sound basis for planning extension activities; and
- providing a sound basis for planning research activities.

While further conceptual and implementational developments have been proposed, these conclusions demonstrate that the application of knowledge-based systems approaches to the formal representation and use of ecological knowledge about agroforestry can facilitate important activities in planning research and extension within agroforestry development programmes.

FURTHER RESEARCH

The AKT software and approach provides an environment for :

- the development of updatable, explicit and comprehensive records of existing knowledge on a defined domain from a defined set of sources; and
- accessing and exploring the contents of knowledge bases developed within the AKT environment.

It has been demonstrated that the approach to knowledge representation in AKT enables the explicit representation of a significant proportion of ecological knowledge about agroforestry, whether from local or scientific sources. Additions to the formal grammar that would enable more faithful representation of a greater proportion of ecological knowledge, for example more effective means of representing temporal knowledge, have been indicated. It has, however, been suggested that such extensions would not be justified without concomitant development of inference mechanisms. Further development of appropriate inference mechanisms and extensions to the formal grammar to enhance its expressiveness may be desirable.

Experience suggests that formal representation does not result in undue distortion of the source knowledge. Nevertheless, further evaluation of the validity of representation achieved by application of the formal grammar would be merited.

Improvements to the functionality of the AKT software to provide more effective support for the user in accessing and exploring the contents of the knowledge base have been indicated in Chapters 7 and 8. Further research to develop means of enabling exploration of the contents of knowledge bases by users with limited skills in computer use is desirable.

It has been demonstrated that the creation of AKT knowledge bases may provide a powerful basis for :

- evaluating the current knowledge held by particular communities on a defined domain; and
- evaluating the current state of scientific knowledge on a defined domain.

Case study evaluation in relation to the activities of Pakhribas Agricultural Centre has demonstrated that these activities may be useful in :

- facilitating the identification of adaptive research priorities for agroforestry;

- planning extension activities; and

- producing appropriate extension materials.

The use of AKT knowledge bases in evaluating the current state of knowledge on a defined domain demands that the representativeness of those knowledge bases of the knowledge held by the source communities can be demonstrated. Some of the factors influencing representativeness have been discussed (§7.6) and means of evaluating the representativeness of a knowledge base are under development (Thapa, in preparation). Nevertheless, further work in this context is necessary.

The AKT software provides a resource for use in the evaluation of the current state of knowledge but only limited decision support in the process of evaluation. Some of the tasks involved in the evaluation of current knowledge in a development context have been indicated in Chapter 8. Further detailed analysis of the processes involved, particularly in planning research and extension activities is merited. This would provide a basis for the specification and implementation, where appropriate, of decision support functionality with the AKT software to facilitate these processes.

The AKT approach has also been proposed as providing a means of synthesising a more useful understanding of the ecology of agroforestry practices through the combination of knowledge from local, scientific and professional sources. While evaluation of the approach has indicated that it does provide an environment in which local and scientific knowledge can be combined effectively, it remains to be demonstrated that this process is useful in development context. Research to compare the utility of local or scientific knowledge alone and the combination of the two in planning research, planning extension and in the management of agroforestry practices is merited.

Finally, an evaluation of the efficacy, efficiency and practicability of the AKT approach in comparison to existing approaches to accessing, managing and using knowledge about the ecology of agroforestry practices is demanded if the approach is to be considered for practical implementation in a development context.

APPENDIX A CHRONOLOGICAL ACCOUNT OF SOFTWARE DEVELOPMENT

The structure and function of software specification implemented in the AKT software resulted from experience in the creation of knowledge acquisition and manipulation software prototypes. This appendix details the development of, and relationship between, these prototypes.

Preliminary demonstration of the utility of representing and reasoning with local knowledge about agroforestry using a knowledge based system approach was undertaken through the development of LOCUS in January 1991. LOCUS was implemented in LPA MacProlog by Robert Muetzelfeldt.

Prototype software development started with the creation of two tools, the Transcript and Template tools (Haggith *et al*, 1992) implemented in HyperCard (Goodman, 1990). Experience in the use of these tools in the support of knowledge acquisition activities resulted in a substantial modification of specifications for intermediate representation (Chapter 3).

New software was demanded to support the exploration of these new specifications. As a result, a rudimentary knowledge base structure was developed in a standard word processing facility. Individual statements of fact about species interactions in multi-layered forest gardens (Southern, in preparation), each tagged according to source, were entered into an appropriate position in the knowledge base, structured under species headings, with subcategories (for example 'pests' and 'light requirements') within each species heading. Using this approach, a knowledge base of 3 000 sentences on 180 species was successfully created during fieldwork in Sri Lanka (Walker and Southern, 1992).

A standard word processor proved to be a successful prototyping environment but did not provide facilities for the data search and retrieval mechanisms required. In particular, means were needed for sorting statements according to identified keywords and their synonyms. However, this trial provided a detailed specification for a further software prototype, TEAK (Tools for Eliciting Agroforestry Knowledge), implemented in HyperCard (Haggith *et al.*, 1992). HyperCard is a very 'user-friendly' environment for the development of particular applications (Goodman, 1990) and provides facilities for the development of appropriate search and retrieval features.

TEAK was based on the representation of knowledge as individual statements of fact. Each statement of fact was recorded on a separate card. The statement was recorded informally with no predefined structure. Associated conditional information and, in later versions, explanatory information was separately recorded on the same card. In earlier versions the opportunity to record notes about the statement was included and an opportunity to include formal statements based on template sentence structures was included. Contextual information was linked to each sentence card. This provided information on the informant and on the circumstances of the particular interview from which the statement arose. Informant and interview details were stored on separate cards.

In addition to this basic facility for recording sentences, TEAK included a keywords facility. Keywords within statements were identified by the user and stored as a separate, hierarchically structured list. The keywords provided a basis for keyword searches across the set of statements with Boolean combinations of keywords.

TEAK was developed between April and September 1992. It was implemented in HyperCard (Goodman, 1990) by Mandy Haggith of the Department of Artificial Intelligence, University of Edinburgh (Haggith *et al.*, 1992).

HyperCard and its associated scripting language, HyperTalk, provide a rapid prototyping environment in which software can be developed, applied and adapted with minimal pre-design. This proved to be an effective vehicle for the early stages of programming. HyperCard does not, however, offer the breadth of facilities for the manipulation of knowledge and reasoning possible with Prolog. Interfacing between the two languages is possible but cumbersome. As a result, HyperCard was used as a prototyping environment, but was not suitable for full software implementation.

TEAK was upgraded and reimplemented in LPA MacProlog 4.5 between October 1992 and January 1993 by Gill Kendon of the Department of Artificial Intelligence.

The approach developed in TEAK provided a straightforward means of entering knowledge and of iterative assessment of that knowledge by a researcher.

Entering informal statements of fact into TEAK proved to be intuitive for postgraduate researchers from developing countries and compatible with fieldwork. During fieldwork in Sri Lanka (Walker and Southern, 1992), it was found that it took from one-and-a-half to two hours to identify and record statements of fact from 45 minute taped interviews. Fieldwork in Nepal (Thapa and Walker, 1992) using TEAK corroborated this.

The fundamental criticism of an interface operating with individual sentences is a lack of overview. Given the identified need for a goal-orientated approach to knowledge elicitation (§7.2.3), the lack of facilities to support the structuring of further elicitation on the basis of assessment of previously recorded knowledge was a serious limitation. This limitation is illustrated by the contrasting success of a diagramming approach in identifying gaps in knowledge already elicited (Thapa and Walker, 1992).

This problem was addressed in TEAK, by providing a means of continually assessing the existing contents of the knowledge base through keyword searches. An ability to select and review sets of sentences according to keywords, and the questions facility (Haggith *et al.*, 1992) mitigated these limitations to a significant degree. Nevertheless, this lack of overview and knowledge synthesis remained a fundamental constraint in the effective use of interfaces for entering individual and disaggregated statements of fact.

TEAK proved to be a useful means of driving fieldwork (Thapa and Walker, 1992), both in terms of the task it represented and the implementation of that task. However, experience has shown that use of TEAK in supporting knowledge acquisition during fieldwork tended to result in knowledge bases containing significant inconsistency, incoherence, repetition and ambiguity introduced by the knowledge base developer through poor management of the growing set of statements. Such knowledge bases were of limited utility. In the same way, a considerable proportion of these problems were shown to relate to the particular purposes for which the software was being used, and the manner in which it was being used. Nevertheless it was argued that implementing the task in the software in a more guided and restricted fashion would go some way towards ameliorating these problems.

In March 1992 Robert Muetzelfeldt undertook prototype development of a diagramming interface for knowledge representation. This prototype was developed in LPA MacProlog and provided a drawing environment in which nodes could be placed on the screen and named and then linked by unlabelled links. This prototype was used as a basis for the development of HyperNet, a diagram based knowledge acquisition tool. HyperNet was implemented by the author in LPA MacProlog between March 1992 and January 1993 (Walker *et al.*, 1993).

In February 1993 Gill Kendon and the author combined the Prolog version of TEAK and HyperNet to create an integrated package called AKT1 (Chapter 6). The integrated toolkit was then debugged and has been in use since April 1993. In September 1993, substantial new functionality was added to this toolkit by Gill Kendon to result in AKT2 (Appendix B), which, thereby, superseded AKT1.

APPENDIX B AKT 2

Between September and December 1993, significant functionality was added to the software described Chapter 6, resulting in AKT2. Specification and implementation was undertaken by Gill Kendon of the Department of Artificial Intelligence of the University of Edinburgh. As a result, description of this process is not undertaken in this thesis.

However, evaluation of the AKT approach in facilitating the development of useful knowledge bases in Chapter 7 includes consideration of the reasoning tasks implemented in AKT2. For this reason, this addendum provides a brief description of that functionality.

AKT2 operates in one of two modes, an acquisition mode and a reasoning mode. The software can only be run in one of the two modes at any one time, each having completely different menus.

In spite of some changes to menu structure and menu item names, the acquisition mode of AKT2 corresponds, in functional terms, exactly to AKT1. All the additional function in AKT2 is in the reasoning mode.

In reasoning mode, the AKT 2 software works with modified versions of the knowledge bases created in the acquisition mode. This modification essentially involves abstracting formal statements and the sort hierarchies from the full knowledge base through the creation of 'extracts'. Extracts may be made from the entire knowledge base or from some subset of the knowledge base, created through use of the selection facilities (§6.5.6). When created, extracts are saved as files, and are therefore autonomous of the knowledge base from which they were created. During the reasoning process, formal statements can be temporarily added to the extract. These can be added in order to provide information demanded by conditions in other statements or to explore the implications of adding knowledge on the reasoning process. They cannot, however, be saved. Furthermore, the statements in the extract cannot be edited. This ensures that the extract remains an abstraction from a full knowledge base rather than diverging from that knowledge base.

In the reasoning mode, AKT2 has four menus. The File menu is analogous to the File menu in the acquisition mode, containing facilities for opening, creating, closing and saving extracts and defaults for searches (Chapter 6). As well as 'Quit', the File menu contains a 'Change mode' option, which can be used to change from the reasoning mode to the acquisition mode, closing the current extract in the process. The Edit menu is precisely the same as in the acquisition mode and AKT1. The View menu has four items 'Browser', 'View statements', 'View log', and 'Extract'. Browser allows the user to see the content of the Browser (object sort hierarchies and glossaries). View statements shows the statements in the extract: one of three options can be selected such that the complete set, only the original statements or only the added statements are shown. View log allows the user to review a log of reasoning activities. View extract shows the entire content of the extract, i.e. statements and sort hierarchies.

The Tools menu currently contains two options, the 'Navigation' and 'Query' tools.

(i) Navigation tool

The navigation tool allows the user to move across the links between statements of fact. As such it is similar to the causal reasoning tool (§4.3). Three types of such linkage are captured in the formal language (causal, comparison and the generic 'link'). A navigation tool can be used for all three.

The navigation tool allows an intuitive exploration of the knowledge base. For causal statements, for example, it allows access to known causes or implications of a particular event. The cause or implications of any one of these causes or implications may then in turn be explored. Linkages may be conditional. Conditionality may be satisfied internally by reference to other statements in the knowledge base or through interaction with the user (§2.3). The interface for the navigation tool is shown in Figure B.1(a).

(ii) Query tool

The query tool takes advantage of the basic pattern matching undertaken by Prolog. A formal statement may contain a mixture of atoms and variables. Prolog can be used to identify the set of statements in the knowledge base that match the atoms specified and give the atoms in those matched statements for the variables in the query.

The query tool fulfils more than a simple data retrieval function as it can make use of reasoning in addressing a query. For example it may respond to the query :

    Does *Alnus nepalensis* fix nitrogen?

Through use of the statement :

    Alders are nitrogen fixing species

and the information, from the sort hierarchy, that *Alnus nepalensis* is an alder species.

The interface for the query tool is shown in Figure B.1(b).

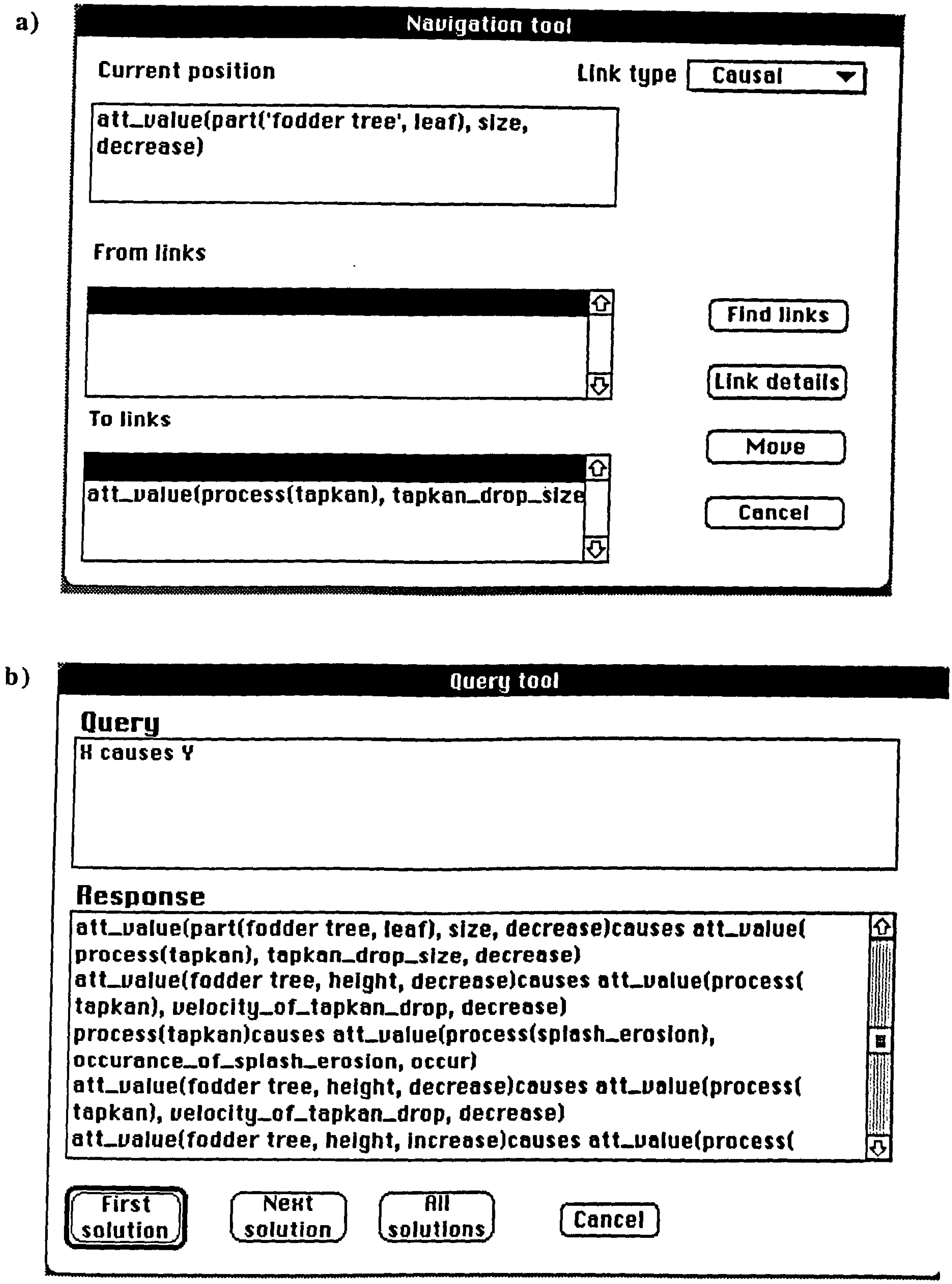


Figure B.1. The interfaces for the navigation (a) and query (b) tools

APPENDIX C - NODE-AND-LINK DIAGRAMS USED IN §5.3.1

A generalised diagram of crop evolution. (Simmonds, 1979)

A model of an upland Javanese farming system. (Van der Poel and Van Dijk, 1987))

Agroforestry decision tree for preliminary matching of prototype technologies to land use systems, in the humid tropics, based on minimal knowledge of three key diagnostic variables and an evolutionary perspective on the role of agroforestry practices in the intensification of smallholder systems in the tropics. (Raintree, 1987)

An influence diagram relating forest growth to wind speed. (Matzuaki, 1988)

Causal diagram of the Dutch agricultural system. (Van Mansvelt, 1988)

Causal relationships drawn by farmers for cogon problem. (Lightfoot *et al.*, 1989)

Cycle of land degradation in Kisii District, Kenya. (Duchhart et al. 1989)

Energy inputs and outputs and material flows in a Wiltshire agricultural system. (Bayliss-Smith, 1982)

Forest and lake ecosystems linked with wood using industry. (Tamm, 1975)

Integrated research and training in agroforestry networking. (Zulberti, 1987)

Plant and animal interrelationships in grazing systems (Maxwell, 1990)

Proposed causal relationships to explain the results of experimentally reducing the population density of *Corvus corone* in a spruce woodland, by shooting. (Slagsvold, 1980)

Removing, replacing and supplementing plant nutrients. (Jollans, 1985)

Rural water balance sub-model for forest and agricultural lands. (Howarth et al, 1991)

Seven types of agrarian society, and their evolutionary relationships. (Bayliss-Smith, 1982)

Some factors that influence whether *Acacia mearnsii* is grown. (Berenscot *et al.*, 1988)

Structure of the Doyle *et al.*, ash/pasture model. (Doyle, 1991)

Structure of the SCUAF carbon model. (Young, 1989)

Summary influence diagram for the RESCAP model. (Monteith *et al.*, 1992)

The food chain system - proximate relative values of product flow (Street, 1990)

The general pattern of mainstream plant-breeding. (Simmonds 1979)

The impact of deforestation on the hydrology, microclimate , biogeochemistry, and the erosion and transport of particulate matter during the first 2 water-years after cutting (Borman & Likens, 1979)

The lumped Stanford IV catchment model. (Tamm., 1975)

The nitrogen cycle at Rothampstead, Hertfordshire, England for the top 22 cm of a cultivated field. (Bayliss-Smith, 1982)

The nitrogen cycle is an example of a biogeochemical cycle with a large gaseous reservoir and many feedback mechanisms. (Odum, 1963)

The plant breeder's decisions on objectives and methods. (Simmonds, 1979)

APPENDIX D - TIMING AND SEQUENCE OF RESEARCH ACTIVITIES

The research described in this thesis was undertaken as part of a large, inter-disciplinary research initiative (see Acknowledgements). The work described was undertaken between April 1990 and March 1994.

The research programme involved methodological specification, software implementation, methodological application and evaluation. These activities were intimately dependent upon each other given the intention of creating a practically applicable methodology. The author was involved in all of these activities. Some software development, notably the development of the diagramming interface, was undertaken solely by the author. The author also undertook a total of four months work in the field in Sri Lanka, Nepal and Thailand as well as spending a significant amount of time supporting the creation of knowledge bases by research students after the completion of their fieldwork. However, the author's principal responsibility was in specification and evaluation, which represent the major contribution of this thesis. The processes of specification and evaluation provided the critical linkage between methodological application and software development. Figure D1 summarises the major events in the research and indicates the central role of specification in the research process.

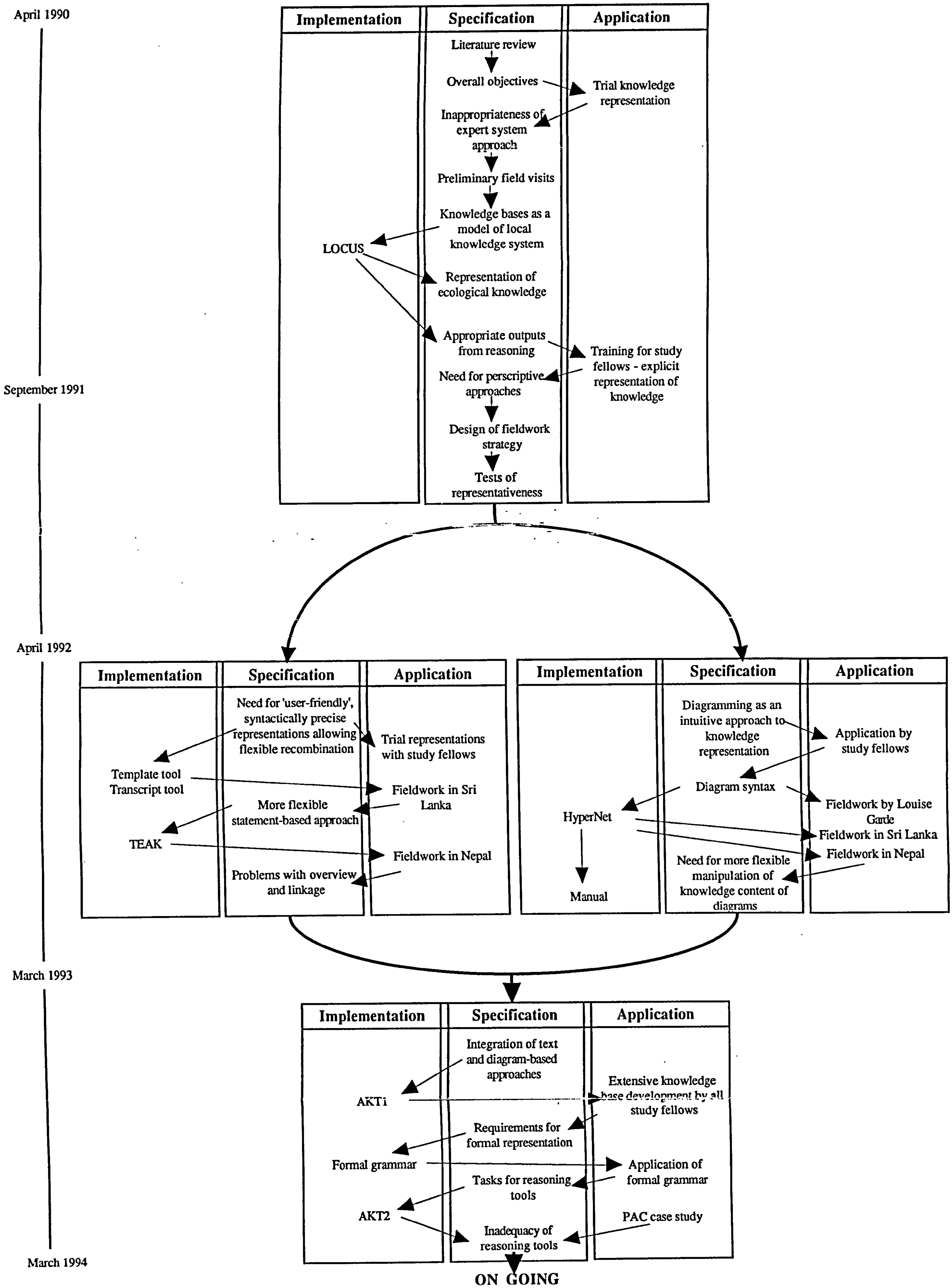


Figure D1 Timing and sequence of research activities

REFERENCES

Agroforestry News. Published by the Agroforestry Research Trust, 17 Arden Drive, Chelston, Torquay, Devon TQ2 6DZ, U.K.

Allen, T.F.H. and Starr, T.B. 1982 *Hierarchy: perspectives for ecological theory*. Chigaco : Chigaco University Press.

Allen, T.F.H., O'Neill, R.V. 1991 Improving predictability in networks : system specification through hierarchy theory. in Higashi, M., Burns, T.P. (eds) *Theoretical studies in ecosystems, the network perspective*. Cambridge University Press.

Amonor, K. 1989. *340 abstracts on farmer participatory research*. Network Paper, Agricultural Administration Research and Extension Network No 5. Agricultural Administration Unit, Overseas Development Institute, London, UK.

Anderson, L.S. and Sinclair, F.L. 1993. Ecological interactions in agroforestry systems. *Agroforestry Abstracts*, 6 (4): 207-247.

Anderson, L.S., Muetzelfeldt, R.I., and Sinclair, F.L. (1993) An integrated research strategy for modelling and experiemental agroforestry. *Commonwealth Forestry Review*, 72(3): 166-174

Balachardran, C.G., Fisher, P.R., and Stanley, M.A. 1989. An expert system approach to rural development. *The Journal of Developing Areas*, 23: 259 - 270.

Barker, D., Oguntoyinbo, J.S. and Richards, P. 1977. *The utility of the Nigerian peasant farmer's knowledge in monitoring of agricultural resources*. London : Monitoring and Assessment Centre, Chelsea College General Report Series, 4.

Barlett, Peggy F. (ed). 1980 *Agricultural Decision Making: anthropological contributions to rural development*. New York Academic Press.

Barrett, J.R. & Jones, D.D. (eds). 1989. *Knowledge engineering in agriculture.American Society of Agricultural Engineers*. Monograph No 8, St Joesph Michigan USA.

Bayliss-Smith, T.P. 1981. *The ecology of agricultural systems*. Cambridge University Press, Cambridge, 112p.

Beer, J., Borel, R. & Bonnemann, A. 1990. On-farm agroforestry research planning in Costa Rica. In : Budd, W.W., Duchhart,I., Hardesty, L.H. and Steiner, F. (eds), Planning for Agroforestry. Developments in Landscape Management and Urban Planning, Elsevier Science Publishing Co. Inc. Amsterdam, pp 58-78.

Bell, J. and Hardiman, R.J. 1989. The Third Role - the naturalistic knowledge engineer. In: Diaper, D. (ed), *Knowledge elicitation : Principles, techniques and applications*. Ellis Horwood Ltd. Chichester.

Benfer, R.A. and Furbee, L. 1990. Knowledge acquisition: lessons from anthropology. *AI Applications in Resource Management*, 4 (3): 19-26.

Berenschot, L.M., Filius, B.M., and Hardjosoediro, S. 1988. Factors determing the occurence of agroforestry systems with *Acacia mearnsii* in Central Java. *Agroforestry Systems*, 6: 119-155.

Berlin, B. 1973. Folk Systematics in Relation to Biological Classification and Nomenclature. *Annual Review of Ecology and Systematics*, 4: 259-271.

Berlin, B. 1978. Ethnobiological Classification. In: Rosch, E. and Lloyd, B.B. (eds). *Cognition and Categorisation*. New York : Lawrence Erbaum Associates, pp 9-28.

Berlin, B., Breedlove, D. and Raven, P. 1968. Covert Categories and Folk Taxonomies. *American Anthropologist*, 70: 290-299.

Biggs S.D. 1985. A farming systems approach : some unanswered questions. *Agricultural administration*, **18**: 1-12.

Binswanger, H.P. 1980 Attitudes towards risk : experimental measurement in rural India. *American Journal of Agricultural Economics*, 62: 395-407.

Bormann, H. and Likens, G.E. 1979. *Pattern and process in forested ecosystems : disturbances, development, and the steady state based on the Hubbard Brook ecosystems study.* New York, Springer-Verlag.

Bratko, I. 1990. *Prolog. Programming for artificial intelligence.* Second edition, Addison Wesley Publishing Company.

Brokensha, D., Warren, D.M., Werner, O. (eds). 1980. *Indigenous knowledge systems and development.* University Press of America.

Buck, L. 1990. Planning agroforestry extension projects, the care international approach in Kenya. In : Budd, W.W., Duchhart, I., Hardesty, L.H. and F. Steiner (eds), *Planning for Agroforestry.* Developments in landscape management and urban planning, Elsevier Science Publishing Co. Inc. Amsterdam, pp 58-78.

Budd, W.W., Hardesty, L. and Hinds, W.T. 1990. A conspectus for Agroforestry. In: Budd, W.W., Duchhart, I., Hardesty, L.H. and Steiner, F. *Planning for agroforestry.* Developments in landscape management and urban planning, 6c. Elsevier Science Publishing Co. Inc. Amsterdam. pp 321-335

Bunderson, T.W., Wakeel, A. el, Saad, Z. & Hashim, I. 1990. Agroforestry practices and potentials in western Sudan. In: Budd, W.W., Duchhart, I., Hardesty, L.H. and Steiner, F. *Planning for agroforestry.* Developments in landscape management and urban planning, 6c. Elsevier Science Publishing Co. Inc. Amsterdam, pp 58-78.

Burns, T.P., Patten, B, Higashi, M. 1991 Hierarchical evolution in ecological networks : environs and selection. in Higashi, M., Burns, T.P. 1991 *Theoretical studies in ecosystems, the network perspective.* Cambridge University Press.

Byerlee, D., Harrington, and Winklemen. 1982. Farming systems research : issues in research strategy and technology design. *Amercian Journal of Agricultural Economics*, **64**: 877-904.

Carruthers and Chambers 1981. Rapid rural appraisal for development. *Agricultural Administration*, **8**: 433-450.

Carter A.S, & Gilmour, D. 1989. Increase in tree cover on private farmland in central Nepal. *Mountain Research and Development*, 9(4): 381-391.

Chambers, R., Pacey, A. and Thrupp, L.A. (eds). 1989. *Farmer First: Farmer innovation and agricultural research.* Intermediate Technology Publications, London.

Chand, S.P., and Gibbon, D. 1990 . *Samuhik Bhraman : A rapid and appropriate method of prioritising and replanning agricultural research in Nepal.* Technical Paper 121, PAC c/o BAPSO, Kathmandu, Nepal.

Chand, S.P., Gurung, B.D. and Rood, P. 1990. *Farmers' traditional wisdom : where does it stand within the present agricultural research system of Nepal?* Pakhribas Agricultural Centre, Nepal, Occasional Paper No. 4, PAC, BAPSO, PO Box 106, Kathmandu, Nepal.

Chandler, P. 1990. *The indigenous knowledge of ecological processes amongst peasants in Fujian, the People's Republic of China.* PhD Dissertation, College of Forest Resources, University of Washington, Seattle, Washington, USA.

Chandler, P. 1991. The indigenous knowledge of ecological processes among peasants in the People's Republic of China. *Agriculture and Human Values*, VIII (1 and 2): 59 - 67.

Conklin, H.C. 1971. *An ethnoecological approach to shifting cultivation.* Bobbs-Merrill, pp 133-42.

Conway, G.R. 1987. The properties of agroecosystems. *Agricultural Systems*, 24: 95-117.

Conway, G.R. 1989. Diagrams for farmers, in Chambers, R., Pacey, A., Thrupp, L.A. (eds), *Farmer First: Farmer Innovation and Agricultural Research.* Intermediate Technology Publications.

Conway, G.R. 1990. Agroecosystems. pp. 205-234. In: Jones, J.G.W. and Street, P.R. (eds), *Systems theory applied to agriculture and the food chain.* Elsevier Science Publishing Co. Inc. Amsterdam.

Cordingley, E., Betsy, S. 1989. Knowledge elicitation techniques for knowledge based systems. In Diaper, D. (ed). *Knowledge elicitation : Principles, techniques and applications.* Ellis Horwood Ltd. Chichester.

d'Anrade, R.G. 1970. *Culture shared and unique.* Paper read at meeting of the American Anthropological Association.

da Costa, M.C.R. 1993. *Women's ecological knowledge of tree-crop interactions in Shimnyanga Region, Tanzania.* Thesis submitted in partial fulfilment for the degree of MSc in Rural Resource Management to the School of Agricultural and Forest Sciences of the University of Wales, Bangor.

Davies, J.J. 1968. On the scientific method, how scientists work. Longmans 166 p.

Dent, J.B. and Blackie, M.J. 1979 . *Systems simulation in agriculture.* Applied Science publishers Ltd. London.

Diaper, D. (ed). 1989. *Knowledge elicitation : Principles, techniques and applications.* Ellis Horwood Ltd. Chichester.

Dobbert, M.L. 1982. *Ethnographic research: Theory and application for modern schools and societies.* Praeger.

Doukidis, G. I., Frank L., Miller,G. (eds). 1985. *Knowledge based management support systems.* Ellis Horwood Books in Information Technology.

Dove, M.R. 1985. *Swidden agriculture in Indonesia : the subsistence stategies of the Kalimantan Kantu.* Mouton, Berlin, New York, Amsterdam.

Doyle, C.J. 1991 Modelling the potential of agroforestry in lowland Britain. *Agricultural Systems and Information Technology Newsletter* 3: 3-4.

Duchhart, I., Stenier, F., Basman, J.H. 1989. Planning methods for agroforestry. *Agroforestry Systems*, 7: 227-258.

Ellen, R.F. 1984. *Ethnographic research : a guide to general conduct.* Academic Press.

English, W., Engelbart, D. and Berman, M. 1977. Display-selection techniques for text manipulation. 8 (1): 21-31.

Furtado, J.I., Prinsley, R.T. and Swift, M.J. 1986 Amelioration of soil by trees in the tropics. In: Prinsley, R.T. and Swift, M.J. (eds) *Amelioration of soil by trees, a review of current concepts and practices.* Commonwealth Science Council, London.

Garde, L. (1992). A study of indigenous ecological knowledge about a livestock / forest production system in the Uttar Pradesh hills of India. MSc Dissertation, University of Wales, Bangor.

Gast, S.J., Stock, M. and Livingston, L. 1988. Customised document generation for forestry extension. *AI Applications,* 2(1): 53-56.

Getahun, A. 1990. Agroforestry for development in Kenya: an overview. In Budd, W.W., Duchhart,I., Hardesty, L.H. and Steiner, F. *Planning for agroforestry.* Developments in landscape management and urban planning :. Elsevier Science Publishing Co. Inc. Amsterdam. pp 183-202.

Gholz, H.L. 1987. *Agroforestry : realities, possibilities and potentials.* Matinus Nijhoff Publishers, Dordrecht.

Gibbon, D. 1993. *Outreach Research Consultancy Mission - PAC.* Koshi Hills Eastern Nepal Overseas Development Group, School of Development Studies, University of East Anglia, Norwich.

Gibbon, D., Schultz, M. 1989. *Agricultural Systems in the Eastern Hills of Nepal: Present Situation and Opportunities for Innovative Research and Extension .* Pakhribas Agricultural Centre, c/o BAPSO, PO Box 106, Kathmandu, Nepal Technical Paper 108.

Giere, R.N. 1984. *Understanding scientific reasoning.* Second ediotion. Holt, Rhinehart and Winston, New York.

Gladwin, C. 1980. Cognitive strategies and adoption decisions: a case study of non adoption of an agronomic recommendation. In: Brokensha, D., Warren, D.M. and Werner, O. (eds), *Indigenous knowledge systems and development.* University Press of America.

Goldberg, A . and Robson, D. 1979. *A metaphor for user-interface design.* Proceedings of the Twelfth Hawaii International Conference on System Sciences, 6(1): 148-157.

Goodman, D. 1990. *The complete Hypercard 2.0 handbook.* Bantam books, NY, USA.

Gorden, R.L. 1969. *Interviewing : strategy, techniques and tactics.* Revised by Williams, R.M. The Dorsey Press. Horewood, Illinios.

Guerrin, F. 1991. Qualitative reasoning about ecological processes: interpretation in ecology. *Ecological modelling* 59: 165-201.

Guillet, D. (ed). 1989a. Expert systems applications in anthropology: Part 1. *Anthropological Quaterly* 62: 57-102.

Guillet, D. (ed). 1989b. Expert systems applications in anthropology: Part 2. *Anthropological Quaterly* 62: 107-144.

Haggith, M., Robertson, D., Walker, D.H., Sinclair, F.L. and Muetzelfedlt, R.I. 1992. *TEAK : tools for eliciting agroforestry knowledge.* Proceedings of the British Computer Society Symposium on IT-enables change in developing countries.

Hart, A. 1986. *Knowledge acquistion for expert systems.* London, Kogan Page

Hayes-Roth, F., Waterman, D. and Lenat, D. (eds). 1983. *Building Expert Systems.* Reading, Mass. Addison-Wesley.

Hildebrand P.E. 1982. Farming systems research : issues in research strategy and technology design. *American Journal of Agricultural Economics,* 64: 877-904.

Hildebrand, P. E. 1990. Farming systems research - extension. In: Jones, J.G.W.. & Street, P.R. (eds), *Systems theory applied to agriculture and the food chain .* Elsevier Science Publishing Co. Inc. Amsterdam.

Hitinayake, H.M.G.S.B. (in preparation) Comparative evaluation of the use of knowledge based systemns and farming systems research and extension (FSR/E) methods in agroforestry development. PhD Thesis, University of Wales, Bangor.

Holmes, M. L. 1993. *Indigenous knowledge of Sukuma women, with particular reference to soils and crops.* Thesis submitted in partial fulfilment for the degree of MSc in Rural Resource Management to the School of Agricultural and Forest Sciences of the University of Wales, Bangor.

Howarth R.W., Fruci, J. R. and Sherman, D. 1991. Inputs of sediment and carbon to an estuarine ecosystem - influence of land use. Ecological applications, 1 (1): 27 - 39.

Howes, M. and Chambers, R. 1980. Indigenous technical knowledge: analysis, implications and issues. In: Brokensha, D., Warren, D.M. and Werner, O. 1980. *Indigenous knowledge systems and development.* University Press of America.

Hunn, E. 1975. A measure of the degree of correspondance of folk to scientific biological classification. *American Ethnologist,* 2: 309-327.

IDS Workshop 1989 . Interactions for local innovation. In: Chambers, R., Pacey, A. and Thrupp, L.A. (eds), *Farmer first, farmer innovation and agricultural research.* Intermediate Technology Publications Limited, London.

IDS Workshop. 1989 Farmers' knowledge, innovations and relation to science. In: Chambers, R., Pacey, A. and Thrupp, L.A. (eds), *Farmer first, farmer innovation and agricultural research.* Intermediate Technology Publications Limited, London.

Jackson, P. 1986. *Introduction to Expert Systems.* Addison Wesley Publishing Company.

Jinadasa, N., (in preparation). Far,ers' indigenous ecological knowledge about the selection of plant ideotypes and breeding objectives in Kandy Home Gardens in Sri Lanka. MPhil Thesis, University of Wales, Bangor.

Johns, N. 1991. *Manual, MacProlog version 4.0.* Logic Programming Associates Ltd., London SW18 3SX

Johns, N. 1992 *MacProlog 4.5 Environment Guide.* Logic Programming Associates Ltd. Studio 4, Royal Victoria Patriotic Building, Trinity Road, London sw18 3SX, England.

Johnson, A. W. 1972. Individuality and Experimentation in Traditional Agriculture, *Human Ecology* 1 : 149-159.

Johnson, N. E. 1989. Mediating representations. In: knowledge elicitation. in Diaper, D (ed), *Knowledge elicitation : Principles, techniques and applications.* Ellis Horwood Ltd. Chichester.

Jollans, J.L. 1985. *Fertilisers in U.K. farming.* Reading Centre for Agricultral Strategy Report No. 9.

Jones, G. W., Street, P. R. (eds). 1990 *Systems Theory Applied to Agriculture and the Food Chain.* Elsevier Science Publishing Co. Inc. Amsterdam.

Jones, P. 1989. Agricultural applications in expert systems. *Agricultural Systems* 31: 3 - 18.

Kang, B.T. and Wilson, G.F. 1987. The development of alley cropping as a promising agroforestry technology. in Steppler, H.A. and Nair, P.K.R (eds). *Agroforestry - a decade of development.* ICRAF, Nairobi.

Kay, A. and Goldberg, A. 1977. *Personal dynamic media.* Computer, March, 1977.

Kilahama, F.B. (in preparation). Trees and indigenous ecological knowledge about agroforestry practices in the rangeland of Kishapu and Negezi dividsions, Shinyanga Region, Tanzania. PhD Thesis, University of Wales, Bangor.

Knight, G.G. 1980. Ethnoscience and the African farmer: rational and strategy Tanzania. In: Brokensha, D., Warren, D.M. and Werner, O. (eds). *Indigenous knowledge systems and development.* University Press of America.

Kronick, J. 1984. Temporal analysis of agroforestry systems for rural development. *Agroforestry systems*, 2: 165-176.

Kumar, K. 1987. *Conducting group interviews in developing countries.* AID Programme Design and Evaluation Methodology Report, No.8, AID, Washington DC.

Kunwar, B.B., and K.C., Sharan. 1993. *Lessons learned from Outreach research activities: An experience of Pakhribas Agricultural Centre.* Paper presented at the Second Regional Outreach Workshop, Pakhribas Agricultural Centre, March 1993.

Lewis, J. 1986. Stella : a model of its kind. *Practical computing*, 66-67.

Lightfoot, C., De Guia, O., Aliman, A., Ocado, F. 1989. Systems diagrams to help farmers decide in on-farm research. In: Chambers, R., Pacey, A. and Thrupp, L.A. (eds), *Farmer first, farmer innovation and agricultural research.* Intermediate Technology Publications Limited, London.

Lundgren, B.O. 1987. Institutional aspects of agroforestry research and development. In Steppler, H.A. and Nair, P.K.R. (eds), *Agroforestry : a decade of development.* ICRAF, Nairobi.

MacDicken K.G., Vergara, N.T. 1990. Introduction to agroforestry. In MacDicken K.G., Vergara, N.T. (eds), *Agroforestry, classification and management.* Wiley Interscience Publication, John Wiley and Sons

MacDicken, K.G. 1990. Agroforestry management in the humiud tropics. In: MacDicken, K.G. and Vergara, N.T. (eds), *Agroforestry : Classification and Management.* John Wiley and Sons. New York.

Marten, G.G. 1988. Productivity, stability, sustainability, equitability and autonomy as properties of agroecosystem assessment. *Agricultural Systems*, 26: 291-316.

Mathias-Mundy, E., Muchena, O., McKiernan, G. and Mundy, P., 1991. *Indigenous technical knowledge of private tree management.* Center for Indigenous Knowledge and Rural Development CIKARD Iowa State University, Ames, Iowa, USA.

Matzuaki, T. 1988. Growth of the windbreak forest in the coastal plain. In: *Modelling Forest Growth Processes.* Proceedings of the Second Uppsala Modelling Workshop, 29th August - 2nd Septrember 1988, Uppsala, Sweden.

Maxwell, T.J. 1990. Plant-animal interactions in northern temperate sown grasslands and semi-natural vegetation, In: Jones, J. G. W. and Street, P.R. (eds), *Systems theory applied to agriculture and the food chain.* Elsevier Science Publishing Co. Inc. Amsterdam, pp23-60.

McCracken, J.A., Pretty, J.N. and Conway, G. 1988. *An introduction to Rapid Rural Appraisal for agricultural development*, International Institute for Environment and Development, London.

McGraw, K. L. 1989. *Knowledge acquisition : principles and guidelines.* Englewood Cliffs Prentice Hall International.

McKinion, J.M., Baker, D.N., Whistler, F.D. and Lambert, J.R. 1989. Application of the Gossym/Comay system of cotton crop management. *Agricultural systems*, 31: 55-65.

McRoberts, R.E., D.L. Schmoldt, H.M. Rauscher 1991. Enhancing the scientific process with artificial intelligence : forest science applications. *AI Applications* 5 (2) pp 5- 26

Meehan, P. 1980. Science, ethnoscience, and agricultural knowledge utilization. In: Brokensha, D., Warren, D.M. and Werner, O. (eds), *Indigenous knowledge systems and development.* University Press of America.

Mollison, B. 1990 *Permaculture : a practical guide for a sustainable future.* Washington D.C., Island press, 579 p.

Monteith, J.L. 1992. RESCAP a RESource CAPture model. In: *Forest/crop/Ecosystem models.* Edinburgh, U.K. Institute of Terresterial Ecology.

Monteith, J.L., Ong, C.K. and Corlett, J.E. 1991. Microclimatic interactions in agroforestry systems. *Forest Ecology and Management,* **45.**

Muetzelfeldt, R.I. 1984. Intelligent knowledge-based systems - potential for forestry. In: Mason, W.L. and Muetzelfeldt, R.I. (eds), *Computers in Forestry*. Proceedings of a conference held at Heriot Watt University, Edinburgh, Dec. 1984. Institute of Chartered Foresters, Edinburgh, pp. 245-251.

Murphy, W.P. 1980 Secret knowledge as property and power in Kpelle society. *Africa* 50(2) pp 193-207

Nair, P.K.R. 1989. *Agroforestry systems in the tropics.* Kulwer Academic Publishers. London.

Nair, P.K.R. 1990. *The prospects for agroforestry in the tropics.* World Bank Technical Paper 131.

Niamir, M. 1990. *Community Forestry, decision making in natural resource management in arid and semi-arid Africa.* Community Forestry Note 4. Food and Agriculture Organization of the United Nations.

O'Neill, R.V., DeAngelis, D.L., Waide, J.B., & Allen, T.F.H. 1986 *A hierarchical concept of ecosystems.* Princeton : Princeton University Press.

Odum, E. 1963. *Ecology.* Holt, Rhinehart and Winston, New York.

PAC 1992 *Annual research Work Programme 1992 - 1993.* Pakhribas Agricultural Centre, c/o BAPSO, P.O. Box 106, Kathmandu, Nepal.

Panday, K.K. 1982. *Fodder Trees and Tree Fodder in Nepal.* Swiss Development Co-operation, Berne Switzerland

Panday, K.K. 1982. *Fodder trees and tree fodder in Nepal.* Swiss Development Cooperation, Berne, Switzerland and Swiss Federal Institute of Forestry Research, Birmensdorf, Switzerland.

Pereira, C.A.D. and Barlow, R.E. 1990. Medical Diagnosis Using Influence Diagrams. *Networks* 20(5): 565-577.

Pickering, D.C. 1988. African agricultural research ad technological development. Proceedings of a meeting. Feldafing, F.R. Germany. The World Bank, Washington D.C., U.S.A.

Plant R.E. and Stone N.D. 1991. *Knowledge Based Systems in Agriculture.* McGraw-Hill, London, UK.

Preechapanya, P. (in preparation). Indigenous ecological knowledge about miang tea gardens in the hill-evergreen forests of Northern Thailand. PhD Thesis, University of Wales, Bangor.

Protsko, L.B., Sorenson, P.G., Tremblay, J.P. and Schaefer, D.A. 1991. Towards the Automatic-Generation of Software Diagrams. *IEEE Transactions on Software Engineering,* 17(1): 10-21.

Raintree, J. B. 1987a. The state of the art of agroforestry diagnosis and design. *Agroforestry systems* 5: 219-250.

Raintree, J.B. 1987b. *Design & Development Users manual: an introduction to agroforestry diagnosis and design.* ICRAF,

Raintree, J.B. 1990. Theory and practice of agroforestry development and design. In : MacDicken, K.G. and Vergara, N.T. (eds), *Agroforestry classification and management.* John Wiley and Sons, pp. 58 - 97.

Rauscher, H.M. 1987. Increasing scientific productivity through better knowledge management. *AI Applications*, 1(2): 21-31.

Richards, A.I. 1939. *Land, labour and diet in northern Rhodesia*. Oxford O.U.P. for the International Institute of African Languages and Culture.

Richards, P. 1980. Community environmental knowledge. In: Brokensha, D., Warren, D.M. and Werner, O. (eds). *African rural development*. Indigenous knowledge systems and development. University Press of America.

Richards, P. 1985. *Indigenous Agricultural Revolution*. Hutchinson & Co.

Rindos, D. 1984. *The origins of agriculture: an evolutionary persective*. Orlando Academic Press,.

Robertson, D.H., Bundy, A, Muetzelfeldt, R., Haggith, M. and Uschold, M. 1991. *Eco-logic, Logic -based approaches to ecological modelling*. MIT Press Cambridge Massachusetts, London, England.

Rocheleau, D.E. 1987. The user perspective and the agroforestry research and action agenda. In: Gholz, H.L. (ed), *Agroforestry: realities, possibilities and potentials*. Martinus Nijhof Publishers.

Sanchez, P.A. 1987. Soil productivity and sustainability. In: agroforestry systems. in. Steppler, H.A. and Nair, P.K.R (eds), *Agroforestry - a decade of development*. ICRAF, Nairobi.

Scherr, S.J. 1990. The diagnosis and design approach to agroforestry project planning and implementation, examples from Western Kenya. In Budd, W.W., Duchhart,I., Hardesty, L.H. and Steiner, F. *Planning for agroforestry*. Developments in landscape management and urban planning : 6c. Elsevier Science Publishing Co. Inc. Amsterdam

SEADD. 1989. *Submission to ODA Projects Evaluation Committee, Nepal, Pakhribas Agricultural Centre, 1988-1995, Volume I - Main Report*. SEADD, Overseas Development Administration, London, UK.

Shachter, R.D. 1988. Probabilistic inference and influence diagrams. *Operations research*, 36(4): 589-604.

Shachter, R.D. 1990. An ordered examination of influence diagrams. *Networks*, 20(5): 535-563.

Shachter, R.D. 1990b. Special issue on influence diagrams - Introduction. *Networks*, 20(5): 487-489.

Shoan, Y., and Goyal, N. 1988. Temporal reasoning in artificial intelligence. In: Shrobe, S. (ed), Exploring artificial intelligence, survey talks from the national conferences on Artificial Intelligence. Morgan Kaufmann, pp. 419-438.

Simmonds, N.W. 1979. *Principles of crop improvement*. Longman, London, New York.

Simmonds, N.W. 1985. *Farming systems research : a review*. World Bank Technical Paper No. 43.

Simon, H.A. 1962. The architecture of complexity. *Proc. Amer. Phil. Soc.* 106, 467-82

Sinclair, F.L., Verinumbe, I., and Hall, J.B. 1994. The role of tree domestication in agroforestry. In Leakey, R.R.B. and Newton, A. (eds), *Tropical trees : the potential for domestication*. HMSO, London in press.

Sinclair, F.L., Walker, D.H., Joshi, L., Ambrose, B., Thapa, B., 1993. *Use of a knowledge based systems approach in the improvement of the tree fodder resources on farmland in the eastrern hills of Nepal*. School of Agricultural and Forest Sciences, University of wales, Bangor.

Singh, R.V. 1982. *Fodder trees of India*. Oxford and IBH Publishing Co. New Delhi, Bombay, Calcutta.

Slagsvold, T. 1980. Habitat selection in birds: on the presence and absence of other bird species with special regard to *Turdus pilaris. Journal of applied ecology* 49: 523-536.

Soemarwoto, O. 1987. Homegardens : a traditional agroforestry system wth a promising future. In: Steppler, H.A., Nair, P.K.R. (eds), *Agroforestry, a decade of development.* ICRAF.

Southern, A.J. (in preparation) Indigenous ecological knowledge about forest gardens in Kandy District, Sri Lanka. MPhil Thesis, University of Wales, Bangor.

Spears, J. 1987. Agroforestry : a development bank perspective. In: Steppler, H.A. and Nair, P.K.R. (eds), *Agroforestry : a decade of development.* ICRAF, Nairobi, Kenya.

Spradley, J. 1979. *The Ethnographic Interview.* New York: Holt, Rinehart & Winston.

Street, P.R. 1990. A systems view of commercial supply and marketing links. In: Jones, J.G.W. and Street, P.R. (eds), *Systems theory applied to agiculture amnd the food chain*, pp 159-209. Elsevier Science Publishing Co. Inc.

Subba, K. J. 1992 *A Comparative Study of Landuse Changes in Dhankuta District Between 1978-1990.* Pilot Land Use Mapping Study, Koshi Hills Development Programme, Dhankuta, Nepal.

Tamm, C.O. 1975. *Man and the boreal forests.* Proceedings of a regional meeting within Project 2 of MAB held Stockhom, Sweden, 13-16 October, 1975. Swedish Naural Sciences Research Council.

Thapa, B. (in preparation) Farmers' ecological knowledge about the cultivation, management and use of farmland tree fodder resources in the middle hills of eastern Nepal. PhD Thesis, University of Wales, Bangor.

Thapa, B. and Walker, D.H. (1992) Visit report : Visit to PAC in support of indigenous ecological knowledge research fieldwork. 8.11.92 - 29.11.92. School of Agricultural and Forest Sciences, University of Wales, Bangor.

Thapa, H.B., Green, T., and Gibbon, D. 1988 *Agricultural Extension in the Hills of Nepal.* Overseas Development Institute, London, UK.

Tick, A.W. 1993. Editorial. *Indigenous knowledge and development monitor.* 1(2).

Van der Poel, P. and Van dijk, H. 1987. Household economy and tree growing in upland Central Java. *Agroforestry systems* 5: 169-184.

Van Mansvelt, J. 1988. The role of lower input technologies in the future. In: Whitby, M. and Olerenshaw, J. (eds), *Land use and the European environment.* Belhaven, London.

van den Hoek, A., & Bekkering, T. 1990. Planning of agroforestry programmes in Java, Indonesia. In: Budd, W.W., Duchhart, I., Hardesty, L.H. and Steiner, F. (eds), *Planning for Agroforestry.* Developments in Landscape Management and Urban Planning 6c, Elsevier Science Publishing Co. Inc. Amsterdam, pp 58-78.

Walker, D.H. and Southern, A.J. 1992. *Report on fieldwork in Sri Lanka April - May 1992.* School of Agricultural and Forest Sciences, University of Wales, Bangor.

Walker, D.H., Muetzelfeldt, R.I. and Sinclair, F.L. 1993. *HyperNet : conventions and a computer program for representing ecological knowledge as node and link diagrams.* Version 1/185. School of Agricultural and Forest Sciences, University of Wales, Bangor.

Walker, D.H., Sinclair, F.L. and Muetzelfeldt, R.I. (1991). Formal representation and use of indigenous ecological knowledge about agroforestry : pilot phase report. June 1991. School of Agricultural and Forest Sciences, University of Wales, Bangor.

Walker, D.H., Sinclair, F.L., Muetzelfeldt, R.I. 1991. *Formal representation and use of indigenous ecological knowledge about agroforestry ; piolt phase report, June 1991.* School of Agricultural and Forest Sciences, University of Wales, Bangor, 111 pp.

Warkentin M.E, Nair P.K.R. and Ruth S.R. 1990 A knowledge based expert system for planning and design of agroforestry systems *Agroforestry Systems*, **11(1)**: 71-84.

Warren, D.M. (ed). 1991. Indigenous agricultural knowledge systems and development. *Agriculture and human values.*, VIII (1 & 2).

Warren, D.M., Slikerveer, L.J., Titilola, S.O. (eds). 1989. *Indigenous knowledge systems : implications for agriculture and international development.* Studies in Technology and Social Change, Technology and Social Change Program, Iowa State University. No 11.

Waterman, D. A. 1986. *A Guide to Expert Systems.* Reading, Mass: Addison-Wesley Publishing Company.

Werner, O., Schoepfle, G.M. 1987a. *Systematic Fieldwork Volume 1: Foundations of Ethnography and Interviewing.* Sage Publications.

Werner, O., Schoepfle, G.M. 1987b. *Systematic Fieldwork Volume 2: Ethnographic Analysis and Data Management.* Sage Publications.

Wesley 1985 Research Directions in Expert Systems. in Doukidis, Georgios I., Frank Land, Gordon Miller eds 1985 Knowledge based management support systems. Ellis Horwood Books in Information Technology.

Wilson, M. 1989. Task models in knowledge elicitation. In: Diaper, D. (ed). 1989. *Knowledge elicitation: Principles, techniques and applications.* Ellis Horwood Ltd. Chichester.

Woodley, E. 1991. Indigenous ecological knowledge sytstems and development. *Agriculture and Human Values,* **VIII (1 and 2)**: 173-178

World Bank. 1990. *Nepal: relieving poverty in a resource scarce economy.* **2**: 143. World Bank, Washington, DC, USA.

Young, A. 1989. *Agroforestry for soil conservation.* Science and Practice of Agroforestry No. 4, Wallingford, U.K. CAB International.

Zulberti, E. 1987. Agroforestry training and education at ICRAF: accomplishments and challenges. *Agroforestry systems,* **5**: 353-374.