# A pollen identification expert system: An application of expert system techniques to biological identification.

A thesis presented in partial fulfilment
of the requirements for the degree of
Master of Science
in Computer Science.

Colin G. Eagle

Massey University

1990

# Abstract

The application of expert systems techniques to biological identification has been investigated and a system developed which assists a user to identify and count air-borne pollen grains. The present system uses a modified taxonomic data matrix as the structure for the knowledge base. This allows domain experts to easily assess and modify the knowledge using a familiar data structure. The data structure can be easily converted to rules or a simple frame-based structure if required for other applications. A method of ranking the importance of characters for identifying each taxon has been developed which assists the system to quickly narrow an identification by rejecting or accepting candidate taxa. This method is very similar to that used by domain experts.

# Acknowledgements

My sincere thanks go to my supervisor, Mr Ray Kemp, for his guidance and useful criticisms during the development and presentation of this research.

I would also like to thank the domain experts, Dr Clive Cornford, Mr Richard Burr and Dr David Fountain of the Botany and Zoology Department for their willingness to provide time and expertise.

Thanks are due to the staff of the School of Information Sciences, who have provided encouragement during the production of this report.

I would like to thank my parents, Gordon and Judy Eagle, for their encouragement and support.

Finally I wish to thank my wife, Susan, for her patience, support and understanding during this time.

# Table of Contents

# List of Figures

# Chapter 1

# Introduction

## 1.1 Purpose

The purpose of the present study is to investigate the suitability of using expert systems technology in the field of biological identification, using pollen identification as an example. In addition, the present study examines the use of a taxonomic data matrix as the core of the knowledge base structure, and also develops a method of assigning importance values to characters.

Chapter 1 describes the objectives of the present study, and presents an investigation into expert systems technology and design, graphical user interfaces and pattern recognition and their relevance to expert systems. Chapter 2 investigates the techniques of biological identification and how expert system techniques are applied to this. Chapter 3 contains descriptions of prototype systems for pollen identification which were intended to determine the practicality of expert systems for pollen identification. Chapters 4 describes the user view, knowledge base organisation and inference engine of the present system. Chapter 5 contains a summary of results achieved and proposals for future developments of the present system.

## 1.2 Objectives of this project

The main objective of the present study is the development of an expert system designed to quickly and accurately identify and count New Zealand pollens based on morphological descriptions given by the user. The system is

designed to run on a computer beside a microscope, assisting the user to interactively identify and count the pollens seen in the microscope.

This study is designed to meet an identified need for a pollen identification system for use in allergen research. Pollen allergens have been identified by the World Health Organization as a research priority. In New Zealand current research (Cornford, Fountain, Burr & O'Leary, 1988) has aimed to build a reference bank of pollens and their extracts, measuring the occurrence of hazardous pollens in the atmosphere, and purifying pollen extracts for use in allergen analysis and treatment programs. This research requires the collection of pollens from throughout New Zealand. Identification of these is primarily carried out by trained but non-specialised staff. These staff would be assisted by an expert system designed to take into account a variety of interacting factors which are crucial to an accurate analysis of pollens. Experienced staff would also benefit from a system which enables them to identify unusual pollens.

In addition to allergen research, there are several other fields where pollen identification may be assisted by an expert system. For example, forensic scientists may need to investigate the approximate area and season in which a crime took place. Apiculturalists can benefit from a pollen identification system to ensure optimum placement of hives, and in palynology pollen identification can aid understanding of plant distribution and geology (Kemp, Greenwood, Tse & Eagle, 1988).

The present study was designed primarily to assist those involved in allergen research. It is intended that the completed system will be used to:

- assist the user to count different types of pollen;

- lead the non-specialised user through an identification process, asking for data which either confirm or negate the most likely candidate pollen;
- assist more experienced users in routine identification and in identifying unusual pollens, via an option which omits the questioning process and allows direct description of an unidentified pollen;
- report when it is not possible to clearly differentiate between two pollens;
- explain the process used to confirm or negate candidate pollens;
- be easily amended to provide identifications in other fields of biological identification;
- incorporate a graphical user interface so that the system is simple and intuitive to use;
- be easily extended to incorporate real-time pollen recognition.

## 1.3 Expert Systems

Expert (knowledge-based) systems are computer programs which can solve 'real world' problems, that is, problems for which a solution requires judgement and experience. The emphasis of expert systems is on the heuristic knowledge which reflects the experience of the expert and the structure of that knowledge, rather than on reasoning from first principles (Michaelsen, Michie & Boulanger, 1983; Wolfgram, Dear & Galbraith, 1987).

An important aspect of expert systems is a capability for explaining their knowledge of the domain and the reasoning processes used to produce results and recommendations. This assists users and system builders to understand the contents of the system's knowledge base and reasoning processes, and

facilitates the debugging of the system during development. It educates users about both the domain and the capabilities of the system, and gives information which assures users that the system's conclusions are correct. Explanation can also help a user to discover when the limits of the system's knowledge are being exceeded (Moore & Swartout, 1988).

In order to make use of judgemental knowledge, expert systems normally include a method for reasoning with uncertainty. This allows better modelling of expert behaviour, including the use of guesses and degrees of belief (Atkinson & Gammerman, 1987).

Other useful aspects of expert systems include the capacity to mimic human reasoning, making the logical progress toward a problem solution easily understood by users. It is also possible to build generalisable systems, that is, an expert system designed to identify one type of biological specimen can, by changing the knowledge base, be used to identify another type of specimen (Woolley & Stone, 1987).

Hayes-Roth, Waterman and Lenat (1983), Wolfgram et al (1987) and Poo and Lu (1989) have identified distinct categories of expert systems designed to solve particular types of problems.

Firstly, fixed instant diagnosis systems (i.e., those in which interpretation of a diagnosis at a point in time depends on the data available), may be used, for example, in medical, electronic, mechanical and software diagnosis (Poo et al, 1989). MYCIN is an example of a medical diagnosis system which attempts to diagnose infectious blood diseases from available knowledge or data supplied by a physician. Clancey (1984) has described various methods of designing fixed instant diagnosis systems.

Secondly, interpretation systems can be used in areas such as surveillance, speech understanding, image analysis and signal interpretation. They attempt

to explain observed data by assigning to them symbolic meanings describing the system state accounting for the data (Hayes-Roth et al, 1983). DENDRAL analyses experimental chemical data in order to infer the plausible structures of an unknown compound (Wolfgram et al, 1987).

Thirdly, prediction systems infer likely consequences from given or hypothetical situations (Wolfgram et al, 1987). This category includes weather forecasting, demographic predictions, traffic predictions and military forecasting.

Planning systems compose sequences of actions for achieving some prescribed effect. This category includes automatic programming, and robot, route, experiment and military planning problems (Hayes-Roth et al, 1983).

Configuration systems construct descriptions of objects in various relationships with one another, and verify that these configurations conform to stated constraints (Wolfgram et al, 1987). These systems include computer configuration (e.g., R1, the DEC VAX computer equipment configuration system), circuit layout, building design and budgeting.

Advice giving systems use recommendations and explanations in attempting to provide the user with a supportive environment for problem solving (Coombs & Alty, 1984; Jackson and Lefrere, 1984). This category includes plan formation and computer programming.

Finally, computer-aided instruction systems incorporate diagnosis and debugging subsystems that address the student as the system of interest. Typically, these systems construct a model of the students knowledge which interprets the students behaviour, diagnose weaknesses in the students knowledge, identify an appropriate remedy, and then plan a tutorial intended to convey the remedial knowledge to the student (Hayes-Roth et al, 1983; Farrell, Anderson & Reiser, 1984; Clancey & Bock, 1988).

The architecture of a typical expert system consists of a fact base, a knowledge base, an inference engine and an explanation facility (Hayes-Roth et al, 1983; Ramsey, Reggia, Nau & Ferrentino, 1986; Poo et al, 1989). (See Figure 1). A fact base may be defined as a store of unchanging knowledge about the domain of interest of the expert system. A knowledge base consists of extensive knowledge regarding the domain of interest, and is used to make inferences about unknown facts, based on information in the fact base. An inference engine is responsible for control of the problem solving process, that is, manipulating the knowledge base, updating the state of the world, and remembering the chain of reasoning being used. It makes use of knowledge in the knowledge base in order to reason about the problem using information in the fact base. In order to provide a more transparent and explainable design, Buchanan and Duda (1983) and Clancey et al (1988), have proposed that inference procedures be represented abstractly, as rule sets, separate from the domain knowledge they operate on. This has advantages for design and maintenance of the system, making it easier to debug and modify, as hypotheses and search strategies are not embedded in rules. The explanation facility of the inference engine consists of an identification of steps used in the reasoning process and justification of each step.

The knowledge bases of expert systems are commonly divided into two types of knowledge representation: rules and frames. Rule-based (production) systems consist of the knowledge and experience of a human expert encoded into a set of rules which consist of antecedents (conditional statements) that define a pattern or state; and consequents, that is, instructions to be carried out in the event that the current state matches the hypothetical pattern described in the antecedent (Woolley et al, 1987). The skill of a rule-based system increases at a rate proportional to the enlargement of its knowledge

base. Rule-based systems are modular, in that each rule defines a small, relatively independent piece of knowledge; this allows relatively simple addition of new rules and updating of old rules (Bratko, 1986). By adaptively selecting the best sequence of rules to execute, and by combining the results in appropriate ways, rule-based systems can solve a wide range of possibly complex problems. They can explain their conclusions by retracing lines of reasoning and translating the logic of each rule into natural language (Hayes-Roth, 1985).

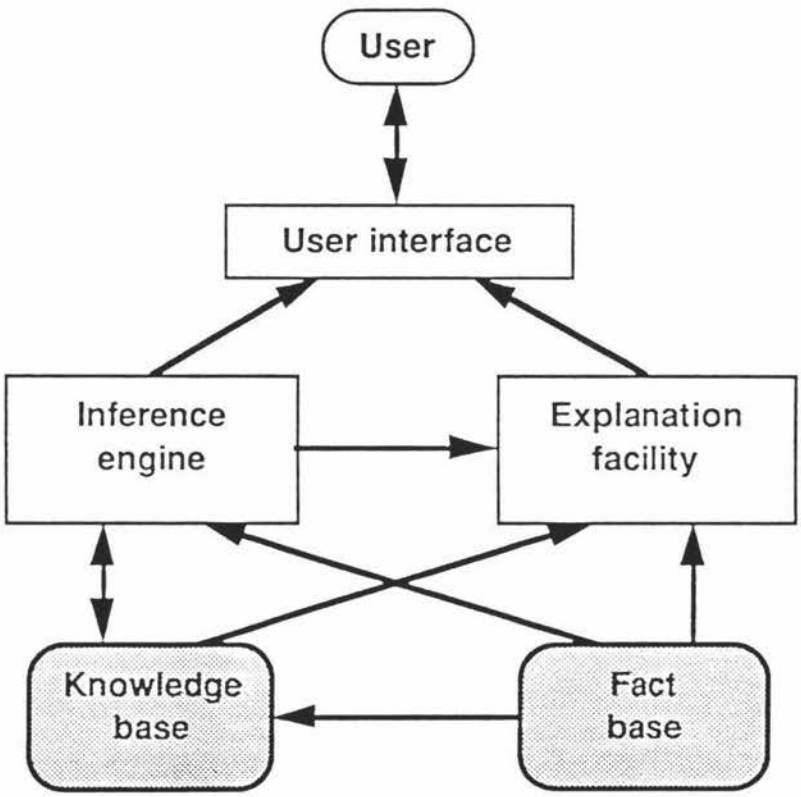


Figure 1: Structure of a typical expert system

(arrows show direction of information flow).

Frame-based expert systems are based on a structured representation of an object or a class of objects (a frame). Frames incorporate sets of attribute descriptions called slots, which are used to describe attributes of the object or class represented by the frame. Constructs are available which allow an

expert system designer to describe relationships between frames (Hayes, 1979; Brachman, 1983; Fikes & Kehler, 1985; Wolfgram et al, 1987). For example, birds can be described as animals in addition to a set of properties which distinguish birds from other classes of animals.

In addition to rule- and frame-based systems, Tschudi (1988) has proposed another type of knowledge representation based on a matrix similar to a taxonomic data matrix. The knowledge in the matrix can easily be encoded to produce rules or a decision tree.

## 1.4 Graphical User Interfaces

Apple Computer (1987) have defined a computer interface as:

> ". . . the sum of all communication between the computer and the user. It's what presents information to the user and accepts information from the user. It's what actually puts the computer's power into the user's hands." (p. xi).

Graphical user (direct-manipulation) interfaces are common on many types of computer system. They provide a human-computer interface which is easier to learn and simpler and more pleasant to use than the traditional command-line interface (Gould & Lewis, 1983; Foley & van Dam, 1984).

Direct manipulation interfaces have been defined by Shneiderman (1983) as a variety of graphical user interface in which the user sees a continuous representation of the world of action. The objects of interest and the permissible actions on those objects are represented on the screen in a visual format which takes into account the user's knowledge of the task domain. Physical actions replace typed commands and actions are rapid, incremental and reversible. These design principles lead to several important benefits. Users with knowledge of the domain find the system easy to learn, users need

learn only a small number of computer concepts, and can therefore concentrate on the task. In addition, designers can reduce the number of situations in errors can be made, users feel free to explore 'what-if' possibilities, and long-term retention is facilitated (Baroff, Simon, Gilman & Shneiderman, 1986).

In expert systems, effective use of direct manipulation interfaces can assist in containing complexity and make the system intuitive and credible to use. This can be done by exploiting the user's expectations regarding how ideas are organized and expressed within the system domain (Potter, 1988). Direct manipulation interfaces have been used in expert system development, allowing designers to display rules and heuristics in graphical format and to graphically display actual and possible interactions between rules (Poltrock, Steiner & Tarlton, 1986; Baroff et al, 1986).

In addition to expert system development, direct manipulation interfaces may be used in the user-computer interface. For example, 'The Student Advisor' (Baroff et al, 1986), assists students in planning course schedules and uses a windows and buttons in order to simplify the interface. The apple problem diagnosis system (Kemp & Boorman, 1987) attempts to determine the cause of inadequate quality or quantity of fruit using 'windows', 'icons', 'mice' and 'pull-down menus' ('wimps') for more effective user interaction and therefore allowing the user to adapt quickly to the system, even though he/she may not use it for extended periods.

## 1.5 Pattern Recognition

Pattern recognition refers to the act of recognising a given object from a complex input stream. For example, identifying a chair from the wider class of 'furniture' (Pao & Ernst, 1982). Three interrelated but distinct processes take

place during a typical pattern recognition process. Data acquisition is the process of converting incoming data from its physical source (pictures, speech, character string, etc.) into an acceptable form for further processing. Pattern analysis is concerned with organising the converted body of data into a form for further processing by determining the different pattern classes which might exist in the data. Finally, pattern classification refers to the process whereby pattern classes are matched with a known class (Chien, 1978). Pattern classification has used expert systems techniques since the early 1960's, particularly where there is imperfect correspondence between input data and a known class (Ballard, Brown & Feldman, 1977; Ogawa, Kurioka, Kitahashi & Tanaka, 1980; Brady, 1982; Magee & Nathan, 1985). For example, galaxy classification (Thonnat, Granger & Berthod, 1985), inspection of mechanical parts (Kanal, 1974), and the interpretation of medical images to provide diagnoses (Ellam & Maisey, 1986).

The application of computerised pattern recognition has been largely directed toward computer vision (e.g., object classification) and speech recognition. A summary of pattern recognition techniques has been provided by Rohlf and Ferson (1983).

# Chapter 2

# Biological Identification

## 2.1 Introduction

Biological identification may be defined as the assignment of a specimen to a particular classification (a group of like objects with which the specimen can be compared) such as a species or genera. Such classifications are generally referred to as 'taxa' (singular: taxon) (Pankhurst, 1978; van Rijsbergen, 1979; Funk, 1982; Sneath, 1982; Clancey, 1984; Brandenburg, 1986). A taxonomic group can be described using a taxonomic data matrix (i.e., a table presenting character states for a number of taxa).

The fundamental unit of information used during identification of a specimen is known as a character. Characters provide the basis for comparing specimens against taxa. For example, 'shape' is a character of a pollen grain. Information regarding characters and their states can be used to identify specimens, for example, the states of pollen 'shape' might be 'circular', 'elliptical', or 'triangular'.

## 2.2 Traditional Methods of Biological Identification

There are two principle methods of biological identification: monothetic, in which only one character at a time is used; and polythetic, in which several characters are used simultaneously. During monothetic identification, all characters used in identification of a specimen must exactly fit the description of the taxa with which it is being compared. A particular single character (a 'key character'), or the lack of such a character, can be sufficient to exclude a specimen from a taxa, regardless of how well the specimen identifies with the

taxa in other ways (Morse, 1975). Monothetic methods may be single-access, in which the sequence of characters to be used is strictly defined, or multi-access (polyclave), that is, the user can select the characters and the order in which they are to be used (Morse, Pankhurst & Rypka, 1975; Pankhurst, 1978).

Among single-access methods, dichotomous keys are the most common means of biological identification and have been used extensively for pollen identification (Kapp, 1969; Faegri & Iversen, 1975; Moore & Webb, 1978; Smith, 1984). The user follows a series of choices between contrasting alternatives, with each choice leading either to another choice or to the name of a taxon (Morse et al, 1975). Diagnostic keys are a more general form of dichotomous key in which more than two choices are available. These are not as widely used as dichotomous keys, but have been used in zoology, microbiology and pharmacognosy (Morse et al, 1975; Pankhurst, 1978).

Single-access monothetic methods are conceptually simple, but the user is obliged to follow the sequence of characters laid down in the key. Distinctions must sometimes be made by means of rather subtle characters early in the key, creating uncertainty in the outcome (Faegri et al, 1975). Ambiguous choices and errors in character observation may produce erroneous results. That is, the wrong choice early in a key can send an inexperienced user on a fruitless search through divisions which do not seem to apply to the specimen being identified (Woolley & Stone, 1987).

Of the multi-access monothetic identification methods, punched cards are most commonly used and have been used in the fields of botany (particularly pollen identification), microbiology and pharmacognosy (Pankhurst, 1978). On body-punched cards, each taxon is assigned a particular position on all cards. Any taxa which agree with all character states will be identified by the alignment of punched holes in the appropriate position on all relevant cards.

Edge-punched cards have character states assigned to particular punched positions around the edges of the cards. Taxa which agree with a given character state are identified by a process of eliminating all non-matching taxa. The process is repeated until only one taxon remains (Faegri et al, 1975; Morse, 1975; Pankhurst, 1978; Sawyer, 1981; Abbott, Bisby & Rogers, 1985; Ecroyd, 1986).

Multi-access monothetic methods can be used with incomplete data and allow the use of striking characters at the beginning of the sequence which can shorten the identification process. Also, taxa and characters can be easily added to multi-access keys. However, errors can not be easily remedied once made (Westfall, Glen & Panagos, 1986). Identification methods combining features of single- and multi-access keys have been proposed but not extensively used (Westfall et al, 1986).

The other principle means of biological identification, polythetic methods use multiple characters simultaneously. Specimens need to be described in detail before being submitted for identification. However, not every character of the specimen need agree with those of the identifying taxon, providing the principle of "overall similarity" is obeyed (Morse, 1975). The smallest set of characters which must be observed in order to separate all taxa is known as the minimum character set.

Identification by comparison (character-set) is an example of a polythetic method. It involves observation of specimen characters and comparison of observed characters with reference taxa groups. This may involve calculating some measure of agreement based on character matches and mis-matches (Pankhurst, 1975). The result of the comparison may be one, or several taxa, from which a subjective choice must be made (Pankhurst, 1978).

Polythetic identification methods have been applied to botany, zoology, and medical diagnosis and microbiology in particular (Pankhurst, 1978). For example, it is preferable to use a polythetic method to obtain identifications from the results of biochemical and physiological testing, as tests can be done in parallel. Polythetic methods do not demand any particular characters from the specimen, and can be used successfully with incomplete material. However, it may be difficult to compare the specimen with every taxon, particularly if there are many taxa to consider. Polythetic methods generally require the user to observe a larger number of character states for the specimen than do sequential methods such as a monothetic method.

## 2.3 Computer Methods in Biology

Computers have been used to solve a variety of problems in biology since the early 1960's (Pankhurst, 1978). For example, creation of taxonomic keys, support of taxonomic databases, creation of classifications, display of plots of taxonomic data, and automation of the identification process.

### 2.3.1 Taxonomic key creation

Computer programs for single-access monothetic key production have been developed by Hall (1975), Payne (1975), and Gunn and LaSota (1977). The method generally used involves a recursive procedure which repeatedly divides sets of taxa into roughly equal mutually exclusive subsets on the basis of one or more taxonomic characters. The most significant difference between various key-construction programs is the method of choosing the dividing character (Morse, 1975; Abbott et al, 1985).

A computer program for body-punched card key (a multi-access monothetic method) production has been developed by Pankhurst (1975, 1978).

The program arranges to produce a card with holes punched in the position by the characters and states of a taxonomic data matrix. An interactive computer program for the construction and revision of identification keys (KCON) has been proposed by Pankhurst (1988b). KCON provides the mechanical aspects of writing and organising the key, while the user contributes subjective taxonomic skills. Pankhurst (1983) has also developed a computer program which finds all possible sets of characters which distinguish a taxon from all the others in a taxonomic data matrix. These sets of characters can be used for describing a specimen in a polythetic identification process. Another computer program (PHYTOTAB) has been developed by Westfall et al (1986) which generates a matrix used for an identification method combining monothetic and polythetic methods. The advantages of using computerised rather than manual methods of key construction are that they are accurate, repeatable, flexible and comprehensive (Pankhurst, 1986).

### 2.3.2 Taxonomic database support

The traditional method of publishing taxonomic information, the printed text, has two shortcomings: it is normally indexed only by scientific name, and quickly becomes outdated. Computerised taxonomic databases can be easily updated and allow data retrieval using any combination of attributes (Felsenstein, 1983; Bisby, 1984). A number of identification programs have been developed as interfaces to taxonomic databases (Margot, Farquhar & Watling, 1984; Gomez-Pompa, Moreno, Gama, Sosa & Allkin, 1984).

Applications of taxonomic databases include the Taxonomic Reference File, which allows users to search using an organism's name or descriptive data (Dadd & Kelly, 1984) and the European Taxonomic Documentation System for the vascular plants of Europe (Heywood, Moore, Derrick, Mitchell & van

Scheepen, 1984). The PRECIS taxonomic system (Gibbs Russell & Arnold, 1989) and the Vicieae Database Project (Bisby, White, Macfarlane & Babac, 1983) are other examples.

Germeraad and Muller (1971) and Dallwitz (1980) have developed systems for describing taxonomic characters to databases. These ensure data is in a consistent format, and allow conversion to other formats for collection or reordering of data. Allkin and Bisby (1988) have proposed that images be used in taxonomic databases to provide an effective means of storing and communicating descriptive information about form, shape and colour. The production of taxonomic databases has been encouraged by the recent development of relational database systems (Abbott et al, 1985; Dextre Clarke, 1988; Pankhurst, 1988a; Beaman & Regalado, 1989).

### 2.3.3 Classification creation

Numerical methods of creating classifications existed long before the advent of computers (Lamarck, 1778), but were tedious and inefficient (Baum, 1986). Computers have assisted the development of generalised algorithms for classification methods which are applicable to many fields. For example, Baum (1986) surveyed 134 classification programs for cultivated plants. They can also be used for comparison of classifications, in order to test a new method, for example, Sackin (1987) has produced a survey of computer programs for creating and comparing classifications.

### 2.3.4 Taxonomic data plotting

The display of taxonomic data is used particularly in palynology, where pollen diagrams provide a graphical description of one or more pollen samples. This allows researchers to easily note changes in samples over time

and geographic areas. The advent of computer graphics has made the process of plotting diagrams easier and more accurate (Dodson, 1972; Faegri et al, 1975; King, 1976; Moore et al, 1978; Melief & Wijmstra, 1984).

### 2.3.5 Identification

In the field of computerised biological identification computer-stored single-access monothetic keys have an advantage over printed keys in that they can be used to edit the key or use computer graphics to illustrate the choices at each stage of the key (Morse, 1975). For example, the fungal key designed by Kendrick (1972) presents the user with a pair of illustrations to choose between. The stages are ordered from gross characters to small-scale characters, and the illustrations build up successively more detailed diagrams as identification progresses. Bossert (1969) used a computer-stored key to identify Polynesian ants.

Computerised multi-access monothetic methods represent an alternative method of identification which allow the user to specify the state of a given character, or request the 'best' character; use of which results in faster, more efficient identification (Pankhurst, 1978). The computer program IDENT4 developed by Morse (1975) can suggest 'best' characters, allow the user to correct errors during identification, and list the possible taxa at each step in the identification. In addition, the user can specify the extent to which specimens must match taxa in order to meet identification criteria. A computer-based interactive multi-access monothetic procedure (ONLINE) developed by Pankhurst and Aitchison (1975) allows the user to choose a character of the specimen. The user then chooses among states of the character supplied by the computer. The process is designed to eliminate taxa which do not match the character state until, ideally, one taxon remains. A

more recent version of ONLINE uses colour graphics to describe character states (Watson, Dallwitz, Gibbs & Pankhurst, 1988).

Computerised identification is most commonly conducted using polythetic methods; the identification by comparison method in particular. Gyllenberg and Niemela (1975) have developed a method, based on the cluster analysis method used in information retrieval, which uses geometric models (van Rijsbergen, 1979; Weiss, 1981). A multidimensional 'identification space' is established with one axis for each character used. A point for the specimen is located in the identification space. The likelihood of the specimen belonging to a taxon can be determined by examining where it lies in relation to the taxa in the identification space.

Examples of the use of computerised identification by comparison include the work of Walker, Milne, Guppy and Williams (1968), Peters (1969), Guppy, Milne, Glikson and Moore (1973), and Pankhurst (1975).

## 2.4 Expert Systems in Biological Identification

Clancey (1984, 1985) analysed a number of expert systems used for identification, and found that all the systems used a heuristic identification process which matched solutions and solution features by direct, non-hierarchical association. However, heuristic relations are uncertain, based on assumptions of typicality, and may simply be poorly understood correlations.

Expert systems which solve biological identification problems can be divided into two classes: those in which the system determines choices of characters and ordering of questions (single-access monothetic); and those in which choice of characters and questions are determined by the user (multi-access monothetic). MYCIN and INTERNIST represent two early examples of the single-access monothetic approach and were designed to assist physicians

with diagnosis and treatment by questioning about symptoms and identifying the disorder before suggesting appropriate treatment (Duda & Gaschnig, 1981; Kane & Rucker, 1988). A single-access monothetic system implemented by Dodson and Rector (1984) allows importance values to be assigned to characters. These assist the system in the ordering of questions.

A simple single-access monothetic expert system for identification has been developed by Woolley et al (1987) using a commercial expert system shell. SYSTEX makes an initial guess about likely taxa using easily observed characters, then attempts to verify that the specimen is one of the likely taxa. The verification procedure requests the user to observe key characters and uses heuristics to determine the number and identity of characters needed to make an identification.

EXPERT KEY is a single-access monothetic expert system designed to minimise the exposure of a novice to technical terminology. It is a rule-based system which has some domain knowledge incorporated in the inference engine, and includes uncertainty handling, including ignorance of the state of a character. It initially asks a series of questions about easily observed and explained characters of the specimen. This information is used to eliminate some species as possible identities and boost the probability of others, before entering a computer-stored dichotomous key. Because the key has to discriminate between a reduced number of species, the path through the key is considerably reduced (Atkinson et al, 1987).

XPER is a multi-access monothetic expert system for biological identification which uses a taxonomic data matrix. It allows the user to choose from a list of available characters and character states until either a match is made with a taxa, or all characters have been described and more than one taxon matches the description. The user is allowed at any stage to request the

strategy being used to match or eliminate taxa (Forget, Lebbe, Puig, Vignes & Hideux, 1986; Lebbe, Nilsson, Praglowski, Vignes & Hideux, 1987).

# Chapter 3

# Prototypes for a Pollen Identification Expert System

## 3.1 Introduction

Prototypes for the pollen identification system used in the present study were developed in order to confirm that expert system techniques are applicable to the domain of pollen identification and to explore features required by the domain experts and users of the final system. Two prototypes were developed during the initial stages of the project; a single-access monothetic approach and a multi-access monothetic approach.

## 3.2 Prototype 1: Single-access monothetic

The first prototype is based on a dichotomous key created by Kapp (1969). Although the original key provides an identification to species level in most cases, for the purposes of a prototype the system identifies pollens to the level of morphological type only. All the characters are treated as key characters, therefore a candidate pollen can be rejected on the basis of just one non-matching character.

### 3.2.1 User View

Prototype 1 simply queries the user about the characters it requires in order to make an identification, then displays the result when identification is complete. There is no explanation facility allowing the user to ask why a particular query is being made or how a result was obtained. The user is constrained to following the sequence of queries set by the system. There are no facilities for indicating that the answer to a question is not known or for

changing the description of a character state once given. As this prototype was designed only for demonstration to domain experts, no attempt to simplify jargon was made. Appendix A contains an example session using this prototype.

### 3.2.2 Knowledge representation

Prototype 1 used a rule-based representation. The rules are based on branches of a conceptual tree created by the dichotomous key. Each node of the tree corresponds to a choice point in the key, the branches from each node being the choices available at that point. The leaf nodes correspond to candidate types. The rules describing candidate types were determined by creating rules from nodes and branches used in tracing the route from the root of the tree to the type node. Figure 2 displays a section of the conceptual tree. The rules generated by that section are shown in figure 3.

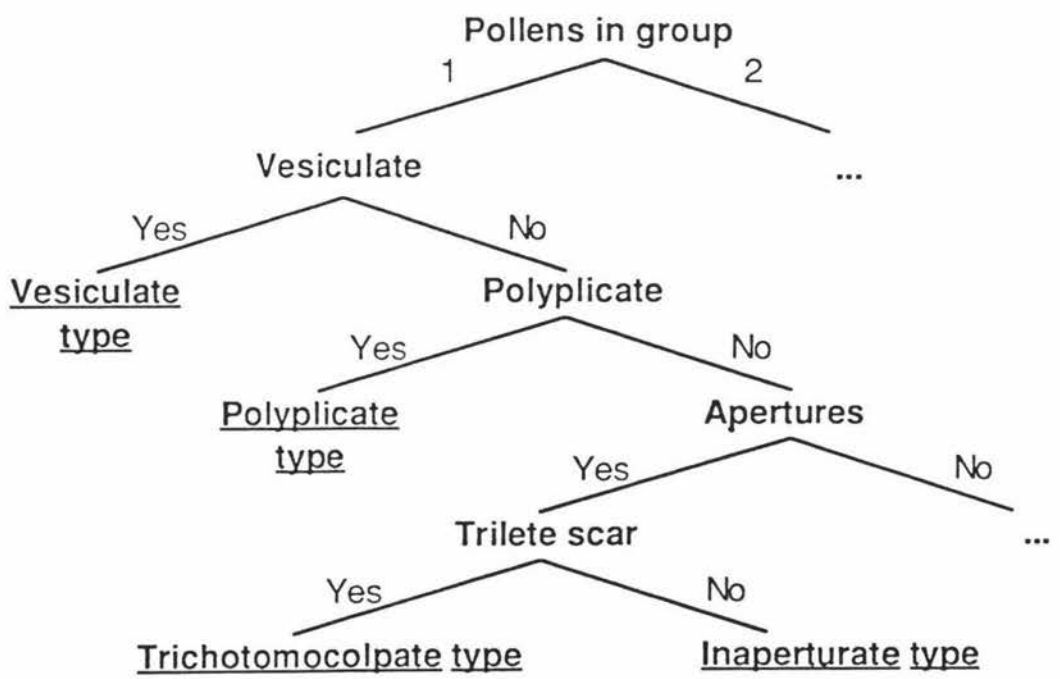


Figure 2: Section of tree formed from dichotomous key.

```
type is vesiculate if:
        pollens in group = 1 and
        vesiculate = yes.

type is polyplicate if:
        pollens in group = 1 and
        vesiculate = no and
        polyplicate = yes.

type is trichotomocolpate if:
        pollens in group = 1 and
        vesiculate = no and
        polyplicate = no and
        apertures = no and
        trilete scar = yes.

type is inaperturate if:
        pollens in group = 1 and
        vesiculate = no and
        polyplicate = no and
        apertures = no and
        trilete scar = no.
```

Figure 3: Section of knowledge base formed from tree shown in Figure 2.

### 3.2.3 Inference engine

The inference engine selects a candidate type from the knowledge base and attempts to verify the selection by confirming each part of the rule corresponding to that type. If any part of the rule fails another candidate type is chosen. The system records the states of characters as they are given, to ensure that if one candidate type fails, only remaining candidates which match the character states recorded so far are considered. The process continues until a match is found, corresponding to a leaf node of the tree.

## 3.3 Prototype 2: Multi-access monothetic

The second prototype is based on a body-punched card key created by Sawyer (1981). It allows the identification of pollens to species level. As in prototype 1, all characters are treated as key characters.

### 3.3.3 User View

Prototype 2 allows the user to select a character they wish to describe from the list of available characters, then choose from a list of possible states for that character. If more than one candidate pollen remains, the system displays the list of remaining candidates before continuing. If one candidate pollen remains, the system displays the identification, including a list of the characters given by the user which match the description of the candidate. If all candidates are negated, the system displays the characters and states given by the user. In addition, it also allows the user to view a complete description of any pollen in the data base.

Appendix B contains an example session using this prototype.

### 3.3.2 Knowledge representation

The knowledge representation structure is based on a taxonomic data matrix. Every pollen has a column in the matrix which describes its morphological structure. Each character has a row in the matrix, where a description of its state for the intersecting pollen is stored. Figure 4 is a section of the matrix and describes a grass pollen.

| Character | Pollen |
|---|---|
| Name | grass |
| Size | medium (30-50μm) |
| Shape | round or irregularly round |
| Aperture number | 1-2 |
| Aperture type | pores only |
| Surface | granular |
| Exine section | thin |
| Other features | thickened or projecting edges |
| Colour | white to grey |

Figure 4: Section of taxonomic data matrix describing a grass pollen.

### 3.3.3 Inference engine

As the description of the specimen pollen is built up by the user, the inference engine keeps account of all pollens which could fit the description. As each character is given a state, the list of candidates is searched, and the pollens which do not have matching character/state pairs are removed. The

process continues until either one candidate pollen remains or all candidates are negated.

## 3.4 Conclusion

After demonstrating the prototypes to both the domain experts and prospective users of the final system, it was decided that there were advantages to each method. The single-access monothetic prototype was found to be useful for inexperienced users, who require help in recognising important features of pollen morphology. The multi-access monothetic prototype was useful for assisting experienced users to make accurate identifications.

# Chapter 4

# Biological Identification Expert System

## 4.1 Introduction

It was decided that the final system should be a monothetic system, and incorporate both single- and multi-access modes in order to lead inexperienced users to an identification, and assist experienced users to identify unfamiliar pollens. The system should have a method of assigning importance values to characters, allowing prominent characters to be used early in an identification so that candidate pollens can be quickly confirmed or negated. There should also be a facility for changing the description of a character state, allowing errors to be easily corrected.

The architecture of the present system consists of a knowledge base, a knowledge conversion facility, an inference engine, an explanation facility and a user interface. (See Figure 5). The knowledge base consists of knowledge regarding the domain of interest, and is structured so that it is easily accessible to domain experts. The knowledge conversion facility is used to convert knowledge from the stored format into a format used by the inference engine and explanation facility. The inference engine is responsible for control of the problem solving process. The explanation facility consists of an identification of steps used in the identification process and justification of each step.

The present system has been implemented on an Apple Macintosh computer using LPA Prolog. The system makes use of the graphical features available in LPA Prolog. (See Appendix F). The present system is

approximately 150 kilobytes in size. Each taxa occupies an additional 10 kilobytes.

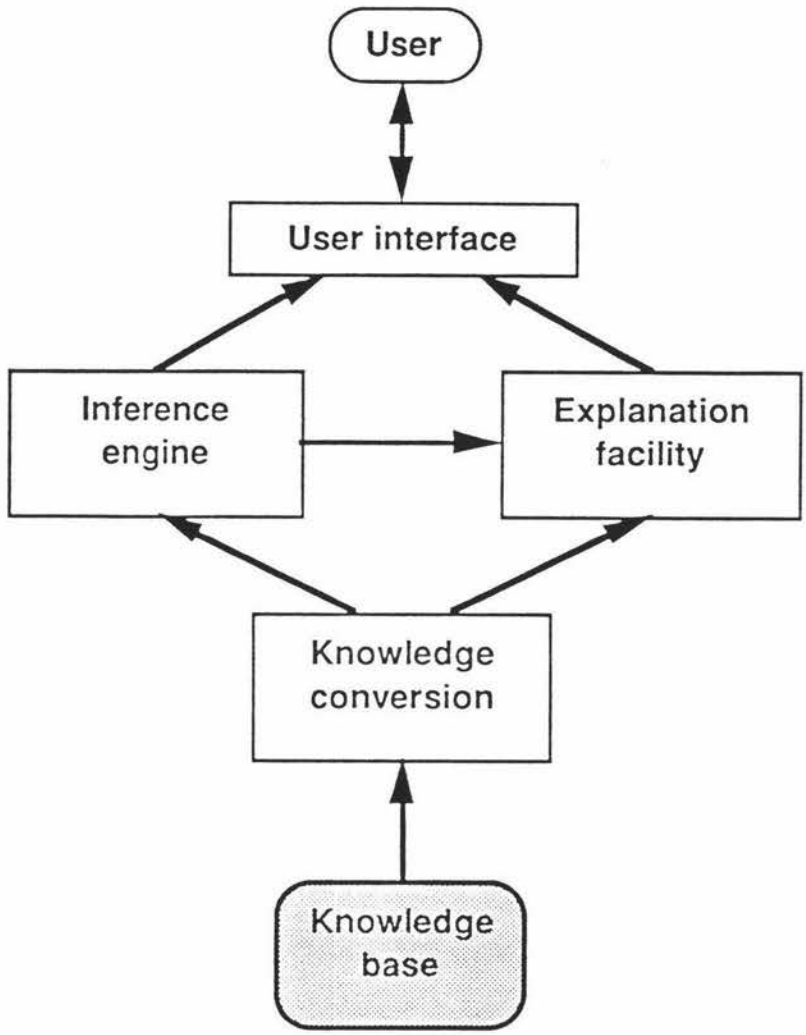


Figure 5: Structure of the present system.

(arrows show direction of information flow)

## 4.2 User View

The present system may be operated in both single-access and multi-access monothetic identification modes. The user can choose which mode they wish to operate, and can change modes at any stage during the identification process. This allows the user to describe the specimen to the system using the

multi-access monothetic mode, then use the single-access monothetic mode to resolve the identification if more than one candidate taxa remains.

The identification process continues until all characters have been used, all candidate taxa have been rejected, or the remaining taxa cannot be distinguished using the characters remaining to be described. At the end of an identification the system displays the remaining taxa, and indicates how well the description of the specimen given by the user matches that of each taxa.

The user interface used by the system is based on 'windows', 'icons', 'mice' and 'pull-down menus' ('wimps'). Choices in the system are presented in the form of menus, while actions the user can take are presented using buttons. Characters which are physical (e.g., shape of a taxa) are shown as pictures. All actions possible at each stage are present on the screen, and can be selected with a mouse, removing the need for remembering and typing complex commands. An complete session using the present system can be found in Appendix E.

### 4.2.1 Single-access monothetic mode

The single-access monothetic mode is primarily intended for users with little identification experience in the domain of the system. It directs the user to describe the specimen by selecting a character and providing the user with relevant states from which to choose. The user can select one or more states corresponding to the character for each of the candidate taxa. Alternatively, the category 'none' indicating that the character state of the specimen does not correspond to any offered by the system; or the category 'unknown', indicating that the character state is not known to the user.

The following is an example of the use of the single-access monothetic mode using the present system with a pollen identification domain. In Figure 6, the user has been asked about the state of the surface on the pollen specimen. He/she does not know the state of the surface on the specimen, so has selected 'Unknown'.

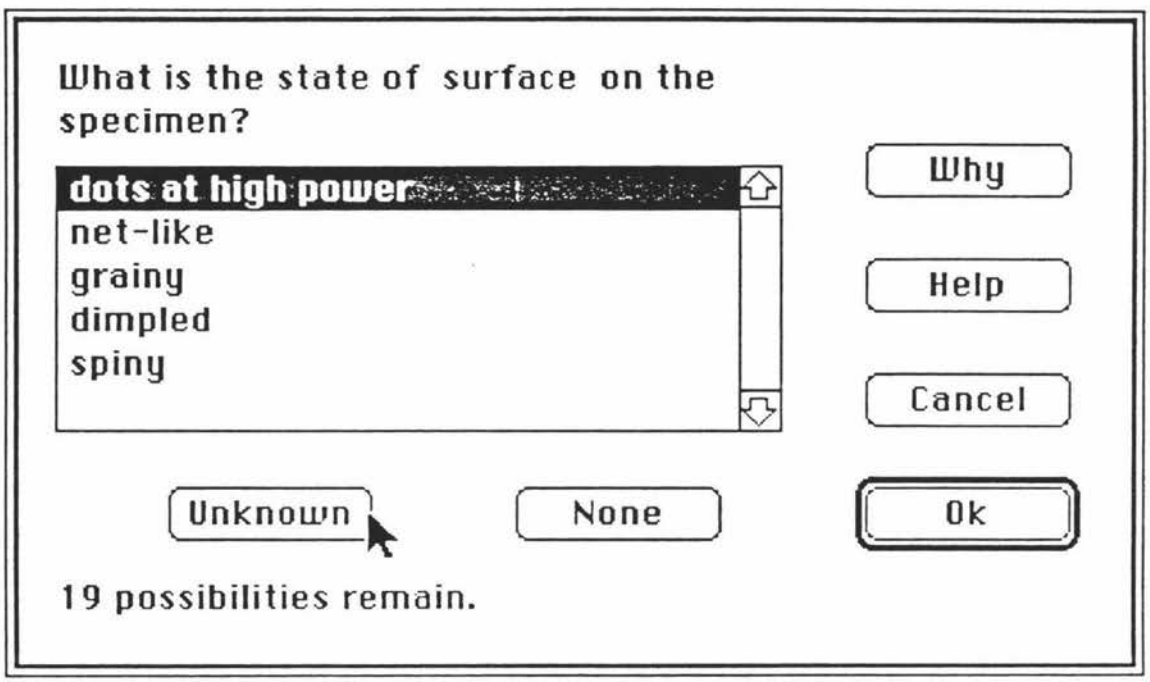


Figure 6: Surface dialog.

Another character, intine appearance, is then queried. The user does not know the state of the intine on the specimen, so again has selected 'Unknown' (Figure 7). Another character, shape, is then asked about. In Figure 8, the user has selected 'triangular'.

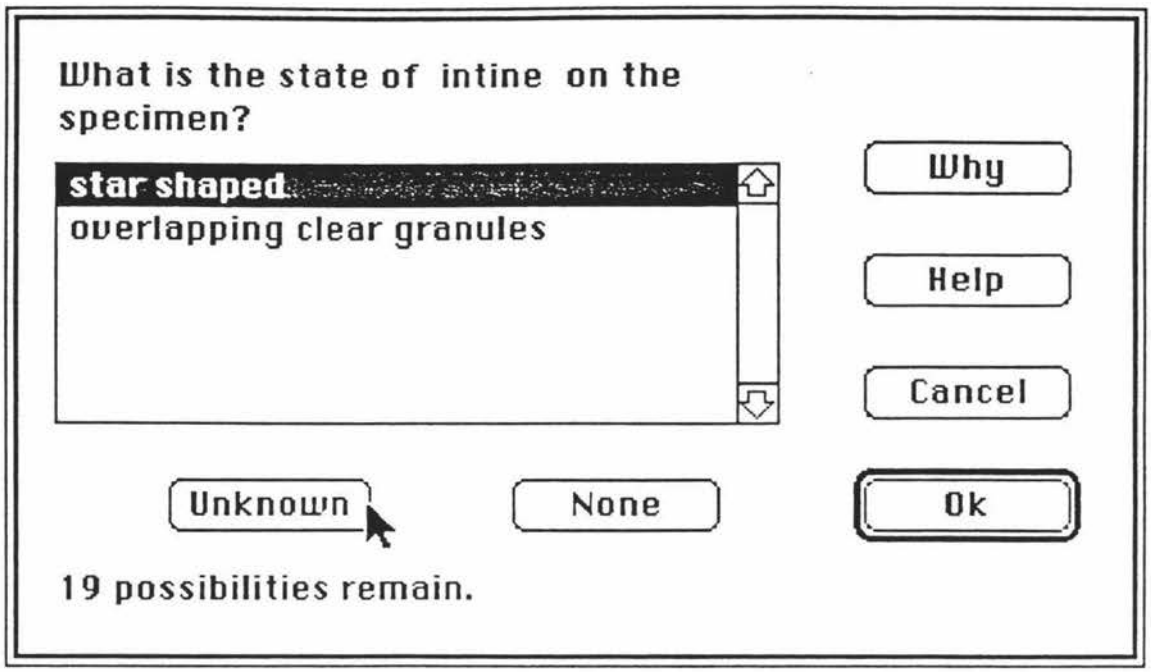


Figure 7: Intine dialog.

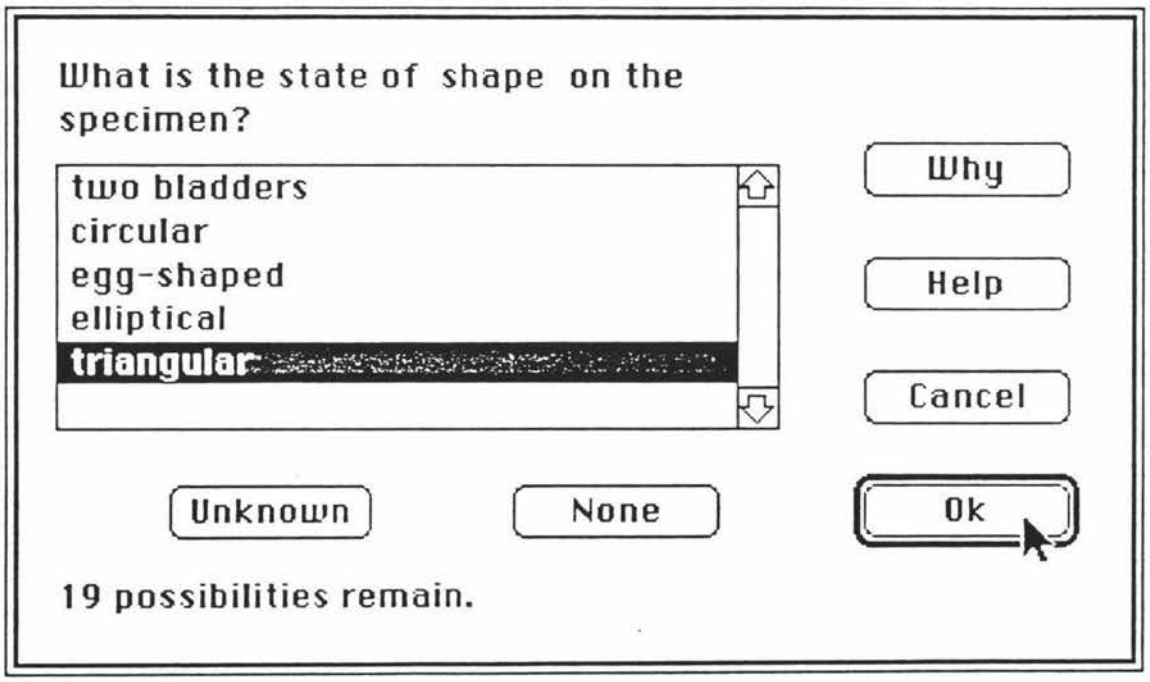


Figure 8: Shape dialog.

The selection of the state 'triangular' for the character 'shape' has enabled the system to identify the specimen as a myrtacea (Figure 9). The 'unlikely'

comment shows that although the system has identified the specimen as a myrtacea, it is unlikely that a myrtacea would be collected during the given season at the given location (see Appendices D & E).

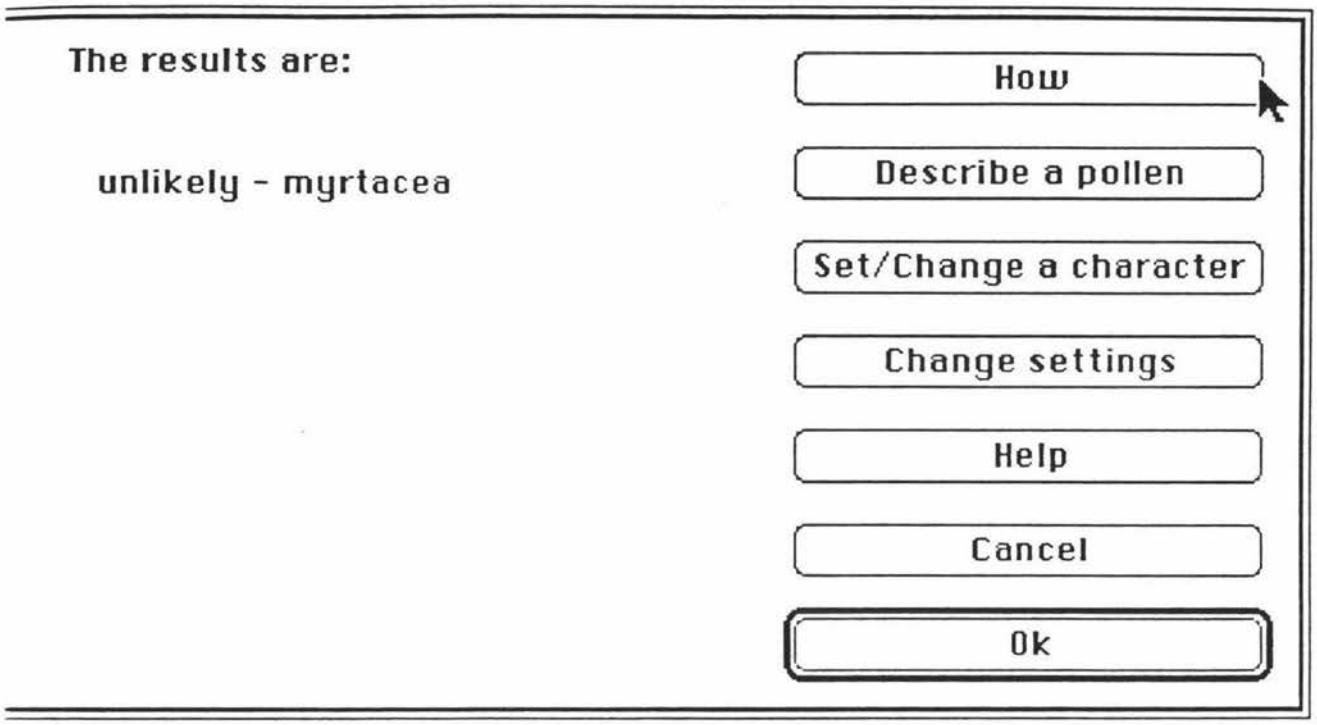


Figure 9: Result dialog.

The above example illustrates a disadvantage in using a single-access monothetic mode: before an appropriate query is made, the user may need to work through a number of characters which do not apply to the specimen.

### 4.2.2 Multi-access monothetic mode

The multi-access monothetic mode is primarily designed for users who have some identification experience in the domain of the system. It allows the user to describe the specimen using a sequence of characters chosen by the user. The system offers a list of all characters in the system, and allows the

user to choose one or more characters. For each character chosen, the system then presents a list of states so that one or more states which match the specimen can be selected. In Figure 10, the user has selected 'shape' as the character he/she wishes to describe.

The system displays the possible states for the character chosen, and allows the user to select one or more which describe the specimen. The state 'triangular' has been selected in Figure 11.

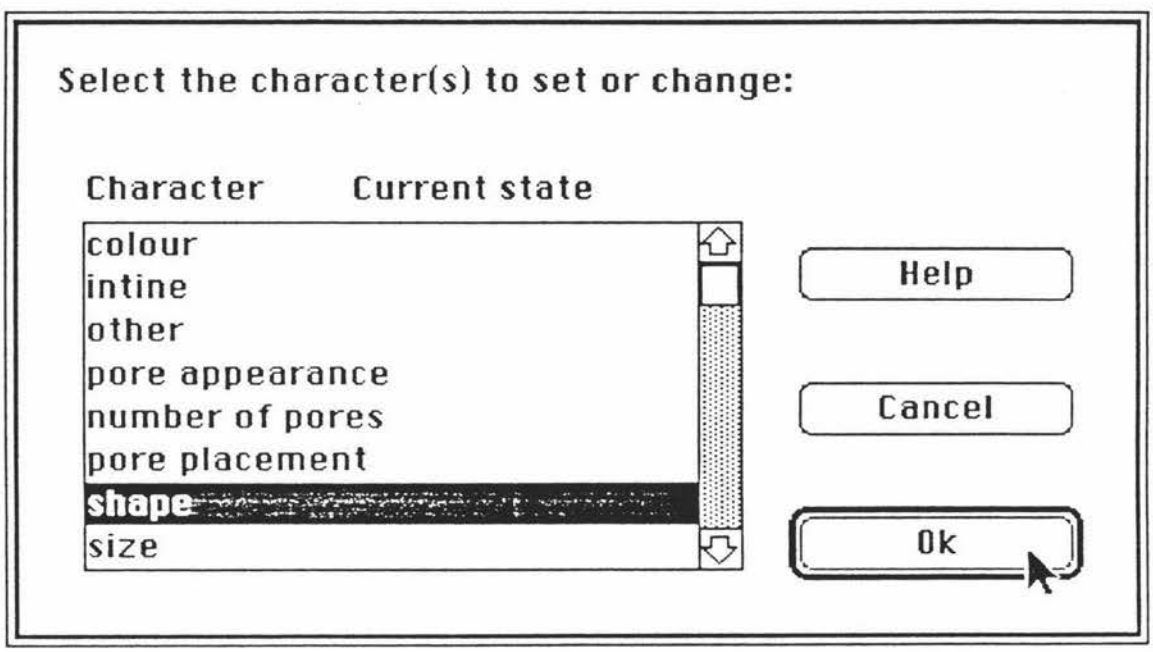


Figure 10: Character selection dialog.

As in the previous example, the system is able to identify the specimen as a myrtacea (Figure 12). This example also illustrates the advantage of the multi-access monothetic mode over the single-access monothetic mode; that is, the user can select characters which apply directly to the specimen and thereby possibly reduce time taken to complete the identification sequence.

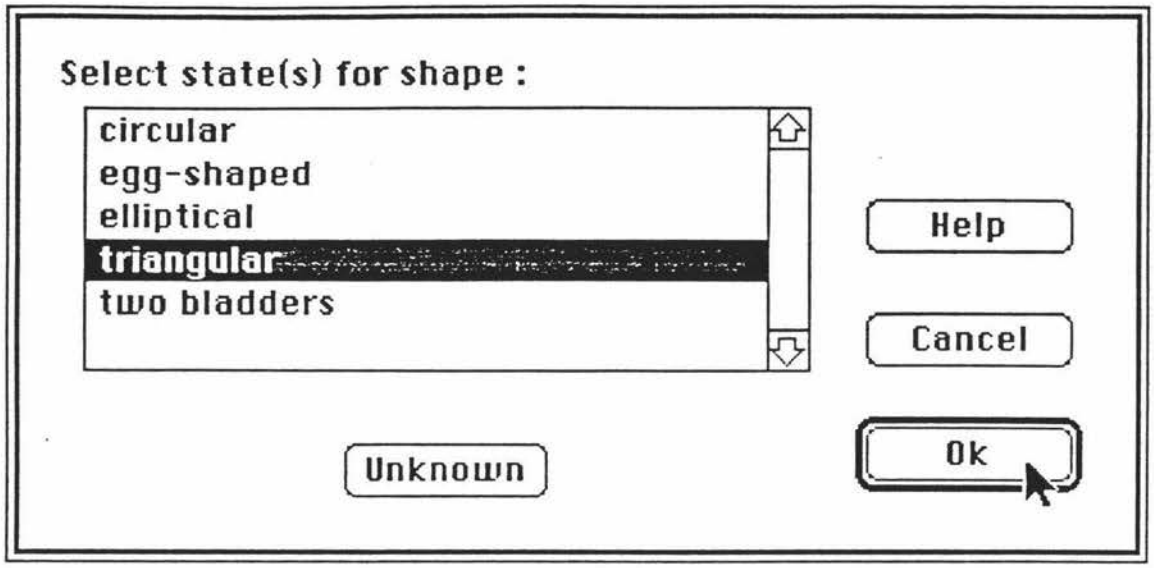


Figure 11: Shape dialog.

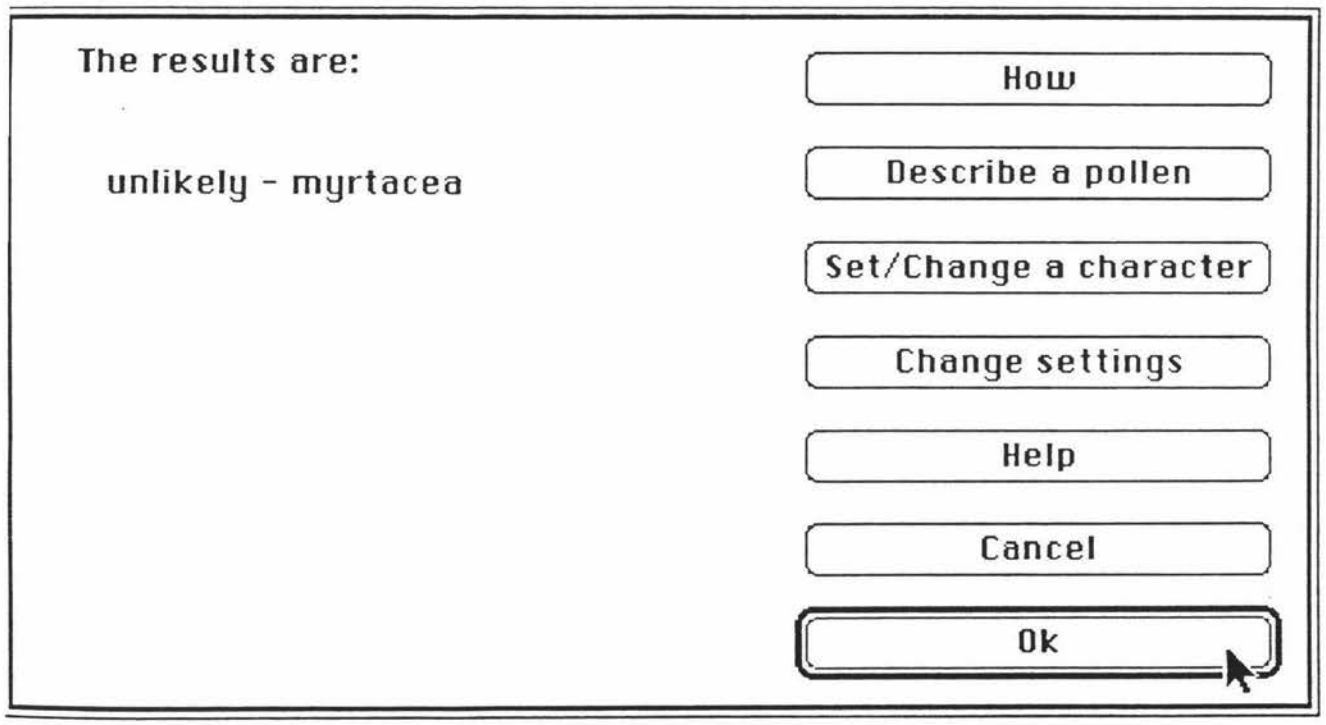


Figure 12: Result dialog.

### 4.2.3 Mixed single-access and multi-access monothetic modes

During an identification process, the user can move freely between single-access and multi-access monothetic modes. The multi-access monothetic mode

can be used to narrow down an identification, then the single-access monothetic mode used to complete the identification, as in the following example. In Figure 13, the user has selected 'number of pores' for description.

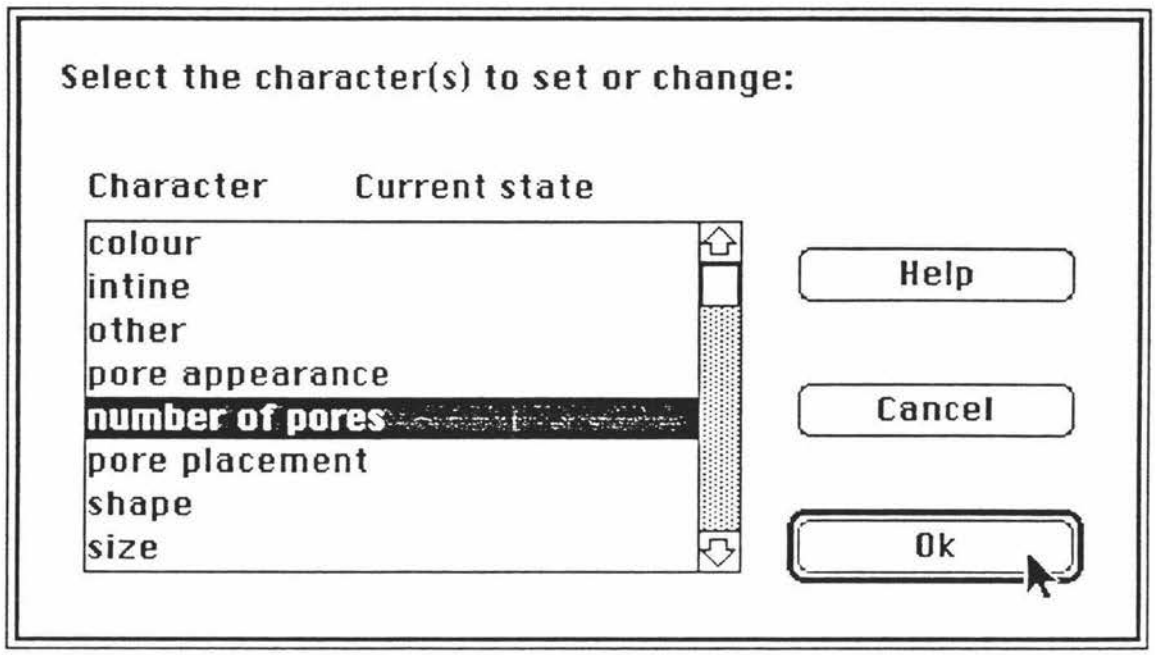


Figure 13: Character selection dialog.

The system queries the user about the number of pores on the specimen: 4 in this example (Figure 14).

The system then displays the pollens remaining, and allows the user to select 'Set/Change a character' to continue in multi-access monothetic mode, or to select 'Continue' in the single-access monothetic mode as shown in Figure 15. The ranking is based on the likelihood of the specimen being identified as the target taxon.

The user is then queried about a character which will distinguish the remaining taxa, 'pore placement' in this example (Figure 16). The user has selected 'At corners'.

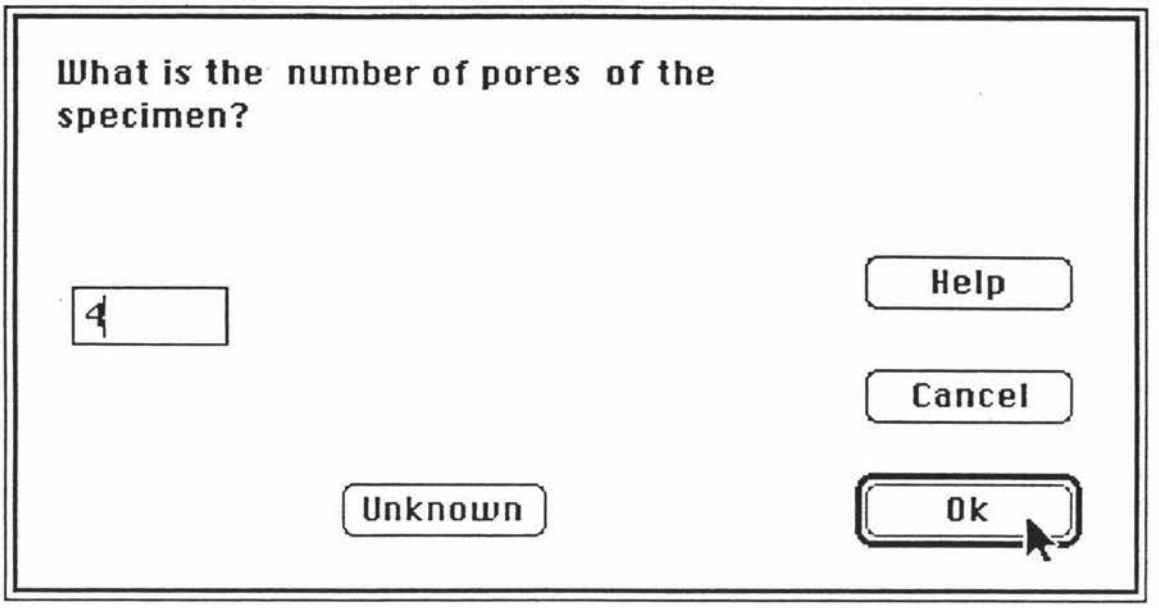


Figure 14: Pore number dialog.

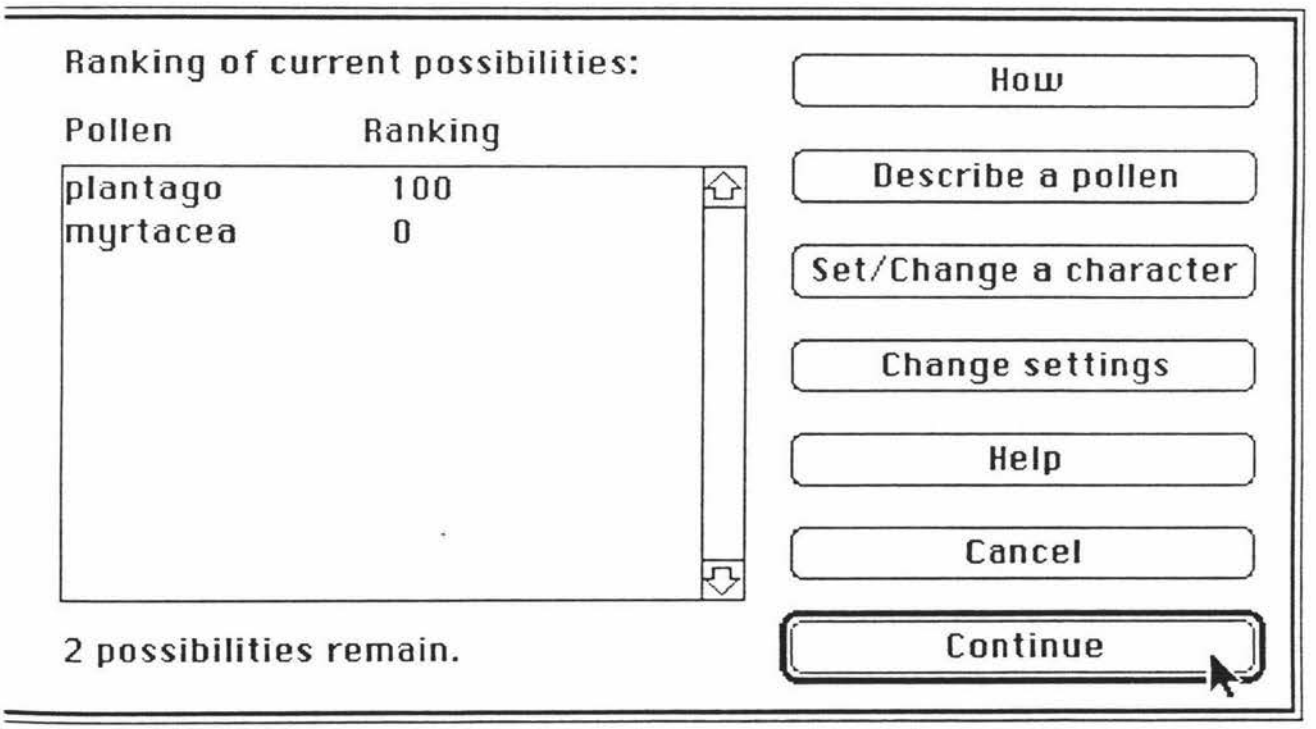


Figure 15: Remaining taxa dialog.

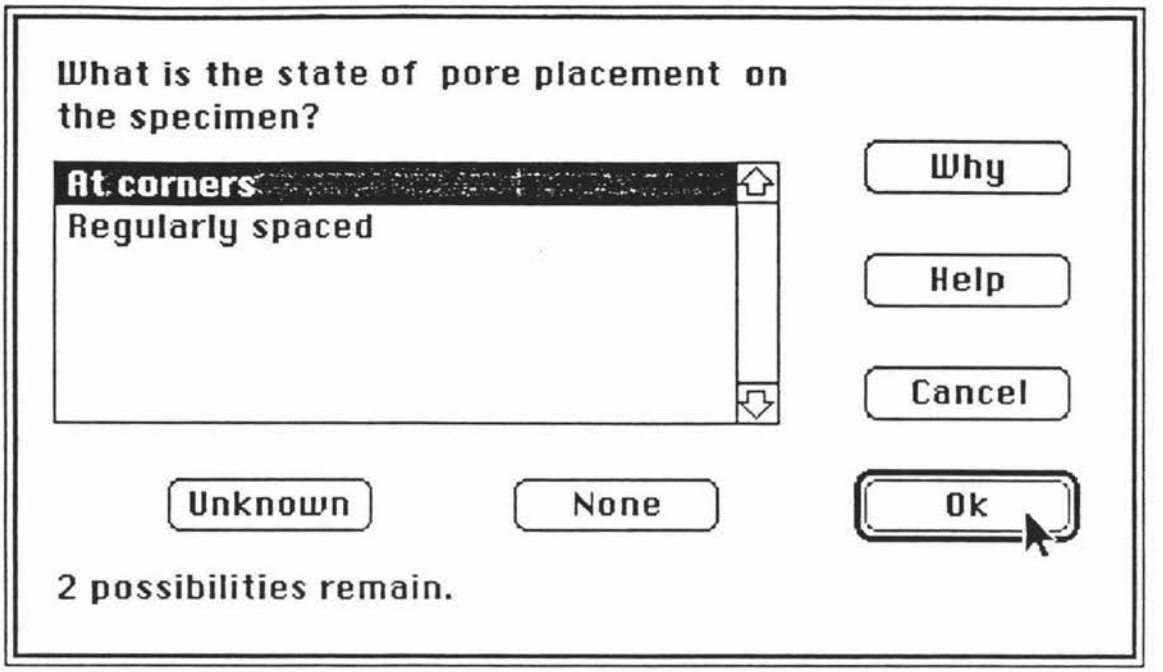


Figure 16: Pore placement dialog.

The specimen has again been identified as a myrtacea (Figure 17).

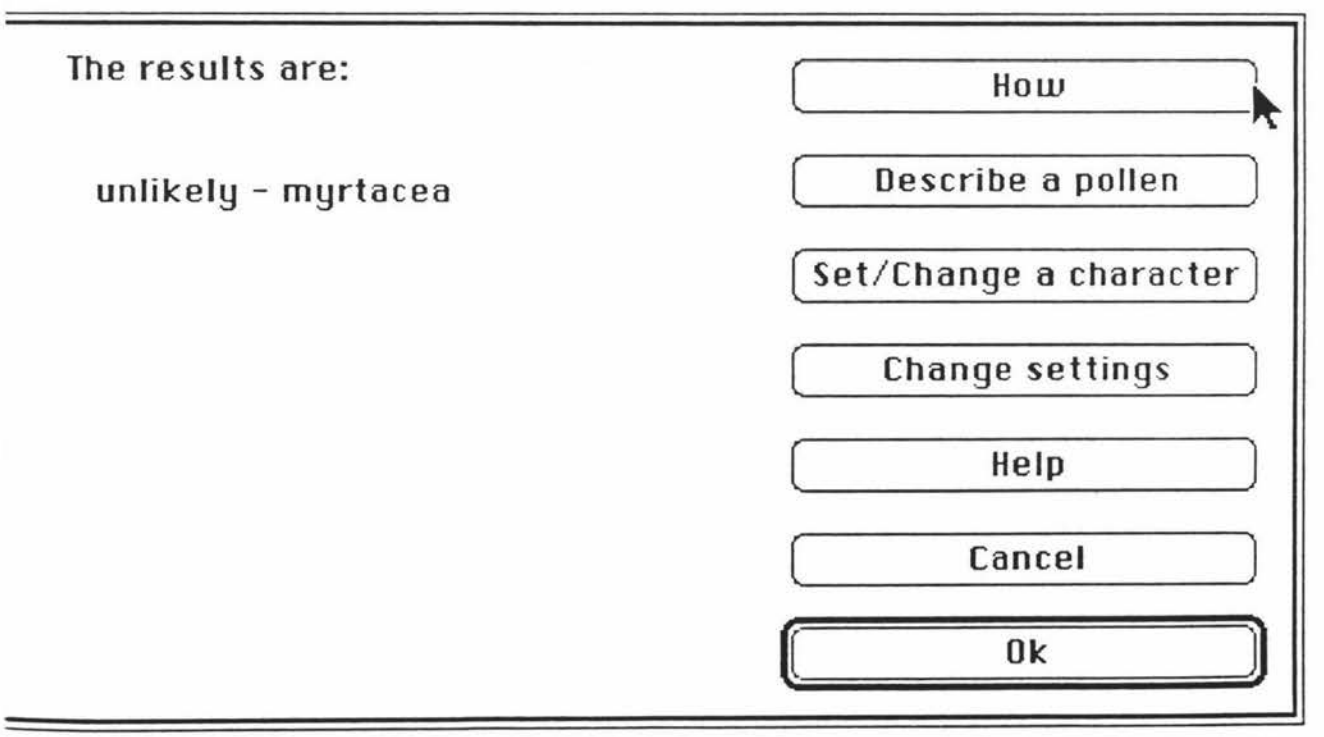


Figure 17: Result dialog.

### 4.2.4 Explanation facilities

During an identification using the single-access monothetic mode, the user can request an explanation for why the system is asking about a particular character. The system displays all taxa which have the character specified in their description, indicating those which will be accepted or rejected depending on whether their states match that given by the user. Figure 18 is taken from the single-access monothetic mode section of the mixed single-access and multi-access monothetic mode identification in Section 4.2.3. It shows that the specimen will be identified as a plantago if the state for 'pore placement' is 'Regularly spaced', or as a myrtacea if the state is 'At corners'.

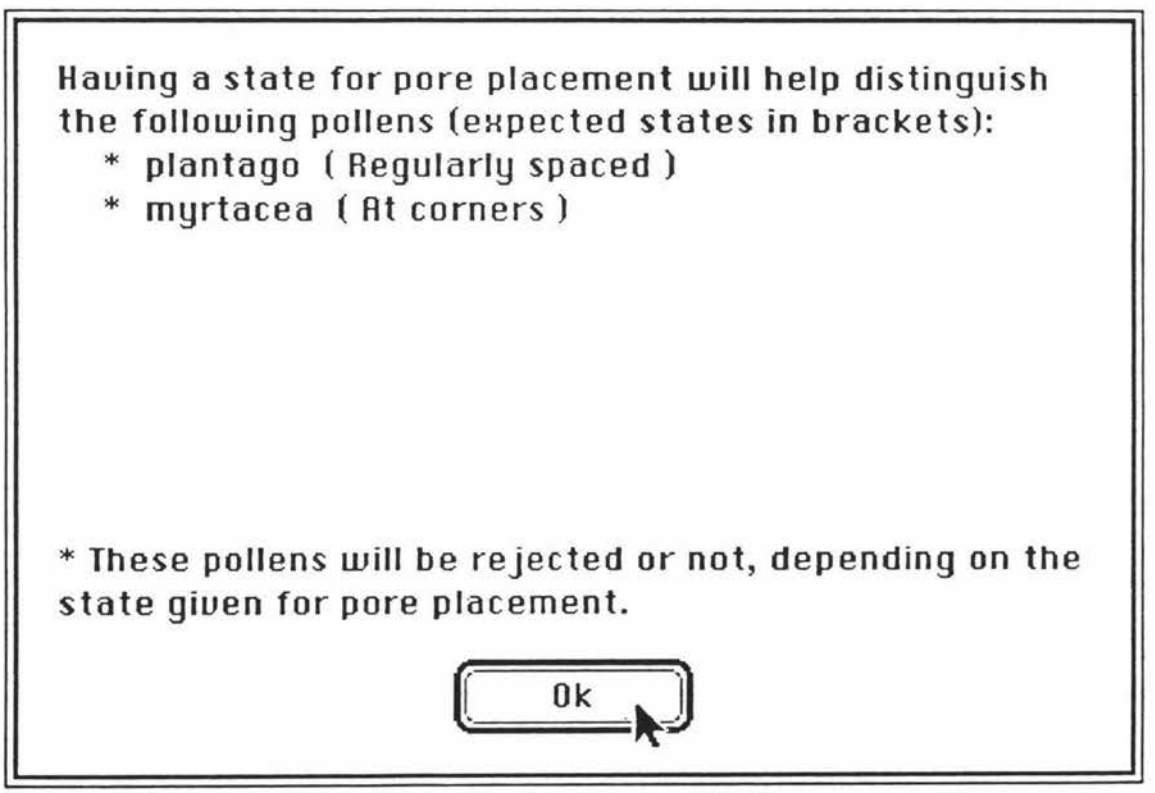


Figure 18: Explanation dialog.

At the end of the identification sequence, the remaining taxa are rank ordered, depending on how well they had matched the specimen and on the

likelihood of the specimen being identified as the target taxon. For example, the present pollen identification system takes into account geographical location and season when ranking taxa.

At any stage during the identification sequence, the user can request a display of all taxa in the system, or of the taxa rejected during the identification process. A detailed description of the character states for selected taxa can be requested. This includes an indication of the character states which match or do not match the description of the specimen. Figures 19 and 20 show the description of myrtacea and plantago at completion of the mixed single-access and multi-access monothetic mode identification example shown in Section 4.2.3. Note the '+' and '-', indicating state matches and mismatches with the description of the specimen.

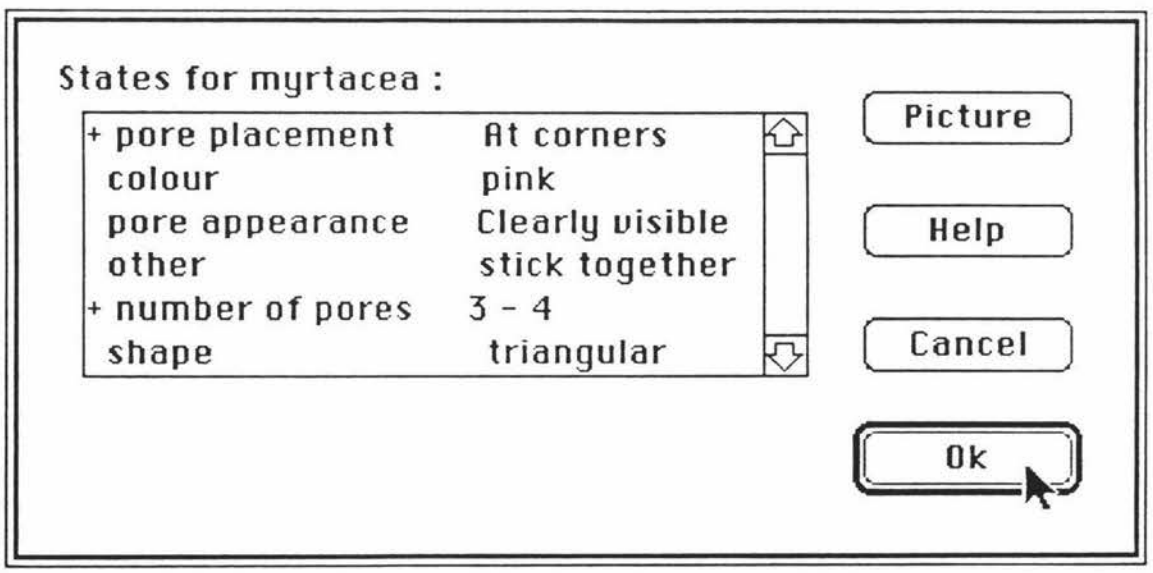


Figure 19: Description dialog.

A report on the characters used in an identification sequence, and the taxa which were accepted or rejected can also be requested. Figure 21 is also taken

from the completion of the mixed single-access and multi-access monothetic mode identification shown in Section 4.2.3.

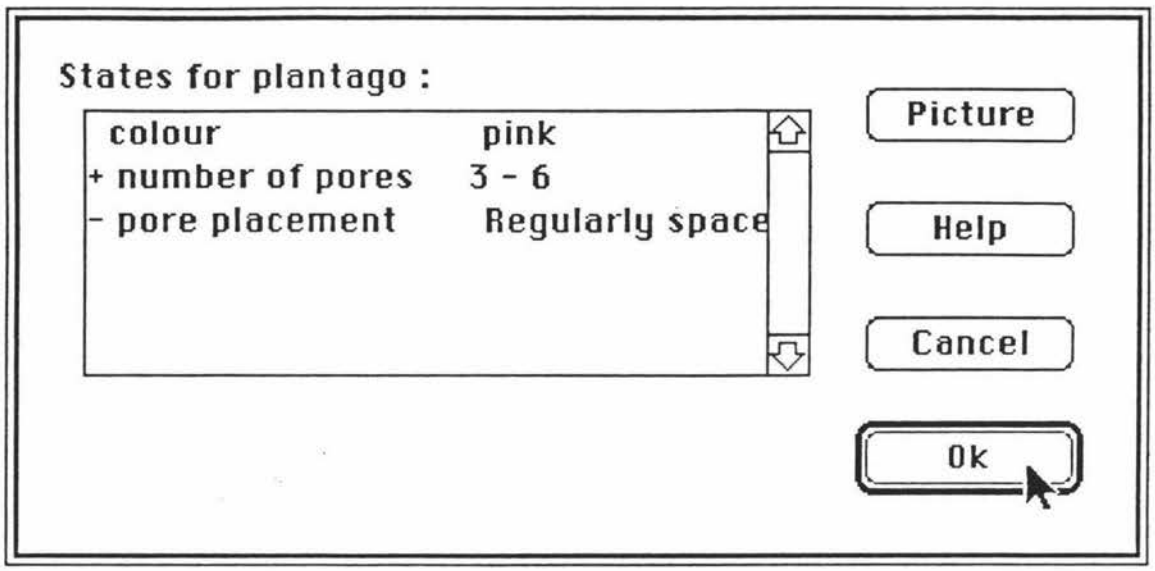


Figure 20: Description dialog.

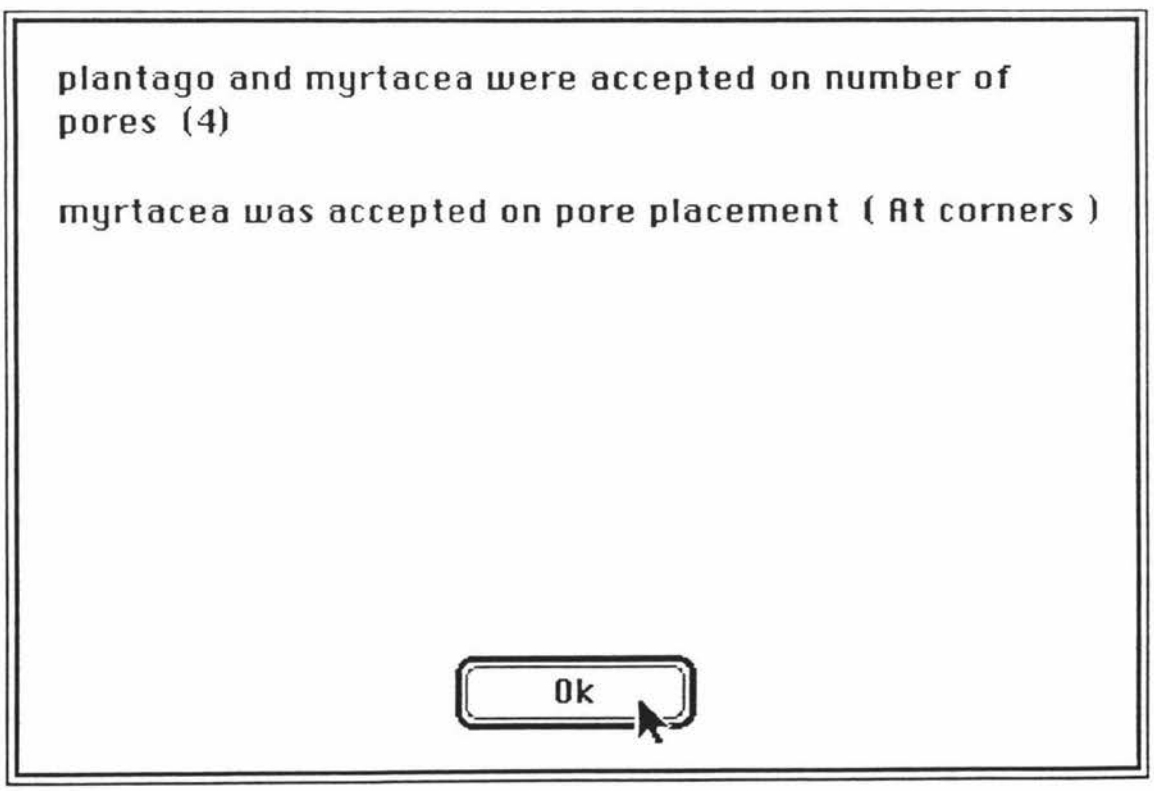


Figure 21: Report dialog.

### 4.2.5 Other features

At any stage during an identification process, the user can request a display of the characters used by the system. This includes a description of the state or states currently assigned to each character. The state of one or more characters can then be changed, allowing errors to be corrected. Figure 22 shows the characters used during the mixed single-access and multi-access monothetic identification process in Section 4.2.3, and the states assigned to them.

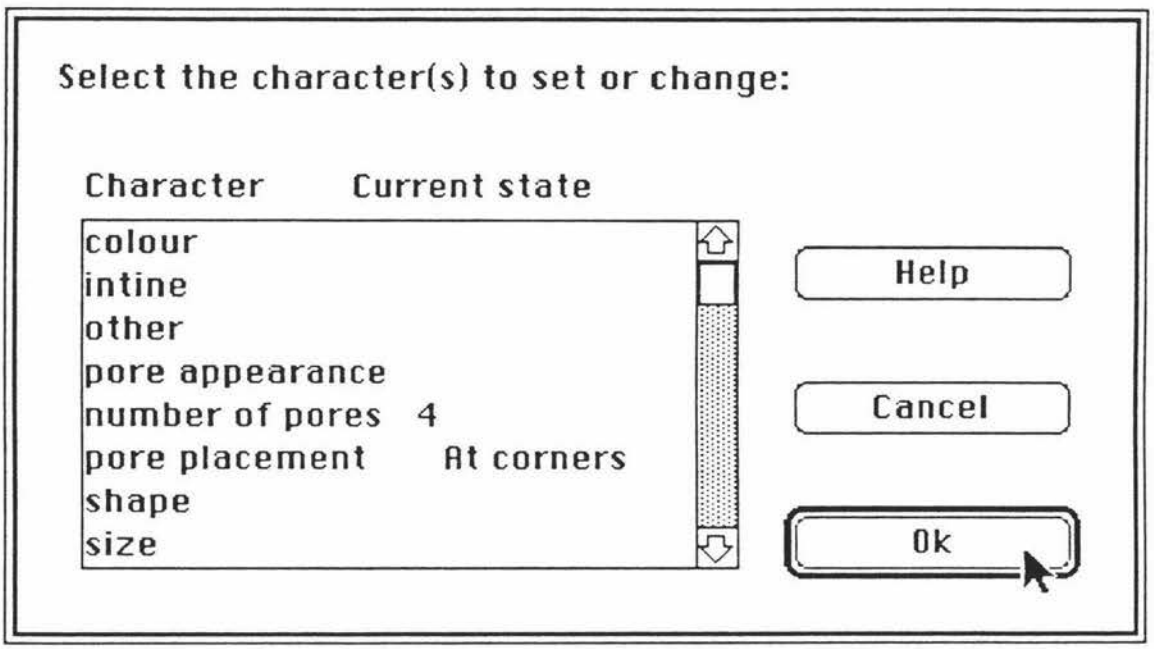


Figure 22: Character description dialog.

Domain experts have indicated that probabilities are very rarely used during identification, therefore the present system does not require the user to specify their certainty about a particular character state, thus maintaining the simplicity of the user interface. As single-access monothetic mode users are not expected to be experienced, the 'unknown' and 'none' choices provide them with a simple method for expressing uncertainty about a character state.

Users of the multi-access monothetic mode are expected to select only those character states about which they are certain.

## 4.3 Knowledge representation

The main objective of the knowledge base design was that domain experts should be able to easily understand the knowledge and structure of the knowledge, in the knowledge base. It was also designed to be easily extended and modified. These objectives were met using a method of ranking characters and use of a taxonomic data matrix as the basis for the knowledge base structure.

### 4.3.1 Character ranking

In order to assist the system with identifications, and with the ordering of questions in the single-access monothetic mode, importance values are assigned to character states. These importance values are supplied by the domain expert and provide an indication of the importance of each character state to identifying the taxon.

There are four categories of importance values:

- essential: the character state must be present in order for the taxon to be identified. Characters with an essential rank are regarded as key characters and are normally easily observed for the taxon in question. For example, 'triangular' is an essential state for the character 'shape' when identifying a myrtacea pollen.
- medium importance: the character state is normally present in the taxon, and can be used to support an identification.

- low importance: the character state is generally present in the taxon, but is not normally used for identification as many other taxa may also have that character state.
- no importance: the character state is almost always present in the taxon, therefore it is not needed for identification. Use of this category ensures that a complete description of the taxa is obtained.

### 4.3.2 Knowledge base structure

The knowledge base is structured on a taxonomic data matrix. This is the traditional method of organising taxonomic data, making the knowledge base easily understood by a domain expert and allowing simple conversion to both rules and frames. It also assists the conversion of existing bodies of biological knowledge to the knowledge base format, allowing the system to be used in many fields of biological identification. The matrix is formed with a taxa by character table. Each entry in the table represents the state that the taxa exhibits for the character, and is empty if the taxa does not have a state for that character.

The importance values of the characters are stored as an appendage to the matrix. For each taxon, there are three lists to which characters are assigned according to their importance value. The 'no importance' category is not specified in the appendage; it is assumed that characters not specified as essential, medium or low importance are not used in the identification. A generalised matrix and appendage may be seen in Figure 23, and Figure 24 shows an example using a section of the matrix which forms the pollen knowledge base.

| | | Taxa | | | |
|---|---|---|---|---|---|
| | $T_1$ | $T_2$ | ... | $T_n$ | |
| Characters | $C_1$ | $S_{1,1}$ | $S_{1,2}$ | ... | $S_{1,n}$ |
| | $C_2$ | $S_{2,1}$ | $S_{2,2}$ | ... | $S_{2,n}$ |
| | $C_3$ | $S_{3,1}$ | $S_{3,2}$ | ... | $S_{3,n}$ |
| | $C_4$ | | $S_{4,2}$ | ... | |
| | $C_5$ | | $S_{5,2}$ | ... | |
| | $C_6$ | | $S_{6,2}$ | ... | |
| | ... | ... | ... | ... | ... |
| | $C_m$ | $S_{m,1}$ | $S_{m,2}$ | ... | $S_{m,n}$ |
| Essential | | $C_1$ | $C_1$ | ... | $C_1$,$C_2$ |
| Medium importance | | $C_2$ | $C_2$,$C_3$ | ... | $C_3$ |
| Low importance | | $C_3$ | $C_4$,$C_5$,$C_6$ | ... | |

Figure 23: Generalised taxonomic matrix.

To facilitate use of the knowledge in the knowledge base in various systems, it can be easily converted to a rule-based or frame-based format. The present system includes a knowledge conversion facility, which converts from the matrix-based knowledge base to the simple frame-based format used in the inference engine. Another conversion facility could be easily created to convert the knowledge to a rule-based format if required.

The knowledge in the knowledge base can be easily converted to a rule-based structure by using each taxon as the consequent, and the characters and states as antecedents. Figures 25 and 26 display example rules, derived from

the matrices in Figures 23 and 24. Note the separation of the antecedents into essential, medium and low importance categories.

| | | Pollen | |
|---|---|---|---|
| | | Myrtacea | Plantago |
| Characters | Size | | |
| | Shape | Triangular | |
| | Surface appearance | | |
| | Intine appearance | | |
| | Pore number | 3 - 4 | 3 - 6 |
| | Pore placement | At corners | Regularly spaced |
| | Pore appearance | Clearly visible | |
| | Furrow number | | |
| | Furrow appearance | | |
| | Colour | Pink | Pink |
| | Other features | Stick together | |
| Essential | | Shape | Pore number |
| | | Pore number | Pore placement |
| | | Pore placement | |
| | | Pore appearance | |
| Medium importance | | Other features | |
| Low importance | | Colour | Colour |

Figure 24: Section of the taxonomic matrix forming the pollen knowledge base.

taxa is $T_1$ if:

essential: ($C_1 = S_{1,1}$) and

medium: ($C_2 = S_{2,1}$) and

low: (C3 = S3,1).

taxa is $T_2$ if:

essential: ($C_1 = S_{1,2}$) and

medium: ($C_2 = S_{2,2}$ and

C3 = S3,2) and

low: (C4 = S4,2 and

C5 = S5,2 and

C6 = S6,2)

...

taxa is $T_n$ if:

essential: ($C_1 = S_{1,n}$ and

$C_2 = S_{2,n}$) and

medium: ($C_3 = C_{3,n}$).

Figure 25: Example rules derived from the matrix in Figure 6.

A simple frame-based structure can be easily created from the knowledge base by considering each column of the matrix as a frame describing a taxon. The importance values from the matrix remain as an appendage to the frame. Figures 27 and 28 display frames formed from the matrix shown in Figure 23 and 24.

```
pollen is myrtacea if:
    essential:   (Shape = Triangular and
                  Pore number = 3 - 4 and
                  Pore placement = At corners and
                  Pore appearance = Clearly visible) and
    medium:      (Other features = stick together) and
    low:         (Colour = Pink).

pollen is plantago if:
    essential:   (Pore number = 3 - 6 and
                  Pore placement = Regularly spaced) and
    low:         (Colour = pink)
```

Figure 26: Example rules derived from the matrix in Figure 7.

| Taxa | Tn |
|---|---|
| $C_1$ | $S_{1,n}$ |
| $C_2$ | $S_{2,n}$ |
| ... | ... |
| $C_m$ | $S_{m,n}$ |
| Essential | $C_1,C_2$ |
| Medium importance | $C_3$ |
| Low importance | |

Figure 27: Example frame formed from generalised matrix in Figure 6.

| Pollen | Myrtacea |
|---|---|
| Size | |
| Shape | Triangular |
| Surface appearance | |
| Intine appearance | |
| Pore number | 3 - 4 |
| Pore placement | At corners |
| Pore appearance | Clearly visible |
| Furrow number | |
| Furrow appearance | |
| Colour | Pink |
| Other features | Stick together |
| Essential | Shape |
| | Pore number |
| | Pore appearance |
| | Pore placement |
| Medium importance | Other features |
| Low importance | Colour |

Figure 28: Example frame formed from the example matrix in Figure 7.

## 4.4 Inference Engine

In single-access monothetic mode, the inference engine is used to select characters about which to query the user and to accept or reject taxa. In multi-access monothetic mode, the user chooses the characters, and the inference engine is used to accept or reject taxa. The algorithm used to accept or reject taxa is the same for both modes.

### 4.4.1 Character selection

The system first selects the most important character for the candidate taxa most likely to be the specimen and then queries the user about that character. Character selection is conducted by sorting candidate taxa according to the likelihood of the specimen matching the taxa. For example, the pollen identification system uses the likelihood of finding a pollen in a given geographical location and season. From the sorted taxa, a list of characters is ordered using the importance values assigned to the characters for each taxa. The list of characters is then searched for the first character which has not been described by the user, ensuring that the system continually attempts to accept or reject the most likely taxa. A structure diagram of the above method may be seen in Figure 29.

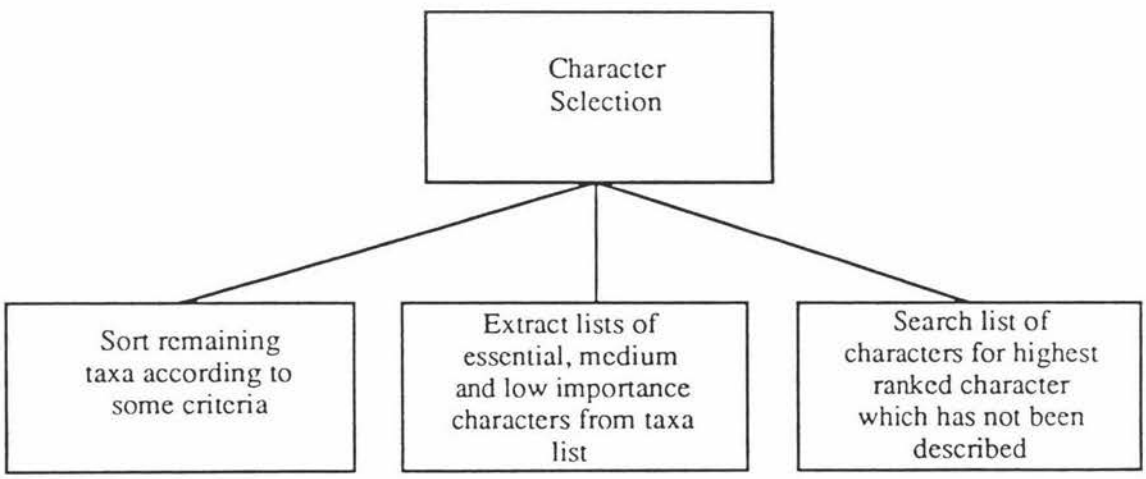


Figure 29: Structure diagram of the method used for character selection.

### 4.4.2 Taxa acceptance or rejection

The main section of the inference engine is concerned with the acceptance or rejection of taxa. Accepted taxa are those being retained for the next stage

of the identification; rejected taxa are those which will not be considered for the remainder of the identification process.

The inference engine initially uses essential characters to accept or reject taxa. This allows the system to quickly narrow down an identification using obvious and easily observed characters. Essential characters are normally sufficient to identify the specimen. However, if all essential characters have been used and more than one taxon remains, medium, and if necessary, low importance characters are used to identify the specimen. Identification is complete when all characters have been used, all taxa have been rejected, or the remaining taxa have the same states for the characters remaining to be described (ensuring that identification stops if two or more taxa remain that are identical in all remaining character states). Figure 30 displays a structure diagram of the taxa acceptance and rejection procedure.

#### 4.4.2.1 Essential characters

When the user describes a character state the system searches the list of possible taxa for those which have the character specified as being essential to identification. If any of the taxa found have a character state which matches the specimen, those taxa are accepted and the identification process continues with that set of taxa. If no taxa with a matching character state are found, all taxa with the character specified are rejected and the identification process continues with those taxa that remain.

If the user selects the category 'Unknown' as an answer to the query, no taxa are accepted or rejected. If the category 'None of the possibilities' is selected, the inference engine rejects all taxa which have that character in the essential set.

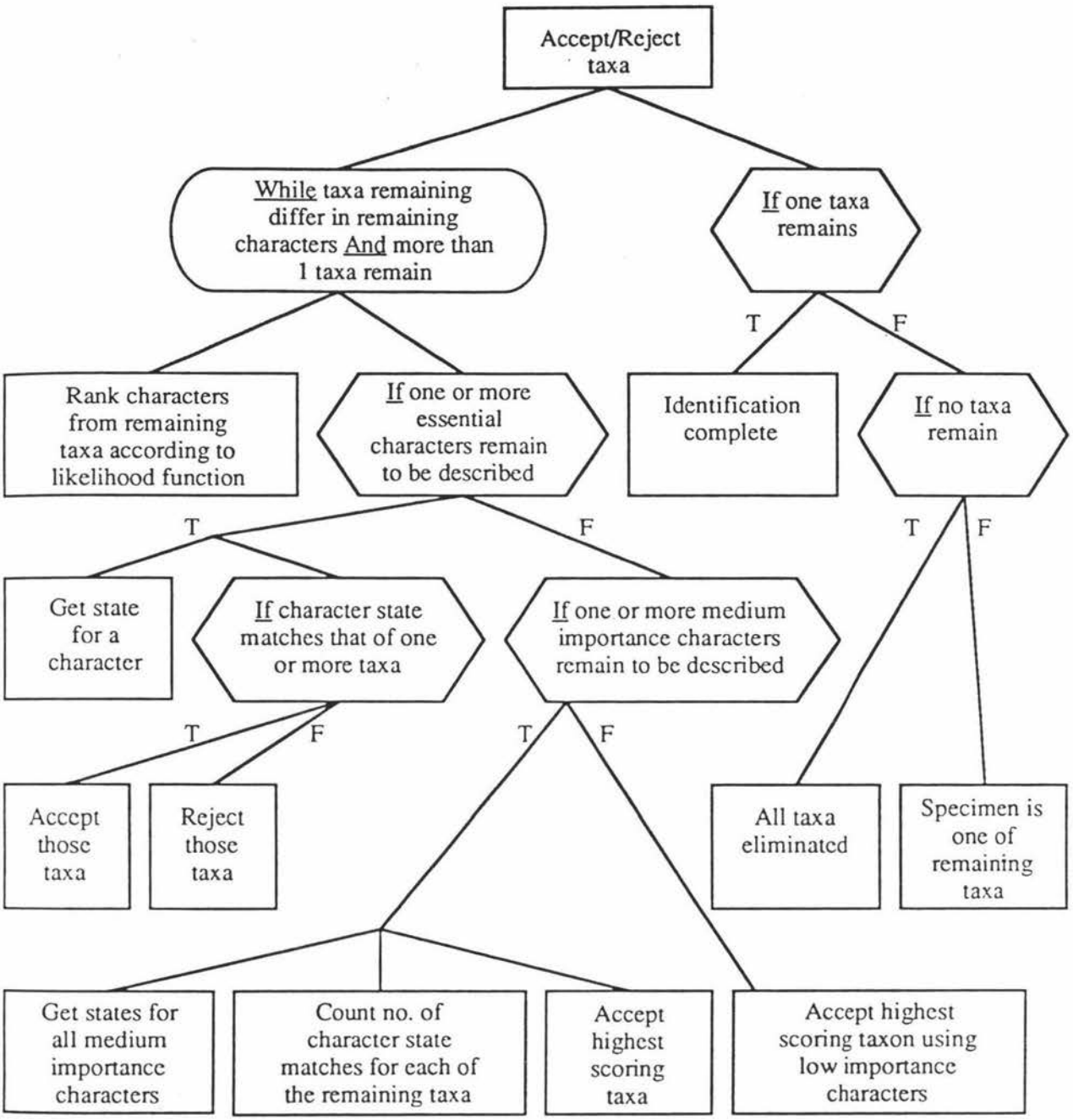


Figure 30: Structure diagram of the taxa acceptance and rejection procedure.

For example, an inference engine using a knowledge base containing just the pollens myrtacea and plantago as defined in Figure 24, will accept and reject the pollens as shown in Figure 31.

| Input: Character / State | Result |
|---|---|
| Pore placement / At corners | Accept myrtacea: identification complete |
| Pore placement / Regularly spaced | Accept plantago: identification complete |
| Pore placement / Unknown | Continue with all pollens |
| Pore placement / None | Reject myrtacea and plantago: continue with pollens remaining |

Figure 31: Table showing pollens accepted or rejected according to various inputs.

#### 4.4.2.2 Medium importance characters

Medium importance characters are used after all essential characters have been described, if more than one taxa remains. The user is queried about all medium importance characters taken from the remaining taxa, and the taxon with the greatest number of matches is selected.

#### 4.4.2.3 Low importance characters

If all essential and medium importance characters have been used, and more than one taxa remains (e.g., if there were an equal number of matching medium importance characters), then the low importance characters are used. The user is queried about all low importance characters from the remaining taxa, and the taxon with the greatest number of matches is the selected taxa. If two or more taxa have an equal number of matches using low importance

characters the system displays all of these, and describes the probability of finding each taxa using the established criteria for ordering taxa. Low importance characters are rarely used, as essential and medium importance characters are generally sufficient to identify a specimen.

# Chapter 5

# Conclusion

## 5.1 Realisation of design goals

The present report details an expert system which has been designed and built in order to assist both experts and non-specialised users in biological identification. The system assists non-specialised users in identification by leading them through the process via a single-access monothetic mode. Expert users can describe the specimen to the system via a multi-access monothetic mode. If candidate taxa cannot be differentiated by characters remaining to be described, the system will recognise this.

The knowledge base of the system can be easily updated in order to add taxa, or changed to provide identifications in other fields of biology. A subsystem was created to facilitate this process, an example of which can be seen in Appendix D.

During the final stages of development of the system, it was demonstrated to experts in the field of biological identification at a Workshop on Computers in Taxonomy and Systematics hosted by the Botany Department of Massey University and was received very favourably.

## 5.2 Future development

Possibilities for future development of the system include automatic taxa identification and the development of a tutorial subsystem. Automatic taxa identification would use the present system as a pattern classification subsystem in a pattern recognition process. Imaging hardware and software would be used to take an image of the specimen, and a pattern recognition

process would attempt to identify the specimen. If the appropriate hardware were available at a reasonable cost pattern recognition could be extended to real-time catching, identification and counting of taxa such as air-borne pollen grains.

A tutorial subsystem could be developed which would provide the user with a graphical display of a taxon, and ask the user to for identification. If the user's response is incorrect, the subsystem would emphasise the main characters until the user can identify the taxon. This subsystem would be simple to construct using the structure of the knowledge base developed in the present study.

## 5.3 Summary

The system developed in the present study has shown that expert system techniques are appropriate to the domain of biological identification. Methods of ranking characters and use of a modified taxonomic data matrix as the knowledge base structure have been developed. The method of ranking characters is used during single-access monothetic identification processes in order to select the most important character to query at each stage, and is used during both single-access and multi-access monothetic identification processes to reject or accept candidate taxa. The use of essential characters is very similar to the method used by domain experts to manually identify taxa.

The use of a modified taxonomic data matrix to form the structure of the knowledge base will assist domain experts to assess the accuracy of the knowledge base contents and make modifications as necessary. Use of a modified form of the traditional taxonomic data matrix allows existing taxonomic data to be easily incorporated in an expert system.

Also included in the present system are two methods for identifying taxa. The single-access monothetic method was found to be useful for users needing to be led through an identification as the system selects the characters and the order in which they are to be described. The multi-access monothetic method was found to be useful in assisting expert users to assess the accuracy of their identifications by allowing them to describe character states in any order.

# Appendix A

## Example session using single-access monothetic prototype

The ordering of questions is based on the conceptual tree shown in Figure 2 (Chapter 3). Note that although some questions appear to have more than one answer, the prototype is based on a dichotomous key with multiple choices shown conceptually as levels of a tree (Figure A.1).

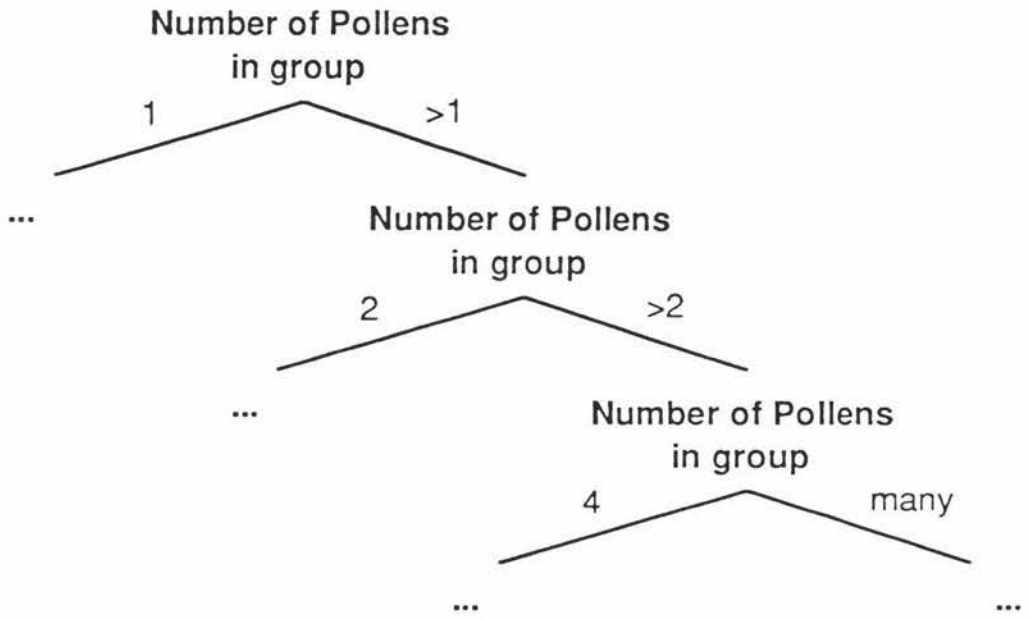


Figure A.1: Method of providing multiple choices within a dichotomous key.

The user is initially queried about the number of pollen grains which are in the group being studied (Answer: 1) (Figure A.2).

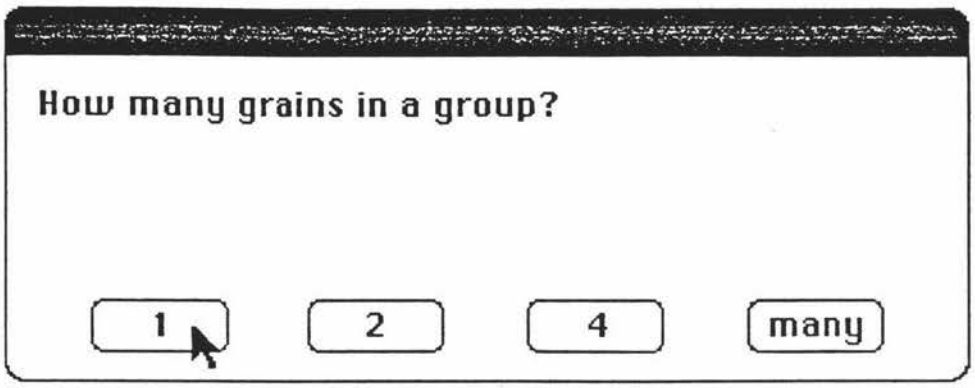


Figure A.2: Group size dialog.

The system then asks a question regarding the shape and surface appearance of the grain (Answer: No wing-like extensions or warty surface) (Figure A.3).

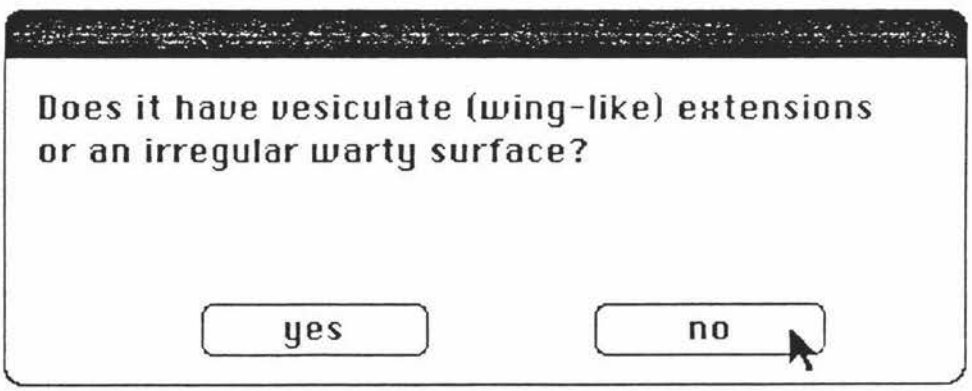


Figure A.3: Vesiculate dialog.

Next, a further question is asked about the surface appearance of the grain (Answer: No meridional ridges) (Figure A.4).

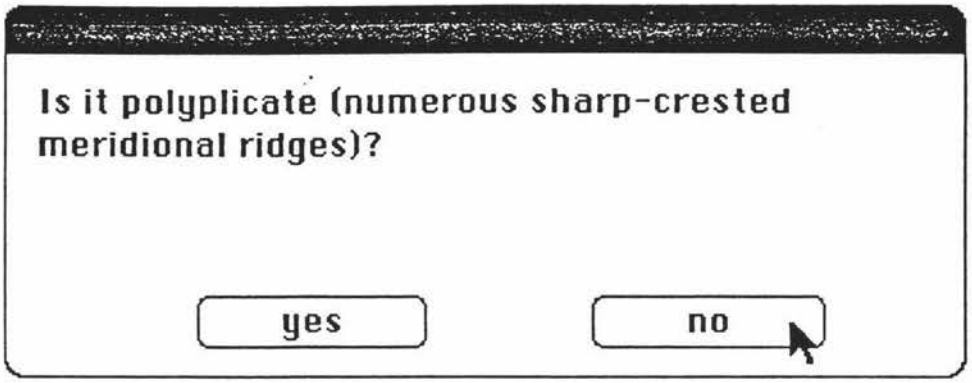


Figure A.4: Polyplicate dialog.

The user is then queried about the number of apertures on the grain (Answer: 0) (Figure A.5).

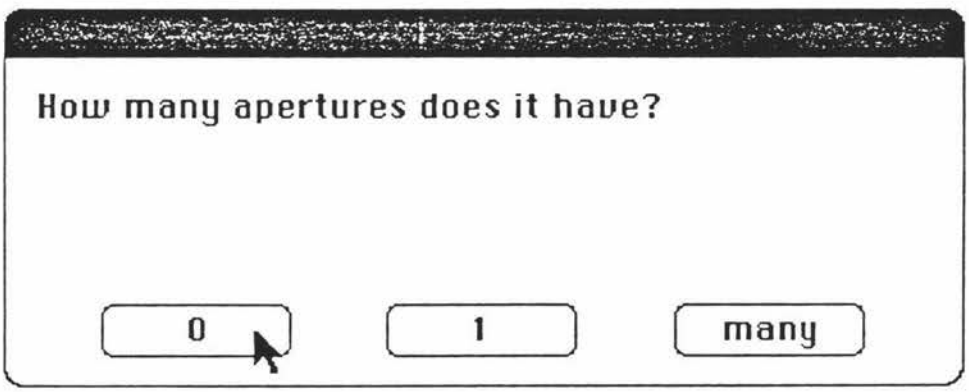


Figure A.5: Aperture number dialog.

Finally, the user is queried about the appearance of a trilete scar on the grain (Answer: Yes) (Figure A.6).

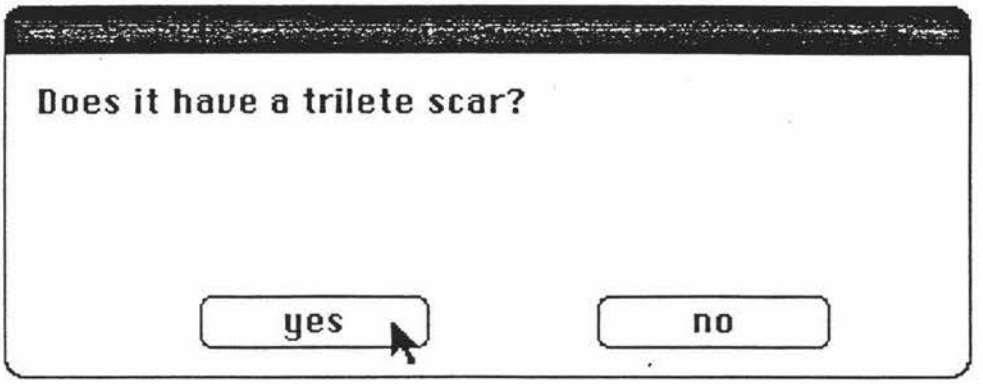


Figure A.6: Trilete scar dialog.

At this stage, the system has reached a leaf node of the conceptual tree, and displays the result of the session. The user can then exit the program or continue with another grain (Figure A.7).

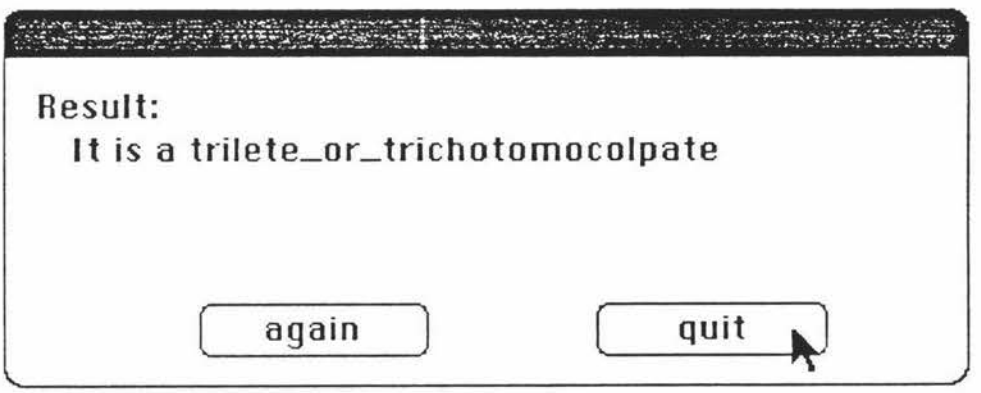


Figure A.7: Result dialog.

# Appendix B

## Example session using multi-access monothetic prototype

### B.1 Introduction

The multi-access monothetic prototype has two sections. One section allows the user to describe the specimen to the system in order to make an identification; the other allows the user to request a description of a pollen from the system.

### B.2 Description of specimen

The user is shown a list of characters, and asked to select the one he/she wishes to describe (Selection: Size) (Figure B.1).

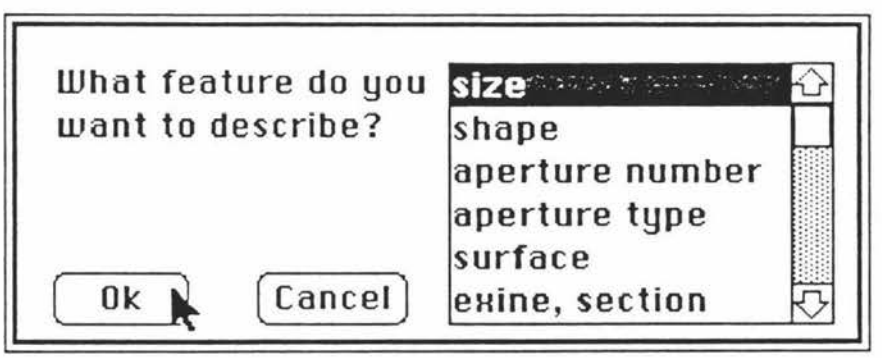


Figure B.1: Character selection dialog.

The user is then shown a list of the possible states for the character size, and asked to select one corresponding to the specimen (Selection: Medium) (Figure B.2).

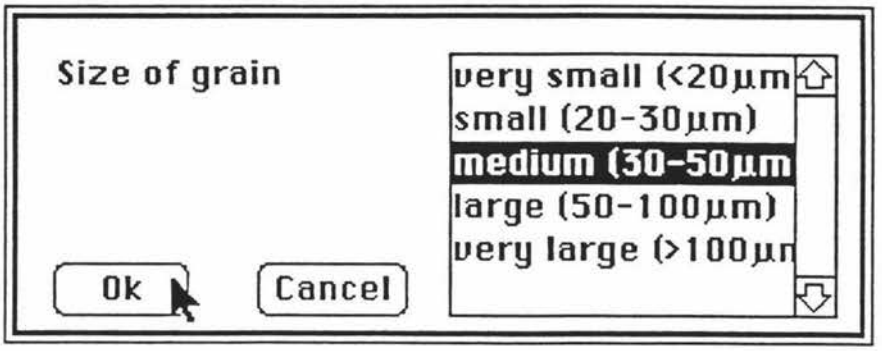


Figure B.2: Size dialog.

After a description of a character state of the specimen the system displays the identity of remaining pollens. If only one pollen remains identification is complete (Figure B.3).

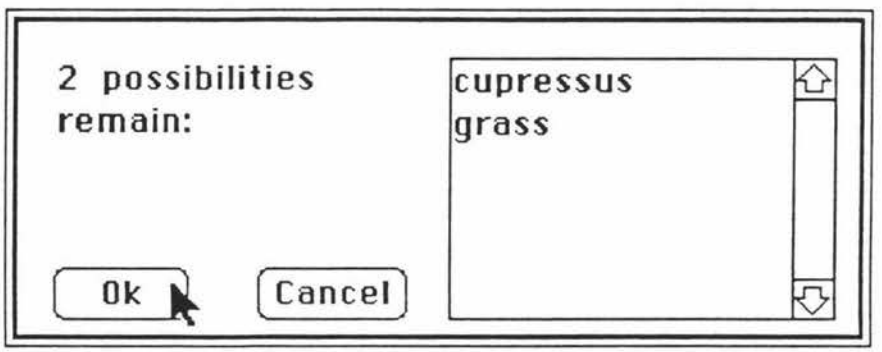


Figure B.3: Remaining pollens dialog.

Next, another list of characters is shown to the user, allowing he/she to select a character for description (Selection: Shape) (Figure B.4).

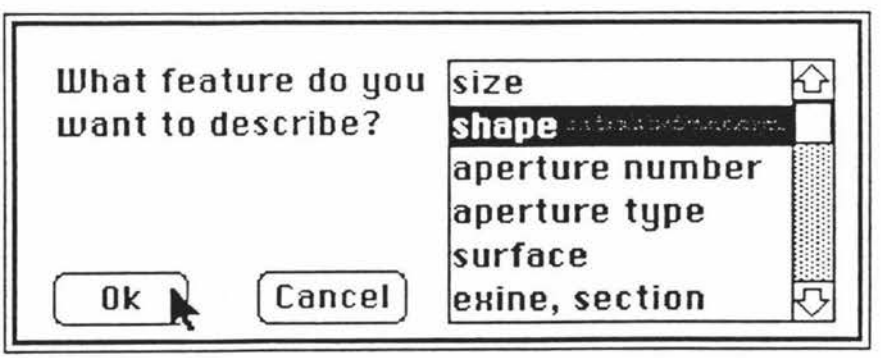


Figure B.4: Character selection dialog.

The user is then shown a list of the possible states for the character shape, and asked to select the one which corresponds to the specimen (Selection: Round or irregularly round) (Figure B.5).

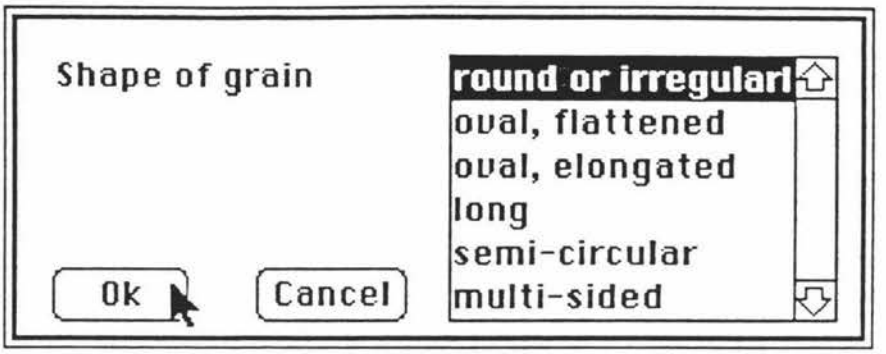


Figure B.5: Shape dialog.

Again, two possibilities remain (Figure B.6).

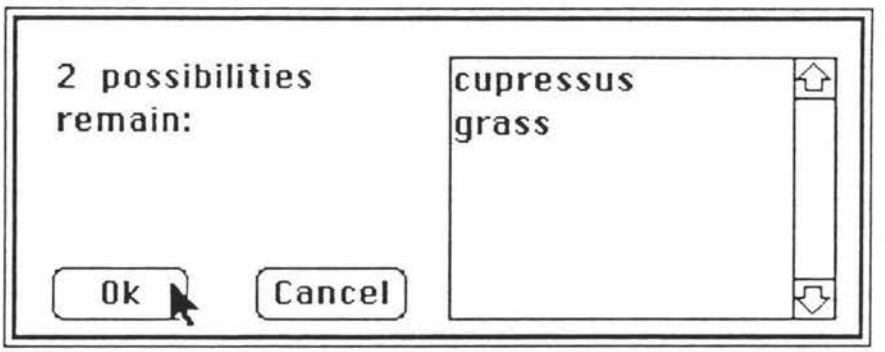


Figure B.6: Remaining pollens dialog.

Another list of characters is shown to the user, allowing them to select a character for description (Selection: Aperture number) (Figure B.7).

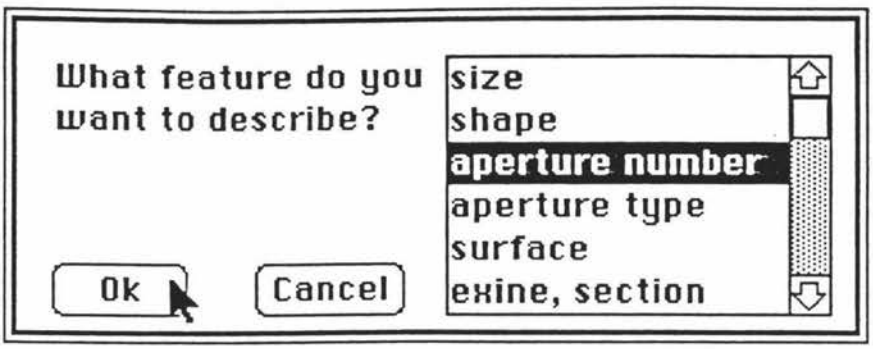


Figure B.7: Character selection dialog.

The user is shown a list of the possible states for aperture number, and asked to select the one which corresponds to the specimen (Selection: 1-2) (Figure B.8).

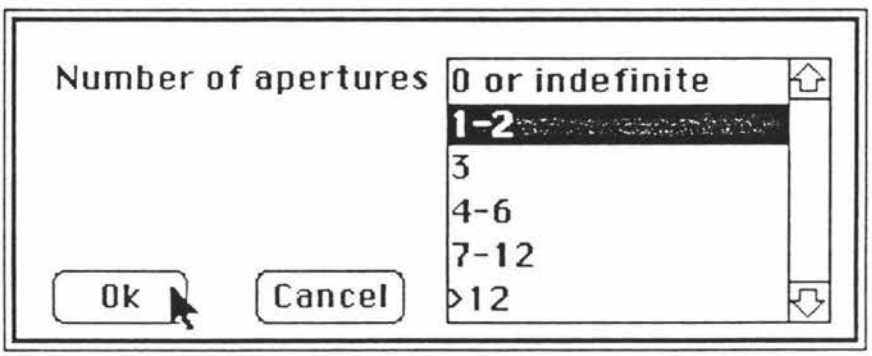


Figure B.8: Aperture number dialog.

Finally, only one pollen remains: the system displays the result, along with the character states which contributed to the identification (Figure B.9).

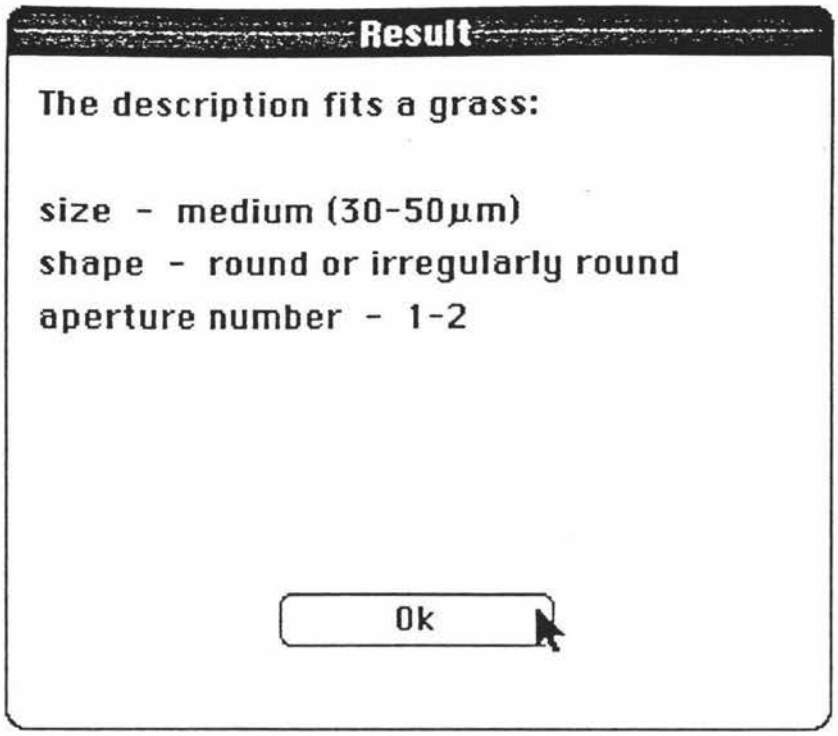


Figure B.9: Result dialog.

## B.3 Viewing description of pollen

The user is shown a list of the pollens in the system, and asked to select the one they would like described (Selection: Grass) (Figure B.10).

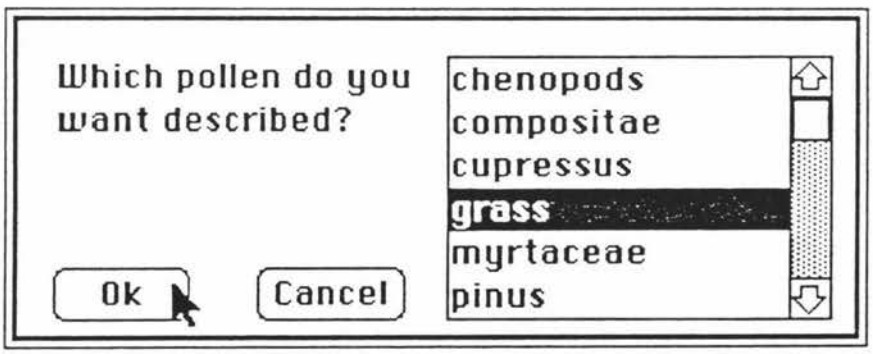


Figure B.10: Pollen selection dialog.

He/she is then shown the character states which describe the selected pollen (Figure B.11).

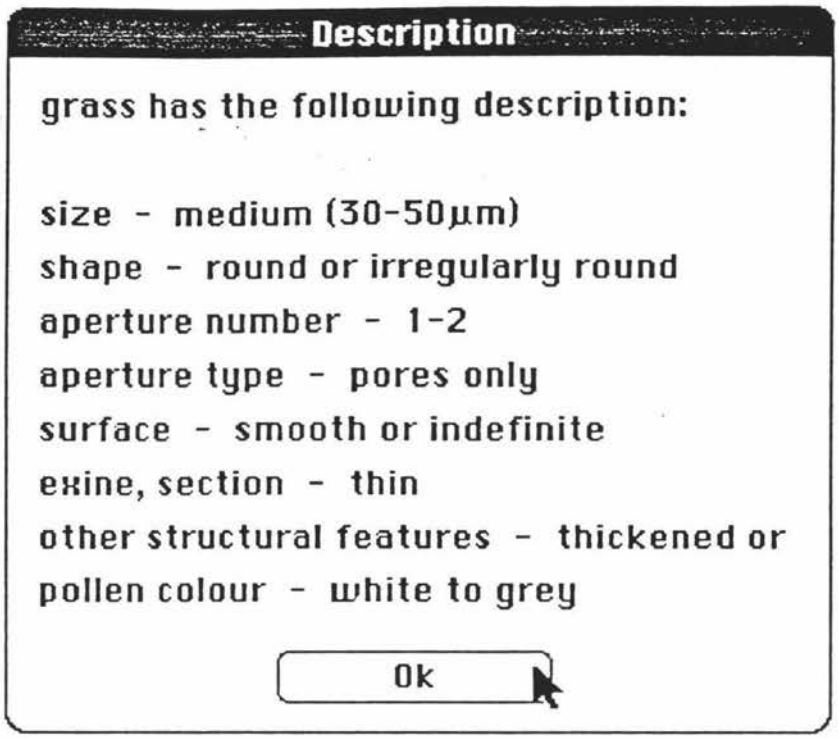


Figure B.11: Pollen description dialog.

# Appendix C

## Characters used in pollen identification expert system

This appendix contains a description of the characters used in the knowledge base of the pollen identification expert system. Not all characters are used for all taxa in the knowledge base. There may be more than one instance of each character for a taxon.

| Name | Description |
|---|---|
| Botanic name | Botanic name of the plant which produces the pollen. |
| Common name | Common name of the plant which produces the pollen. |
| Size | Size of the pollen. This can be a constant value, a range of values, or the string 'variable'. |
| Shape | Shape of the pollen. This is a character string. |
| Surface appearance | Surface appearance of the pollen. A character string. |
| Intine appearance | Character string describing the intine of the pollen. This is the innermost of the two layers forming the wall of the pollen grain. |
| Pore number | The number of pores on the pollen. Pores are apertures which are approximately round, less than twice as long as they are wide. This can be a constant value, a range of values, or the string 'variable'. |
| Pore placement | The placement of pores on the pollen grain. A character string. |
| Pore appearance | The appearance of pores on the pollen grain. A character string. |

| | |
|---|---|
| Furrow number | The number of furrows on the pollen grain. Furrows are boat-shaped apertures, more than twice as long as they are wide. This can be a constant value, a range of values, or the string 'variable'. |
| Furrow appearance | The appearance of furrows on the pollen grain. A character string. |
| Colour | The colour of the pollen grain after staining. |
| Other features | Unusual features of the pollen, which are not covered by other characters. |
| Location | A list, commencing with a location, and 12 probabilities, corresponding to the probability of the pollen appearing at the given location in each month of the year. |

# Appendix D

## Adding taxa to the knowledge base

### D.1 Introduction

The subsystem for adding taxa allows domain experts to easily update the contents of the knowledge base. It uses a direct manipulation interface: all actions the user can take are presented in the form of buttons; choices of character states are presented using scrolling menus. The subsystem can be easily modified to change or add characters, allowing other domains to use the system.

For every character, the user is asked to select one or more states that best describe the taxon they are specifying. The user can select 'unknown' if the taxon being described does not exhibit a state for that character. If the user selects a state, he/she is then asked to specify the importance of that character in identifying the taxon.

### D.2 Example session using pollen system

The user is initially asked for the botanic name of the pollen being described (Myrtacea in this example).

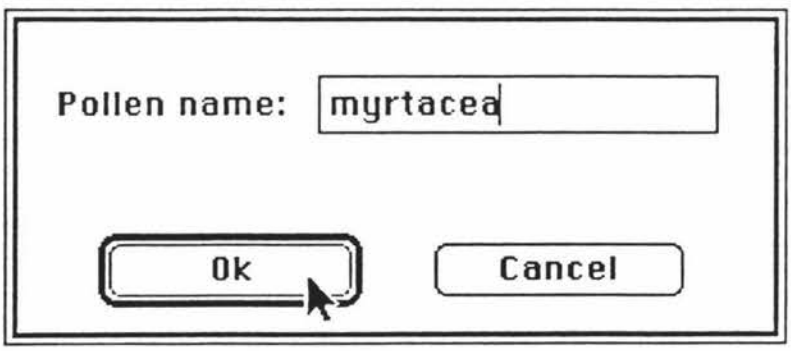


Figure D.1: Botanic name dialog.

Common names for the pollen are then requested. The user can continue by selecting 'Unknown' to indicate that they have no information about common names, 'Add another common name' to continue adding names, or 'Ok' to indicate that all names the user knows about have been specified. For example, a common name for myrtacea is eucalyptus.

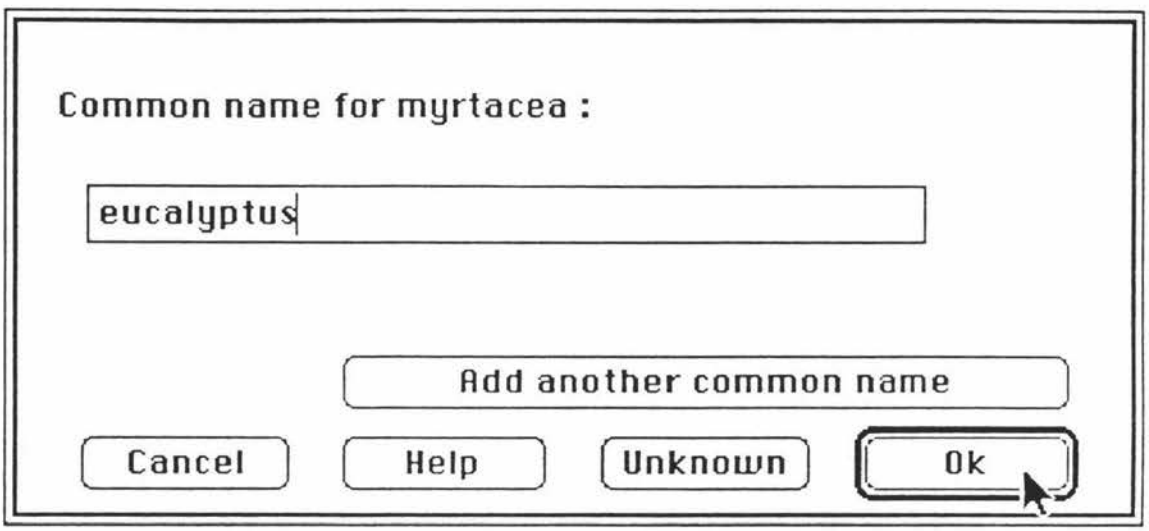


Figure D.2: Common name dialog.

An indication of the size of the pollen is then requested. The user has selected 'Unknown', to indicate that the size of myrtacea is not known at this stage.

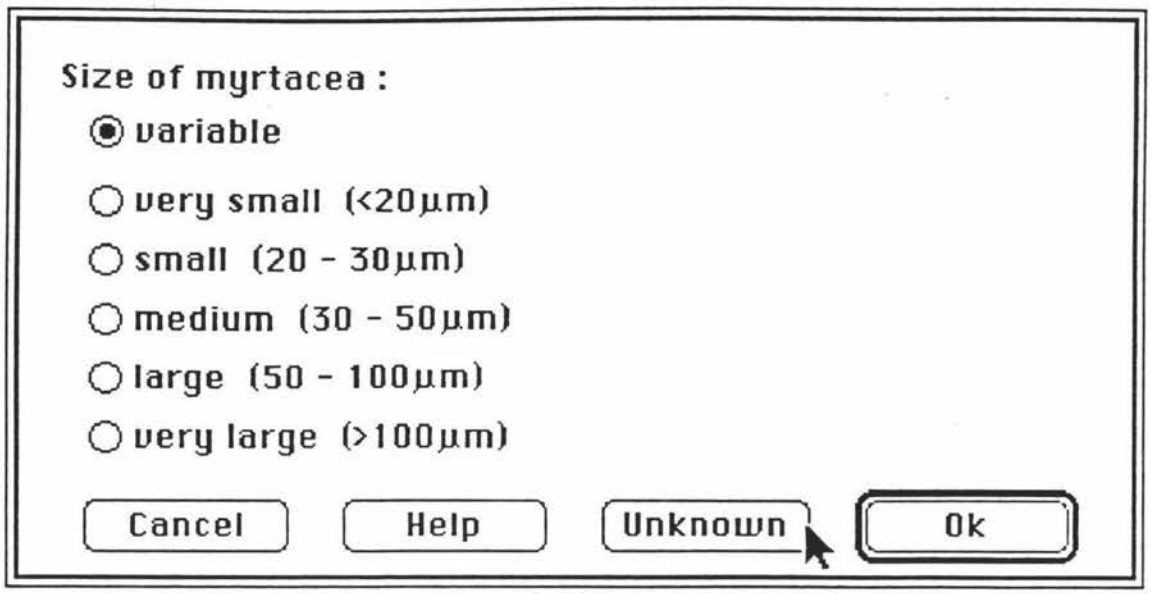


Figure D.3: Size dialog.

Next, an indication of the shape of the pollen is requested. The user has selected 'Triangular', to indicate that the shape of myrtacea is triangular.

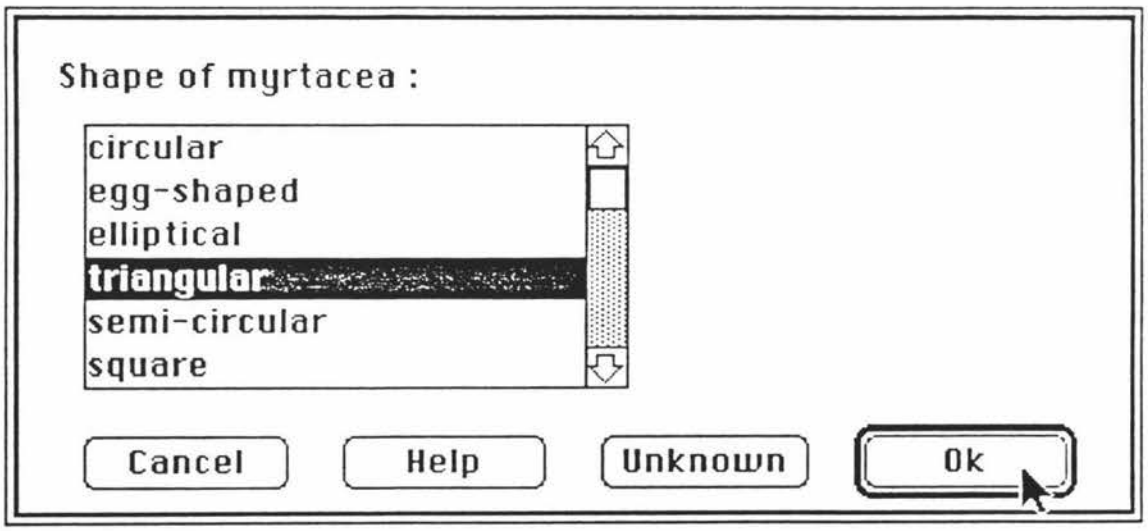


Figure D.4: Shape dialog.

As a state was selected for the character 'shape', the user is asked to specify the importance of shape in identifying the pollen. In this example, 'Essential'

has been selected, indicating that the shape 'triangular' must be specified in order for a specimen to be identified as myrtacea.

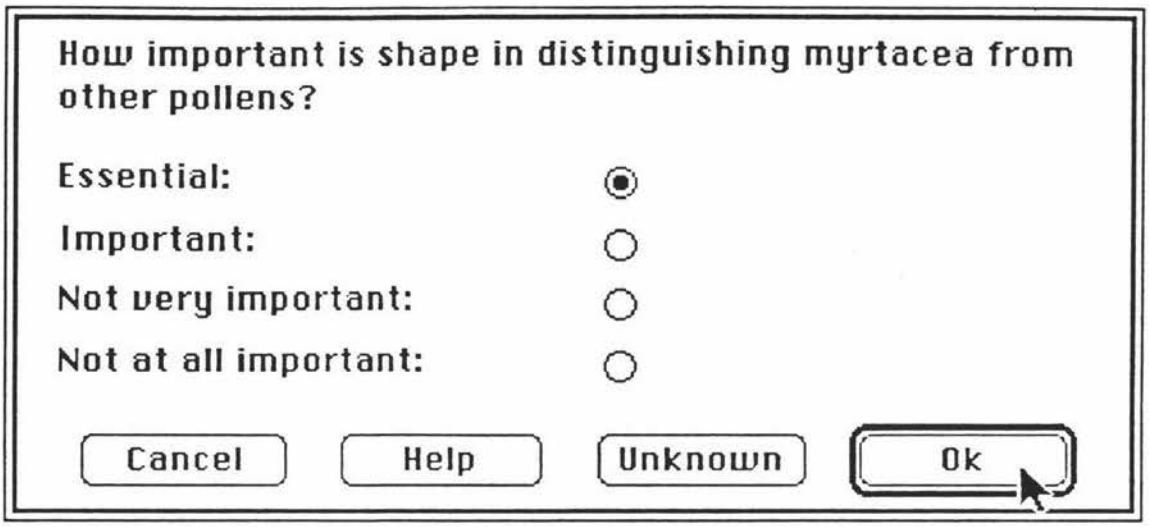


Figure D.5: Shape importance dialog.

An indication of the surface appearance of the pollen is then requested. The user has selected 'Unknown', to indicate that the surface appearance of myrtacea is not currently known.

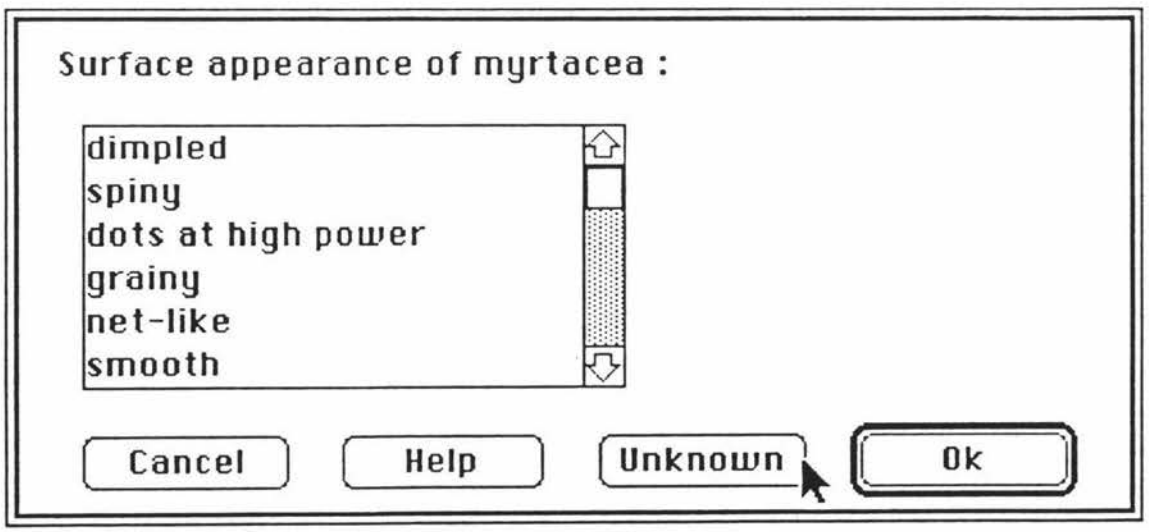


Figure D.6: Surface appearance dialog.

An indication of the intine appearance of the pollen is then requested. The user has selected 'Unknown', indicating that the intine appearance of myrtacea is not known at this stage.

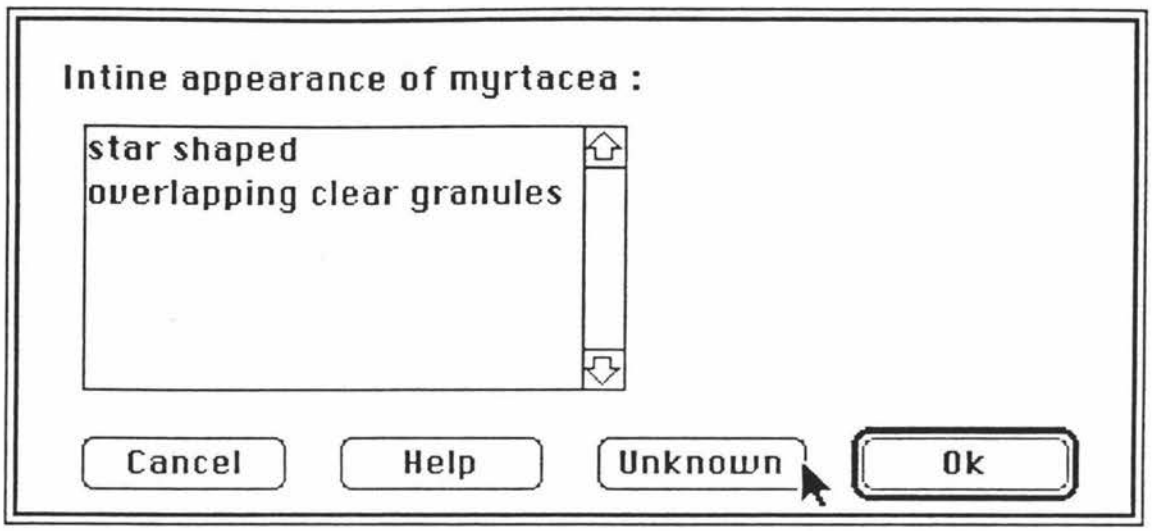


Figure D.7: Intine appearance dialog.

The user is then asked to specify the number of pores appearing on the pollen. The user has selected 'Range', and entered values which indicate a range of 3 - 4.

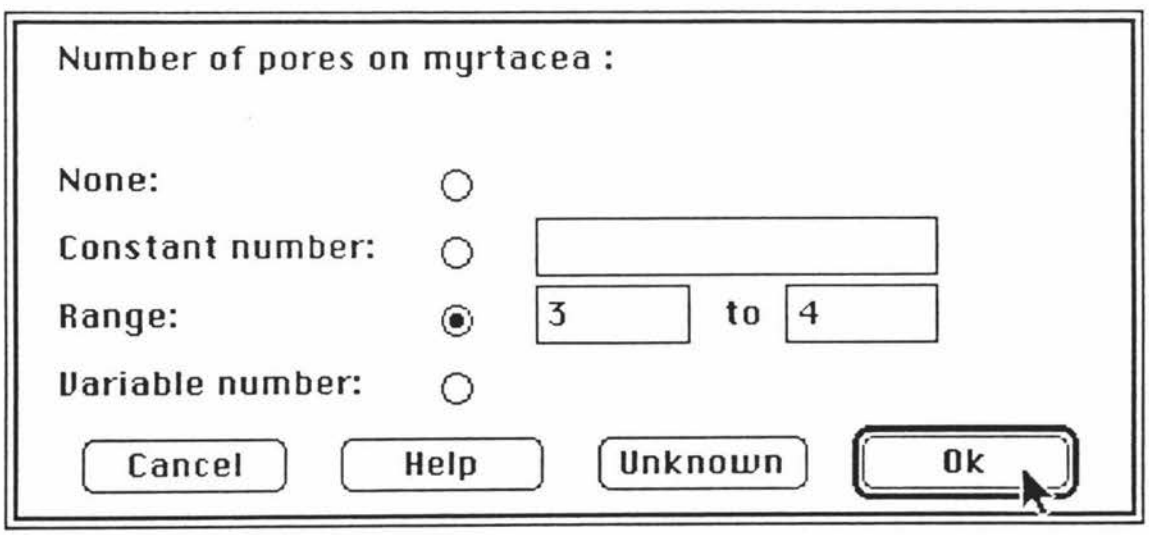


Figure D.8: Pore number dialog.

As a state was selected for the character 'number of pores', the user is asked to specify the importance of the number of pores in identifying the pollen. In this example, 'Essential' has been selected, indicating that the number of pores on the specimen must be in the range 3 - 4 in order for a specimen to be identified as a myrtacea.

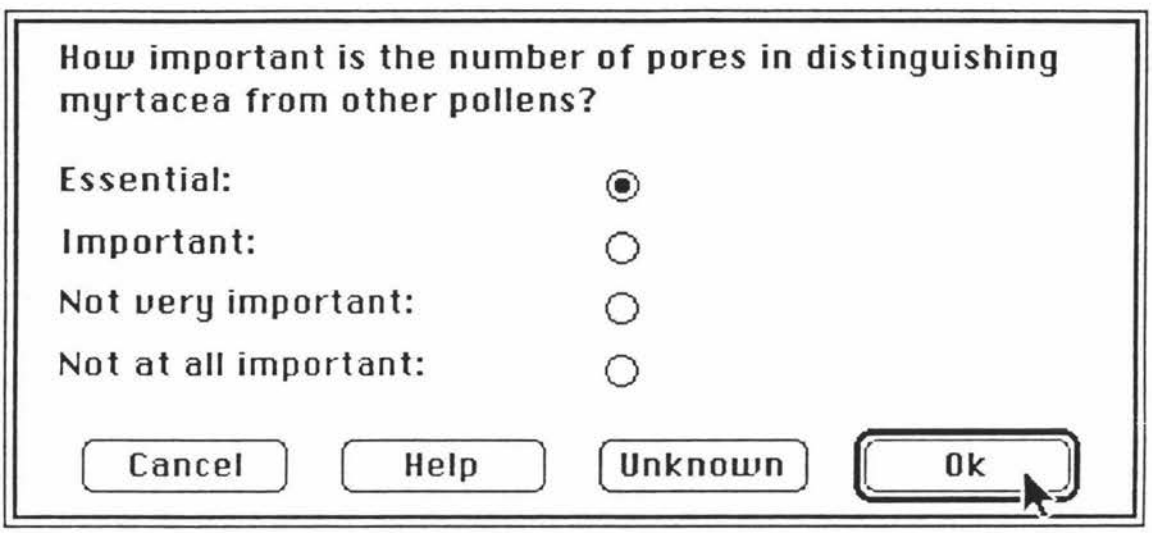


Figure D.9: Pore number importance dialog.

The user is then also asked to describe the appearance of the pores on the pollen. In this example, he/she has selected 'Clearly visible', indicating that the pores on myrtacea are clearly visible.

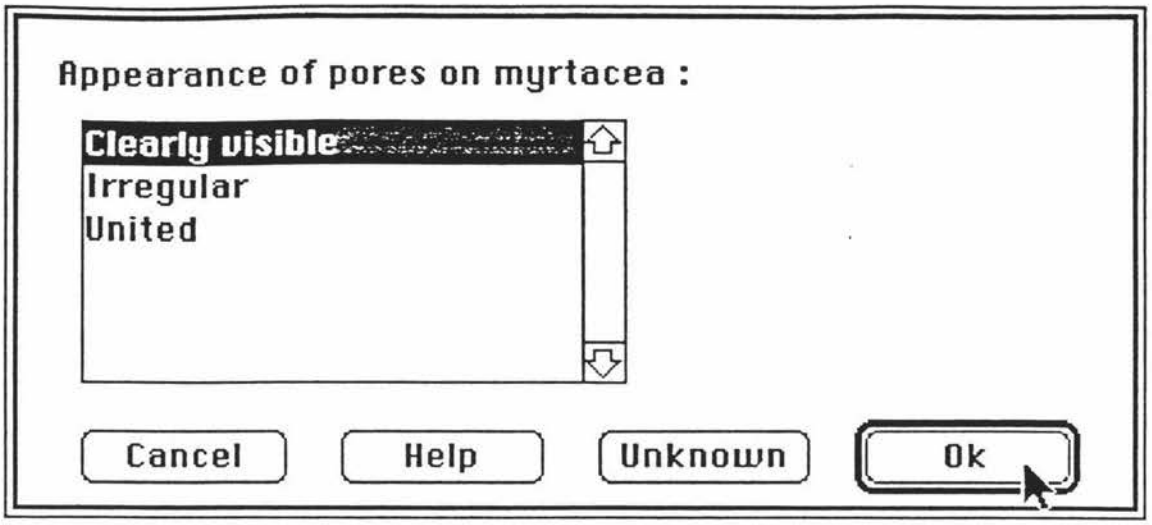


Figure D.10: Pore appearance dialog.

Because a state was selected for the character 'pore appearance', the user is asked to specify the importance of pore appearance in identifying the pollen. In this example, he/she has selected 'Essential', indicating that the pores on the specimen must be clearly visible in order for a specimen to be identified as a myrtacea.

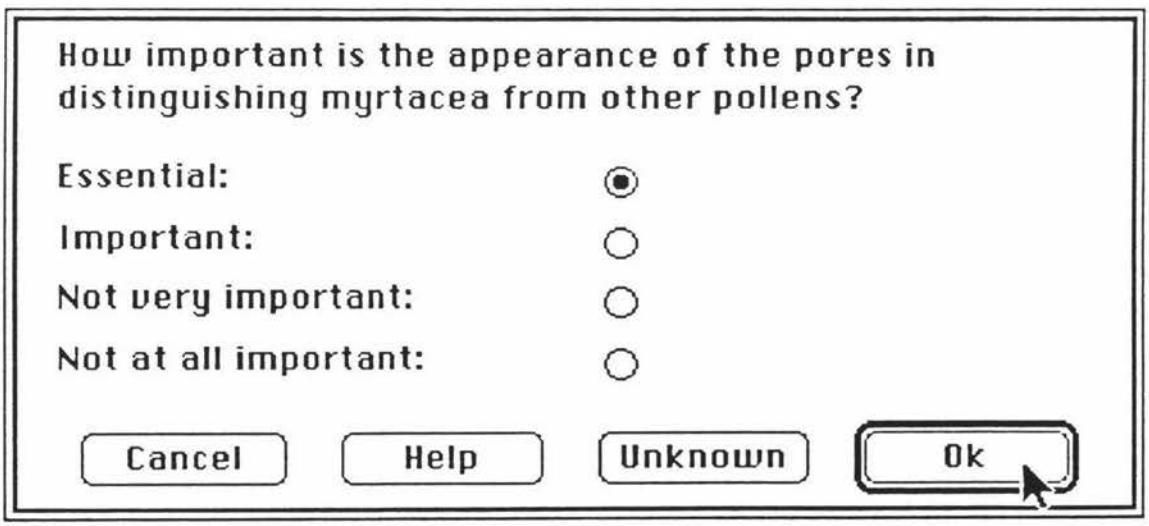


Figure D.11: Pore appearance importance dialog.

Next, the user, having given a state for the character 'number of pores', is asked to describe the placement of the pores on the pollen. In this example, he/she has selected 'At corners', indicating that the pores on myrtacea appear only at the corners of the pollen.

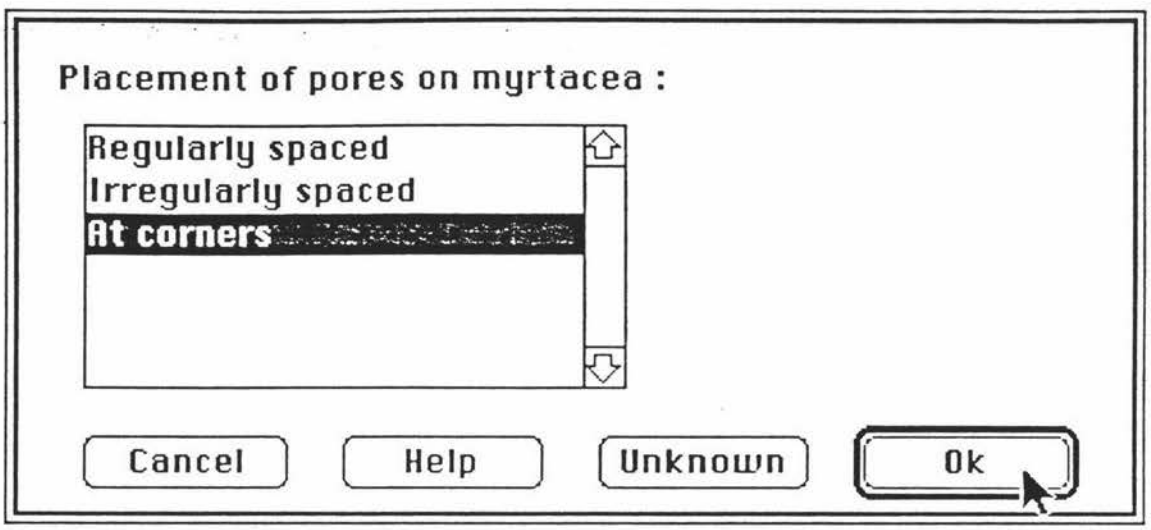


Figure D.12: Pore placement dialog.

As a state was selected for the character 'pore placement', the user is asked to specify the importance of pore placement in identifying the pollen. In this example, he/she has selected 'Essential', indicating that the pores on the specimen must appear at the corners in order for a specimen to be identified as a myrtacea.

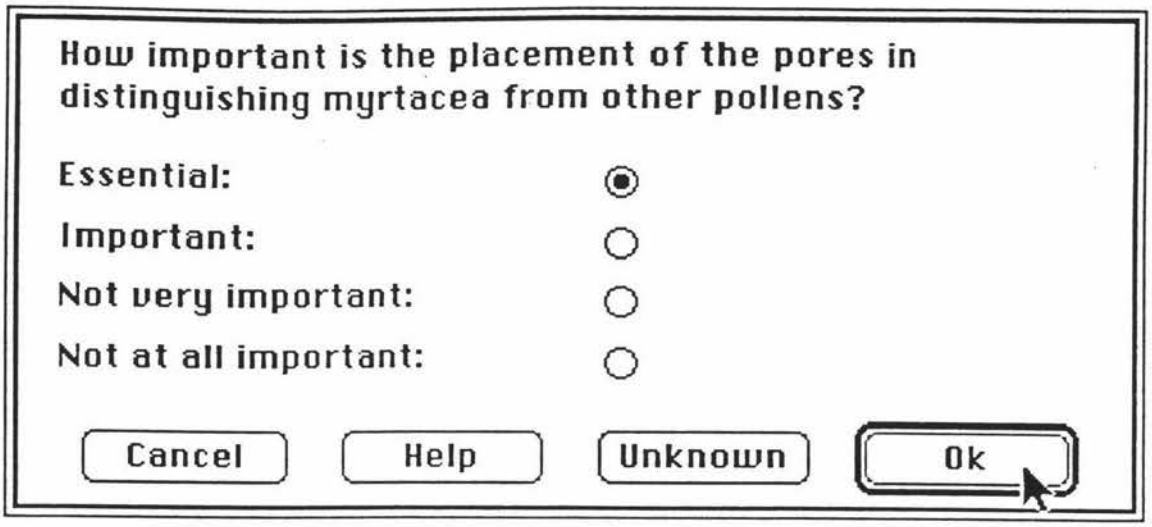


Figure D.13: Pore placement importance dialog.

An indication of the number of furrows on the pollen is then requested. The user has selected 'Unknown', indicating that the number of furrows of myrtacea is not known at this stage.

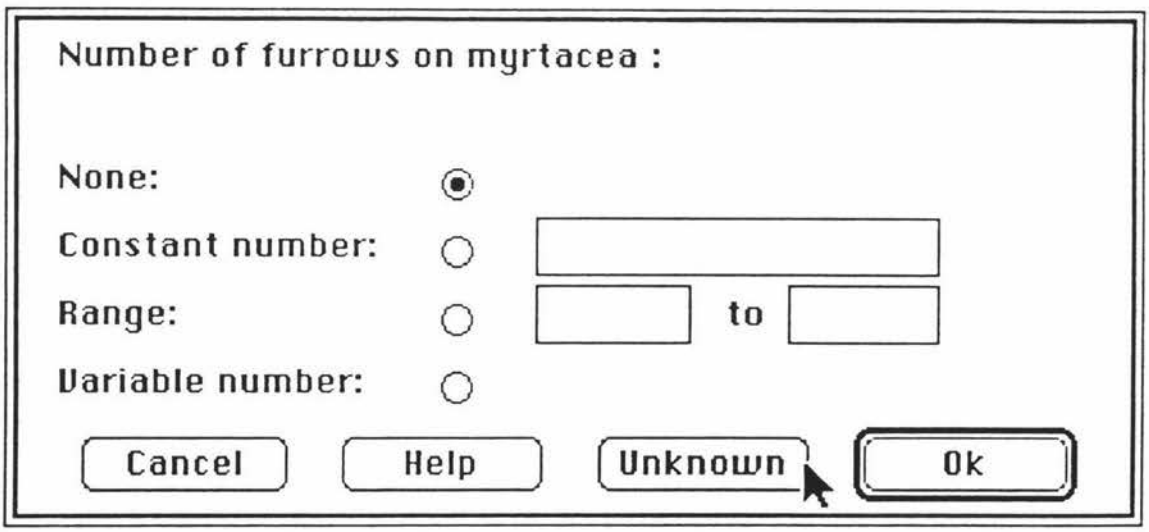


Figure D.14: Furrow number dialog.

In addition, an indication of the colour of the pollen is requested. The user has selected 'Unknown', indicating that the colour of myrtacea is not known at this stage.

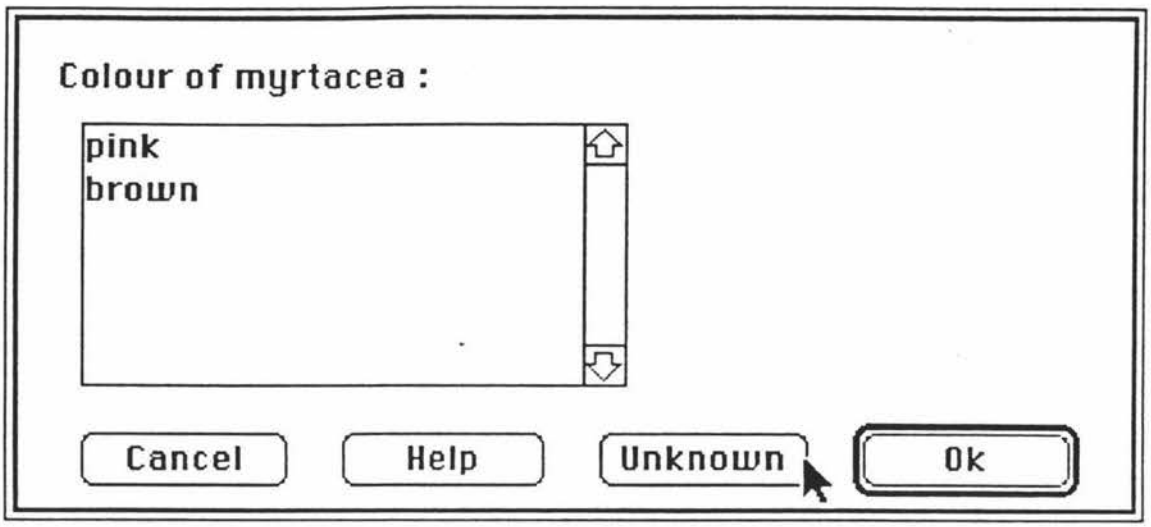


Figure D.15: Colour dialog.

Features of the pollen which have not yet been recorded are then requested. The user can continue by selecting 'Unknown' to indicate that he/she has no information about other features, 'Add another feature' to continue adding features, or 'Ok' to indicate that all features the user has information about have been specified. A feature of myrtacea is that they commonly stick together.

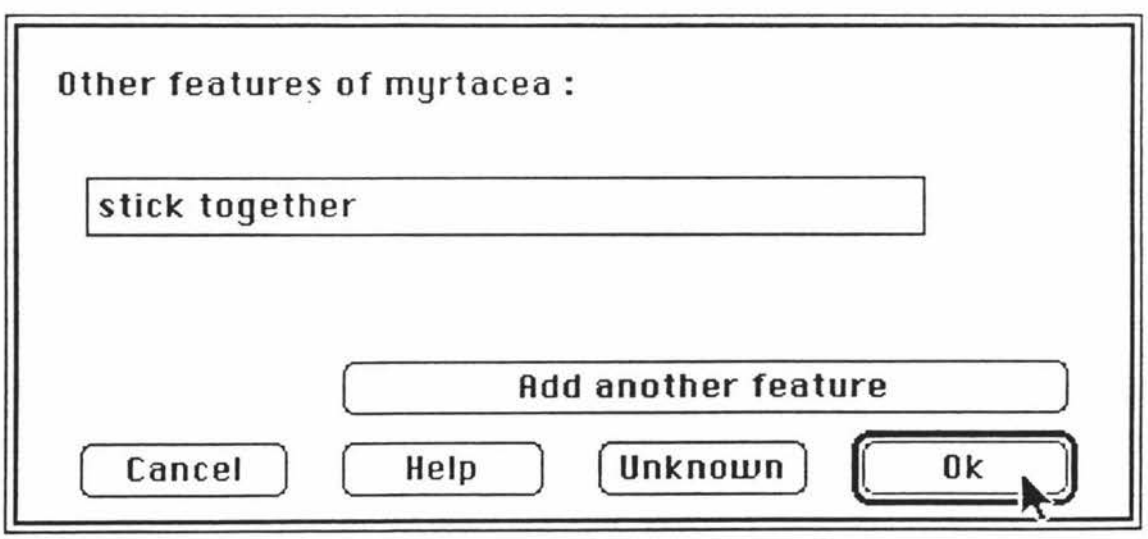


Figure D.16: Other features dialog.

As a state was selected for the character 'other features', the user is asked to specify the importance of these features in identifying the pollen. In this example, he/she has selected 'Important', indicating that these features should only be employed if all essential characters have been used.

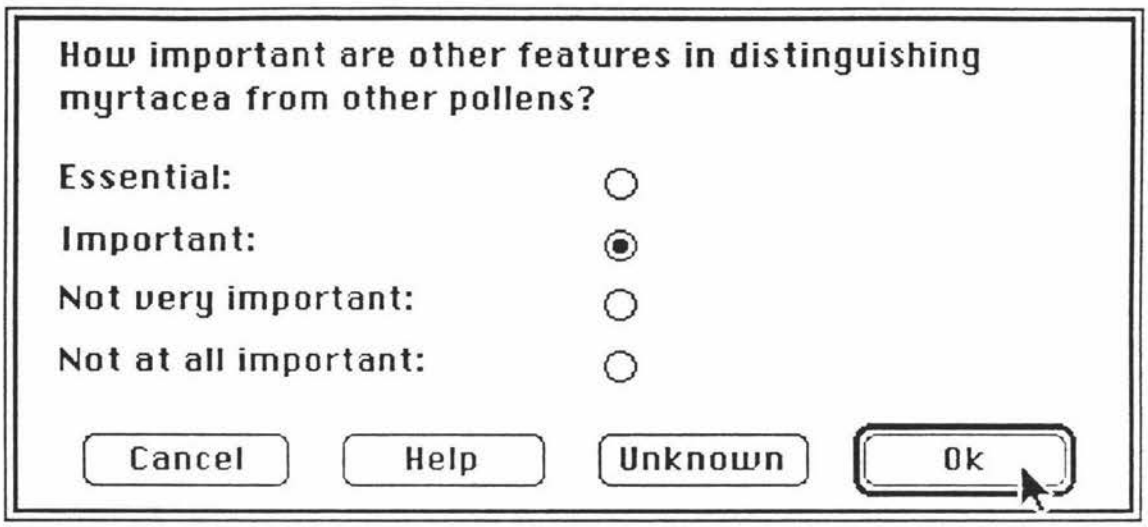


Figure D.17: Other features importance dialog.

Finally, an indication of the locations at which the pollen is known to appear is requested. The user has selected 'Palmerston North', to indicate that myrtacea is known to appear only at this location.

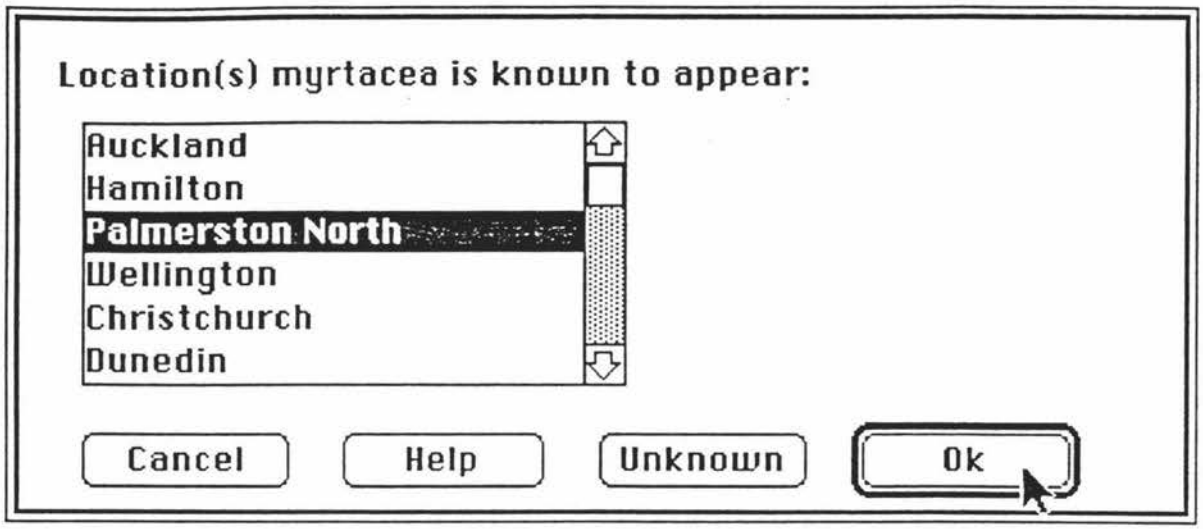


Figure D.18: Location dialog.

As a state was selected for the character 'Location', the user is asked to specify the times of the year at which the pollen is known to appear at that location. In this case, myrtacea has a medium chance of being collected in March, a low chance in April, and is not seen during the remainder of the year.

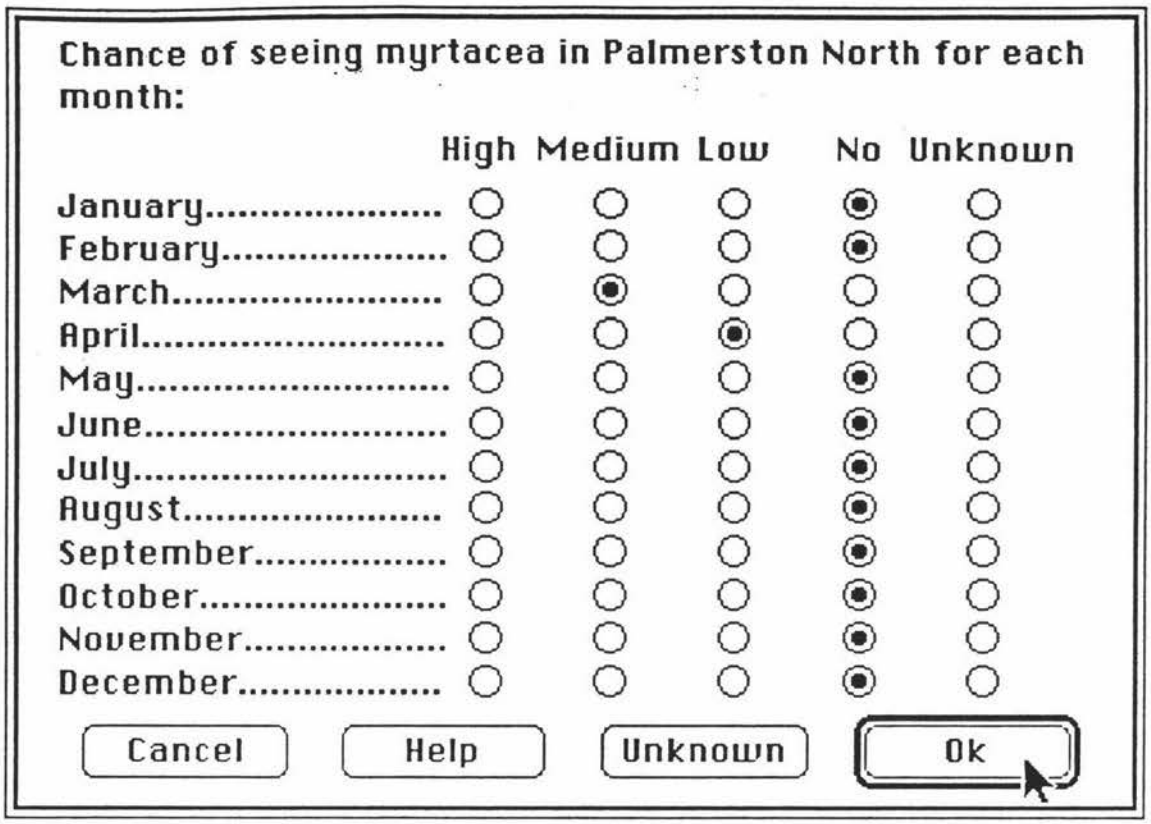


Figure D.19: Location/Month dialog.

# Appendix E

## Example session using the present system

## E.1 Counting subsystem

### E.1.1 Introduction

Pollens are collected on vaseline-covered rods which are placed in exposed areas and changed approximately every 24 hours. The rods are placed on microscope slides and the pollens identified and counted. The purpose of the counting subsystem is to assist both non-specialised and experienced staff to accurately count the numbers of different types of air-borne pollens which have collected on a rod.

### E.1.2 Example session

The user is initially asked for the date of collection. The '-' and '+' buttons allow the user to change the date from the default if required. November 25, 1989 has been selected in Figure E.1.

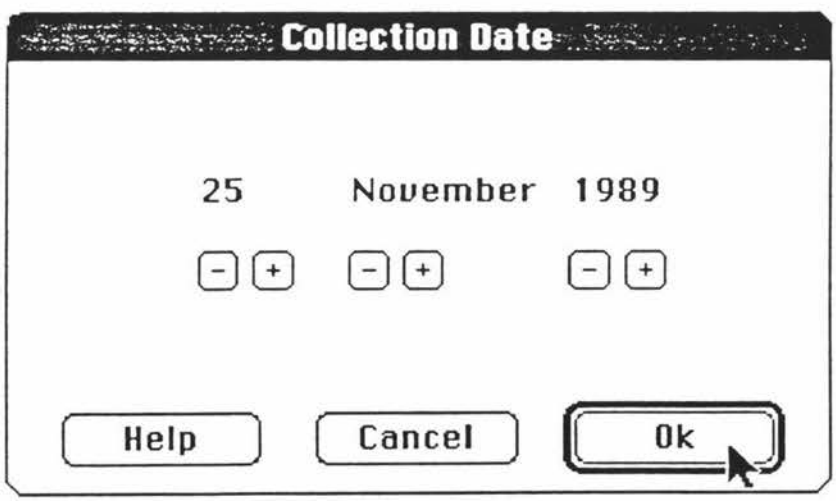


Figure E.1: Collection date dialog.

The user is then asked for the location of the collection device. Palmerston North has been selected in Figure E.2.

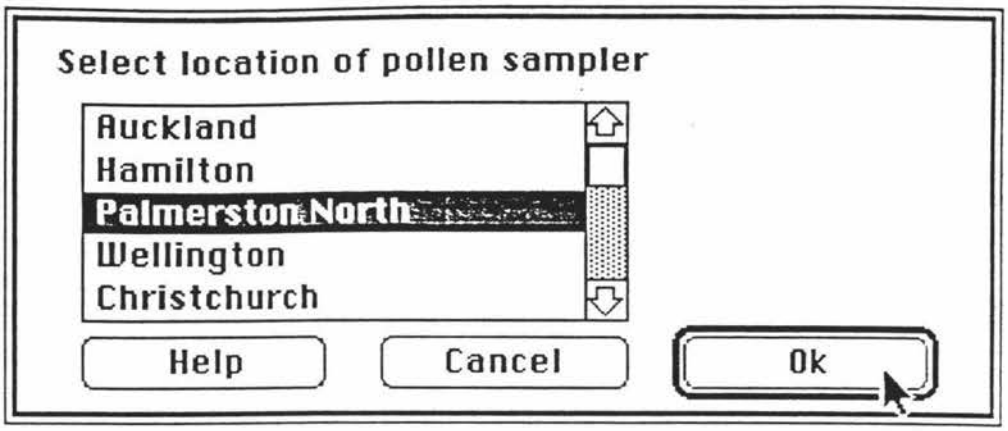


Figure E.2: Location dialog.

The most likely pollens for the given date and location are selected from the knowledge base, and pictures of these are displayed on the computer screen (Figure E.3). The user counts pollens by placing the cursor on a picture using the mouse, and clicking the mouse button. The system 'beeps' when a pollen is counted, allowing the user to count without looking up from the microscope.

The total pollen count may be requested at any time. The system then displays the counts of the various types of pollen, total number of pollens and spores, and the density of these in grains per cubic metre of air (Figure E.4).

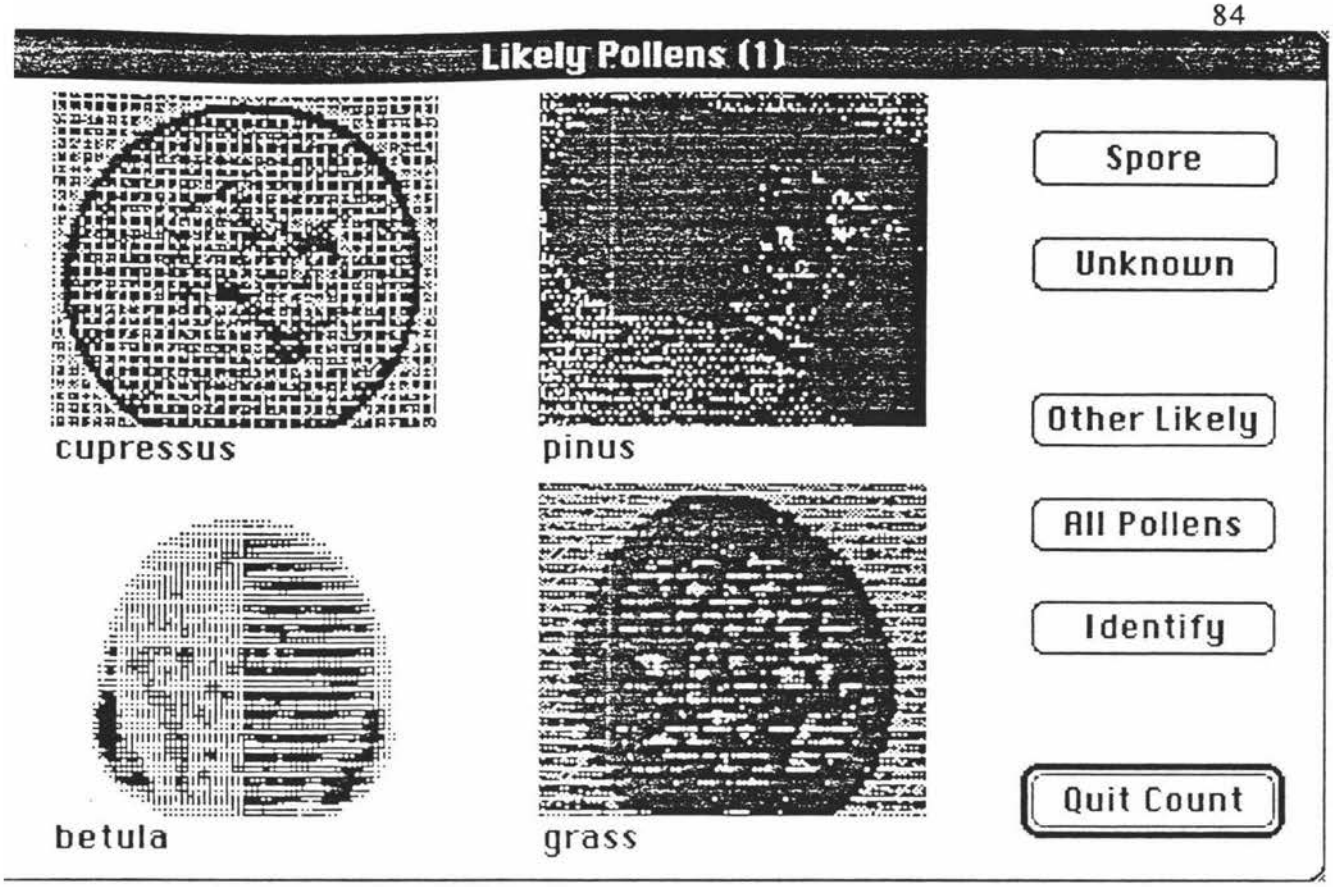


Figure E.3: Likely pollens dialog.

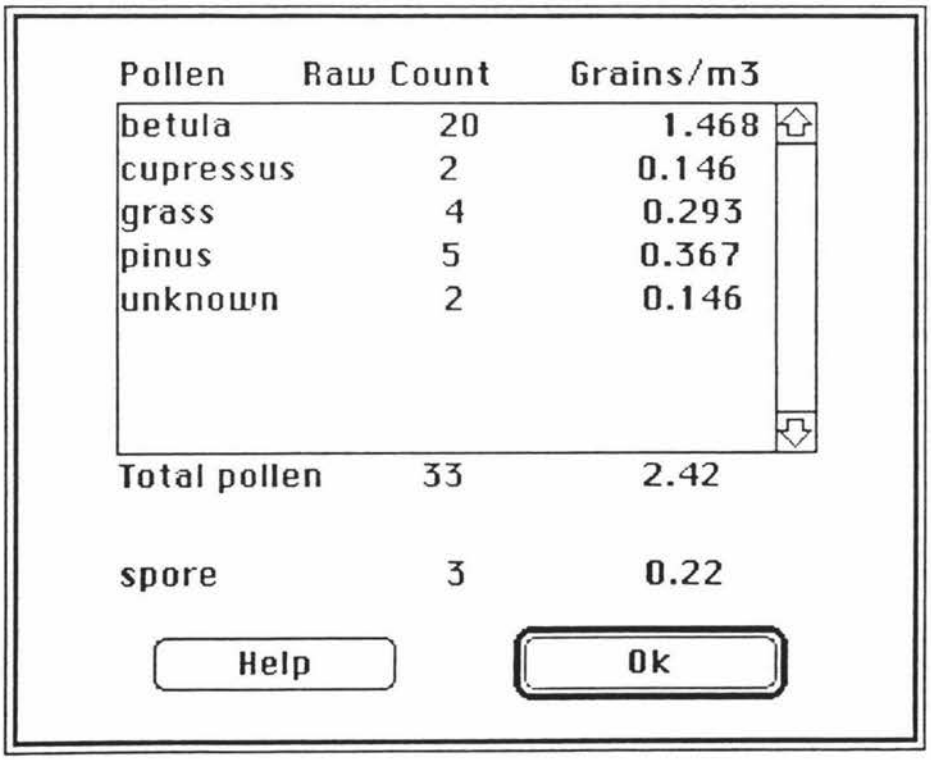


Figure E.4: Count result dialog.

## E.2 Identification subsystem

### E.2.1 Single-access monothetic mode

Selecting the 'Identify' button in the count dialog begins the identification process. The system begins by asking about the character 'surface', this being the most important character for the most likely pollen during November in Palmerston North. The user can select one of the states given to describe the character, 'Unknown' to indicate that they cannot determine the state of the character, or 'None' to indicate that they can determine the character state, but it does not match any of the given states. In this example, the user does not know the state of the surface on the specimen, so has selected 'Unknown'. Note the statement at the bottom showing the number of possibilities remaining in the system.

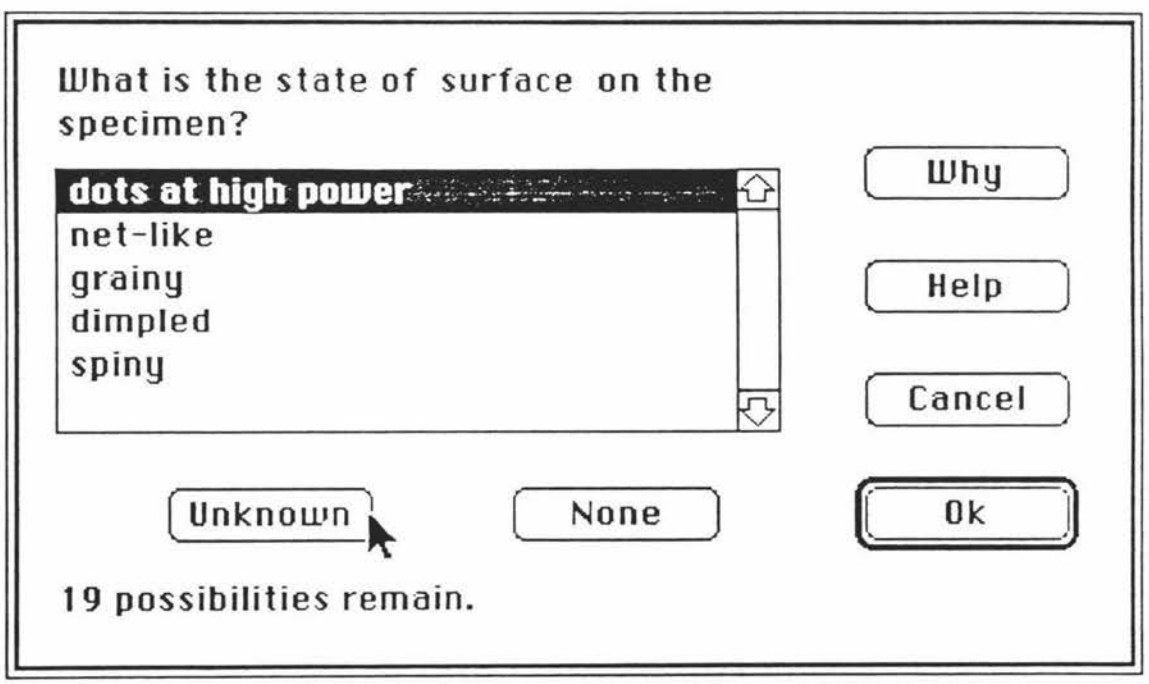


Figure E.5: Surface dialog.

As 'Unknown' was selected, the system continues with all the pollens, and queries the user about the next most important character, intine appearance. In this example, the user does not know the state of the intine on the specimen, so again has selected 'Unknown'.

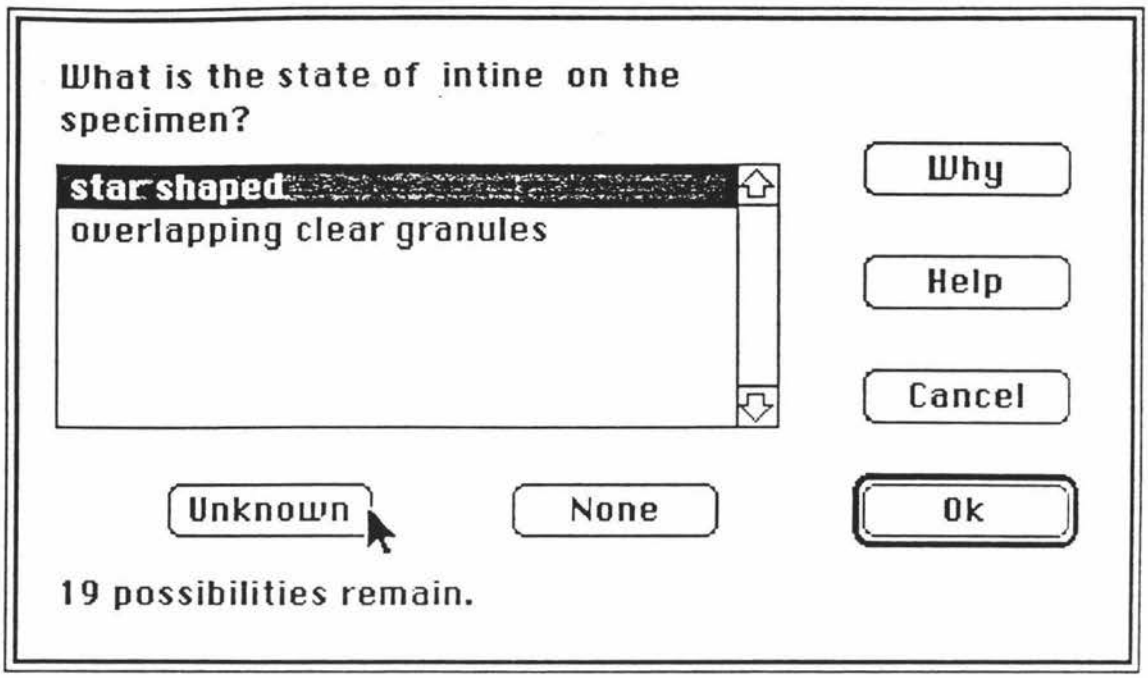


Figure E.6: Intine dialog.

As 'Unknown' was selected, the system continues with all the pollens, and queries the user about the next most important character, shape. In this example, the user has selected 'triangular'.

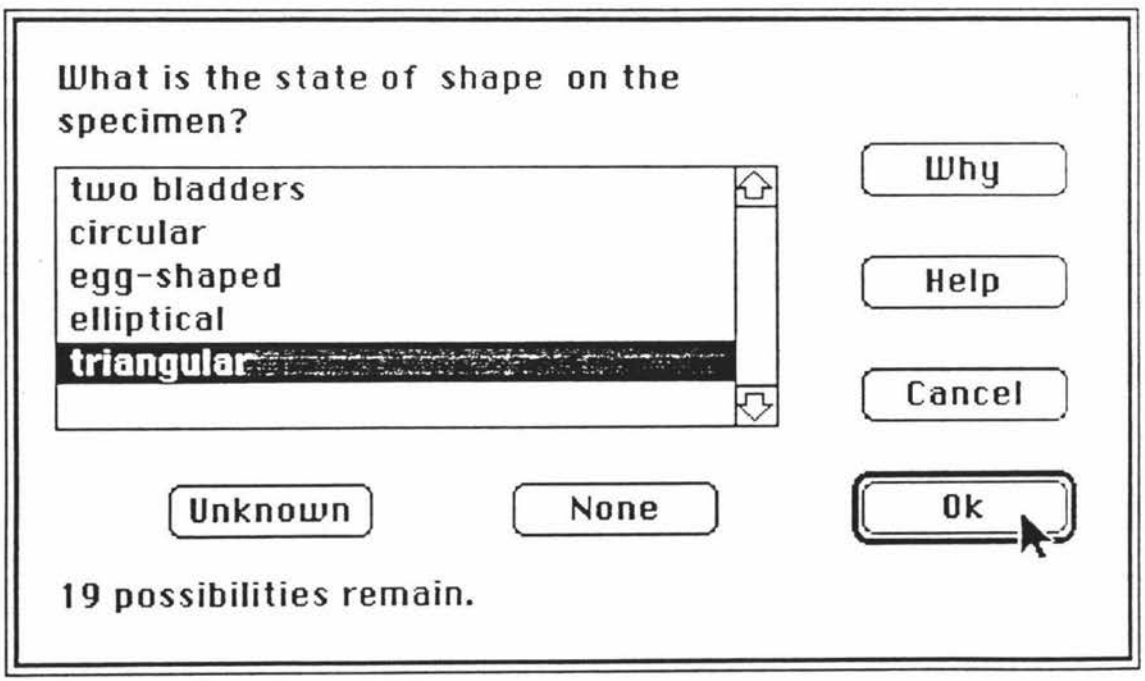


Figure E.7: Shape dialog.

As myrtacea is the only pollen present in the knowledge base which has 'shape' as an essential character, and which matches the state 'triangular', the system is able to identify the specimen as a myrtacea. The 'unlikely' comment shows that although the system has identified the specimen as a myrtacea, it is unlikely that a myrtacea would be collected in Palmerston North during November (see Appendix D).

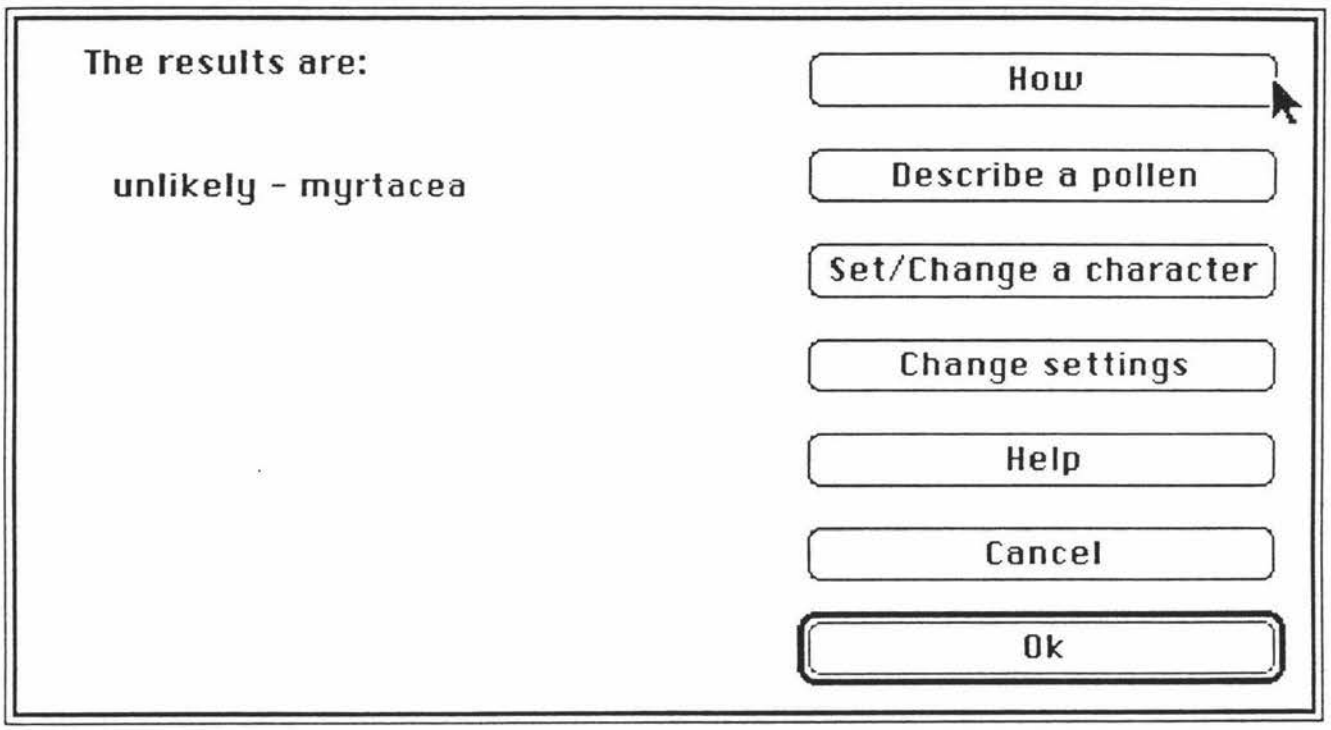


Figure E.8: Result dialog.

Selecting the 'How' button in the result dialog displays an explanation of how the identification was completed. Characters which are 'unknown' are not shown in the explanation dialog.

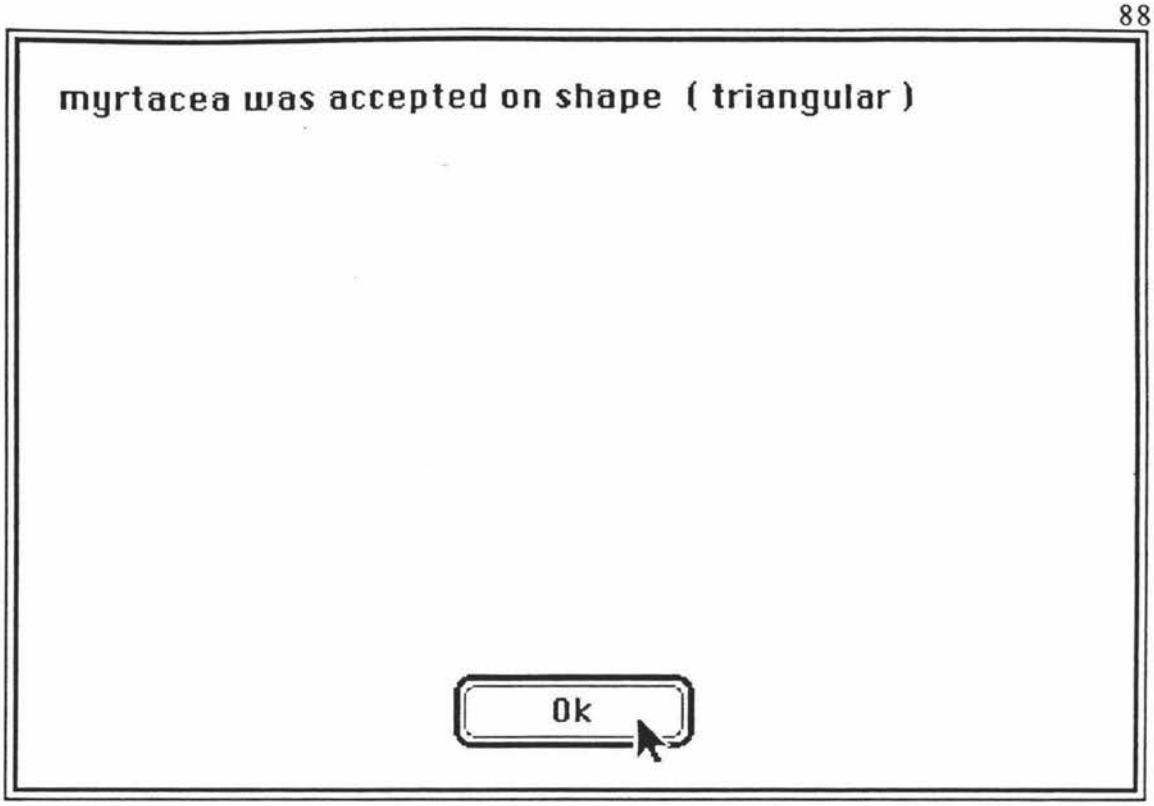


Figure E.9: Explanation dialog.

Selecting 'Describe a pollen' in the results dialog allows the user to select one or more pollens in order to view their descriptions. In this example, the user has selected myrtacea.

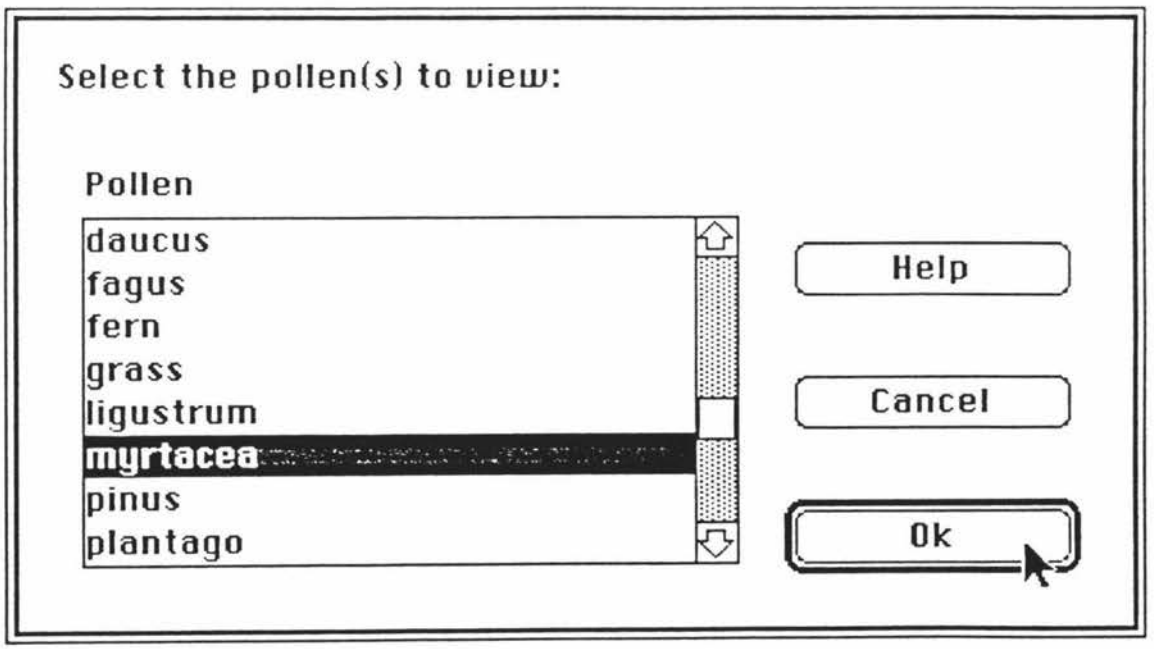


Figure E.10: Pollen selection dialog.

Note the '+' on shape, indicating a match with the description of the specimen.

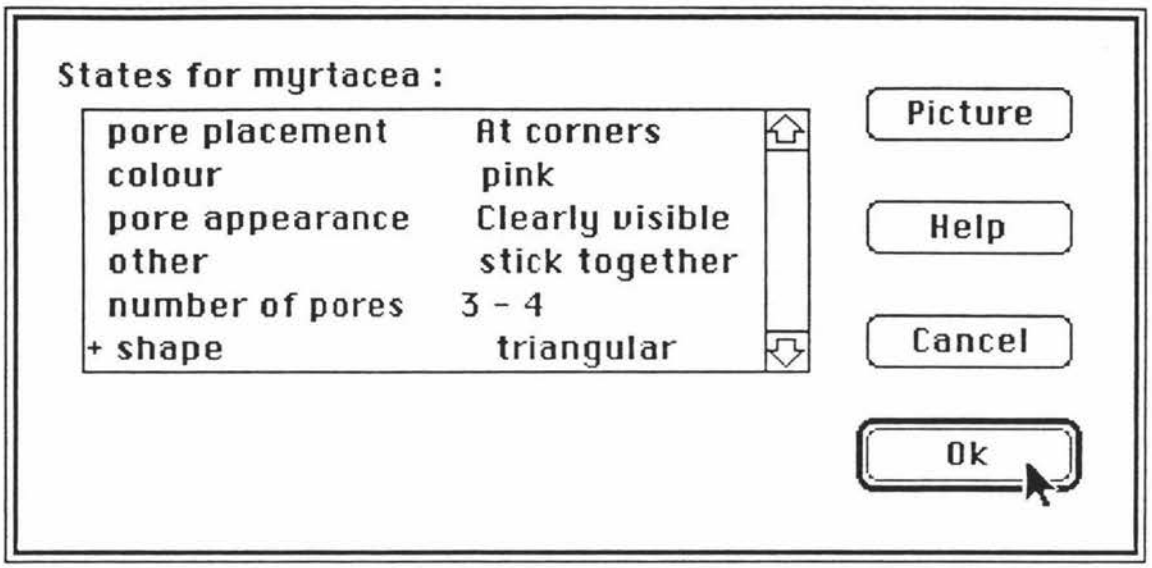


Figure E.11: Pollen description dialog.

### E.2.2 Multi-access monothetic mode

Selecting the 'Change settings' button in the result dialog allows the user to specify how far into the identification sequence the system is to use the single-access monothetic mode. If this number is low, most taxa must be rejected using the single-access monothetic mode before the user is permitted to select a character for description. If the number is larger than the number of taxa in the system, the user can use the multi-access monothetic mode from the beginning of the sequence.

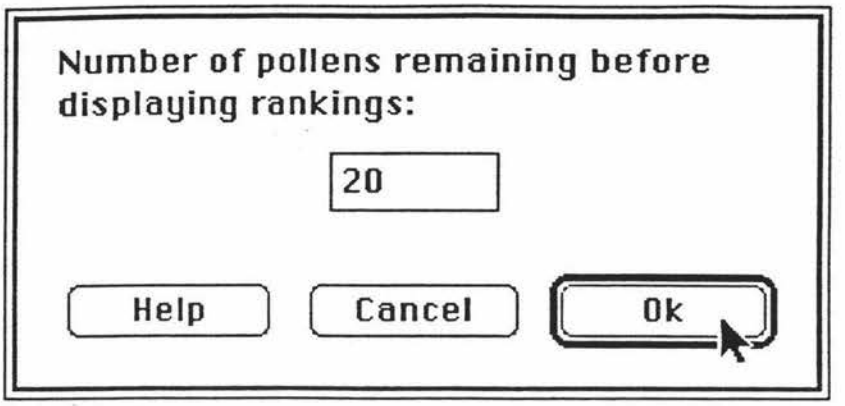


Figure E.12: Change settings dialog.

Selection of the 'Identify' button in the likely pollens dialog (Figure E.3), when the number set using 'Change settings' is larger than the number of taxa in the system displays the following dialog. This allows the user to 'Continue' in the single-access monothetic mode, or to 'Set/Change a character' to use the multi-access monothetic mode. The user has selected 'Set/Change a character' in this example.

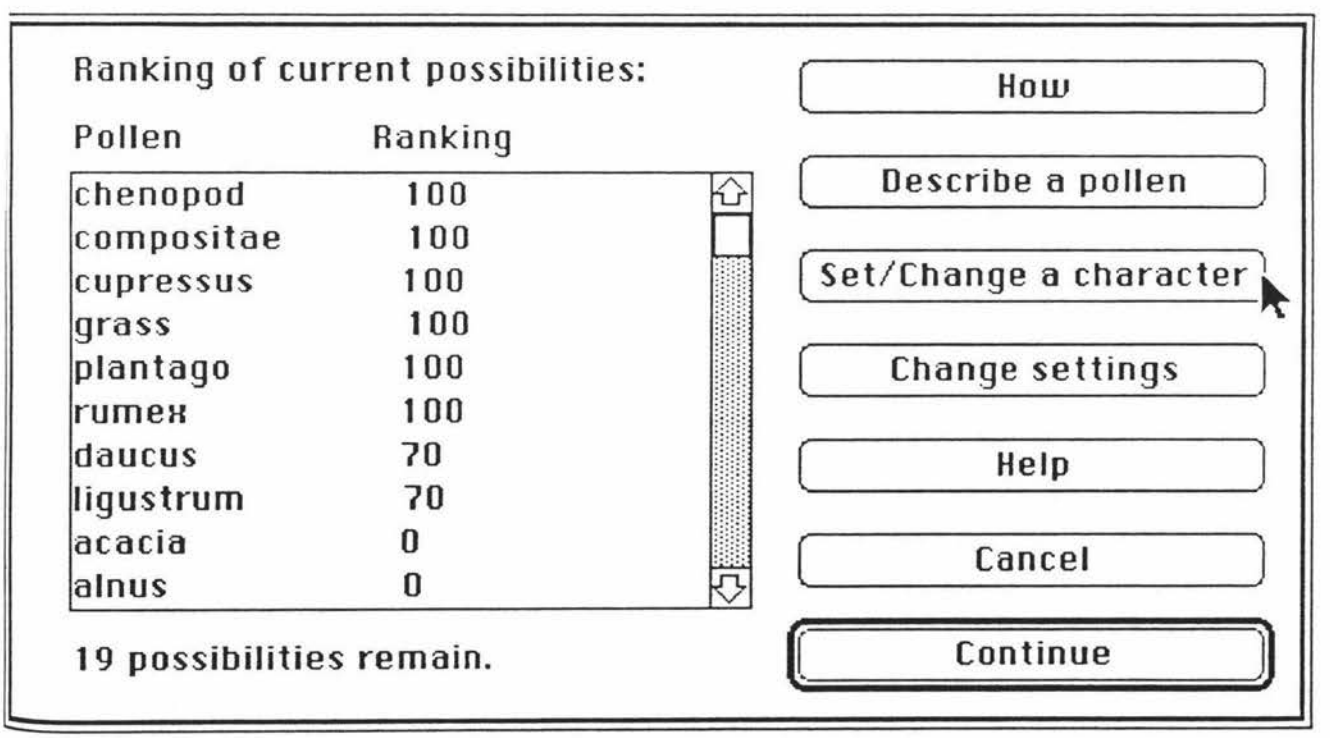


Figure E.13: Pollen ranking dialog.

Selecting 'Set/change a character' allows the user to describe a character state, or to change a previously given character state. The user has selected the character 'shape' in this example.

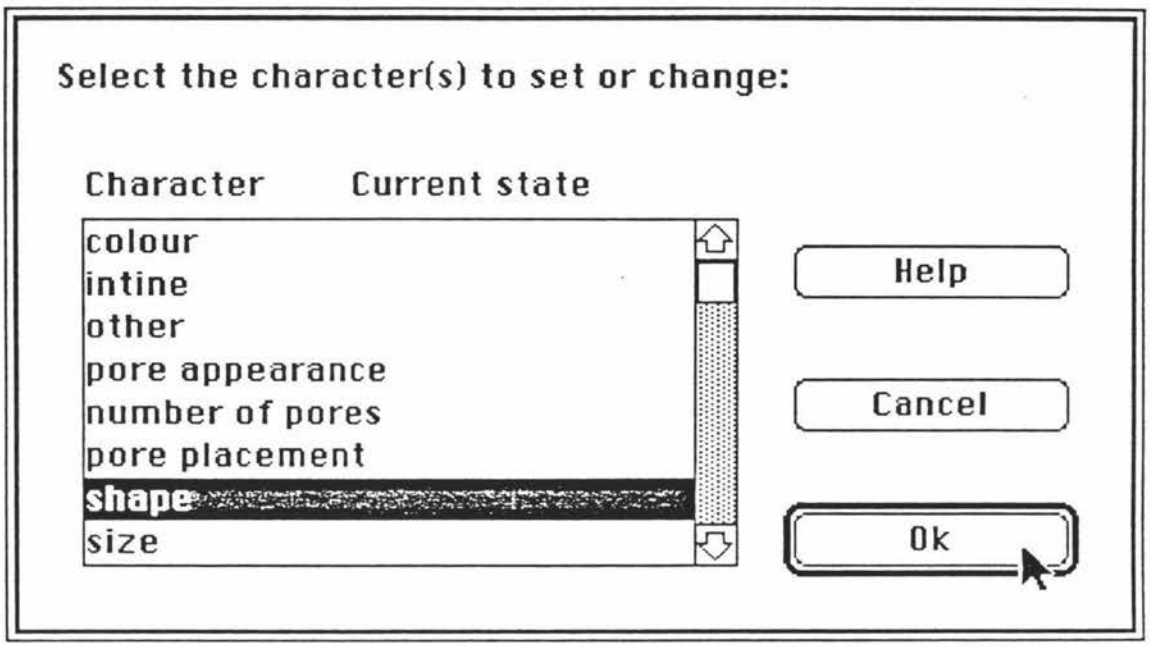


Figure E.14: Character selection dialog.

The system displays the possible states for the character chosen, and allows the user to select one or more which describe the specimen. The state 'triangular' has been selected in this example.

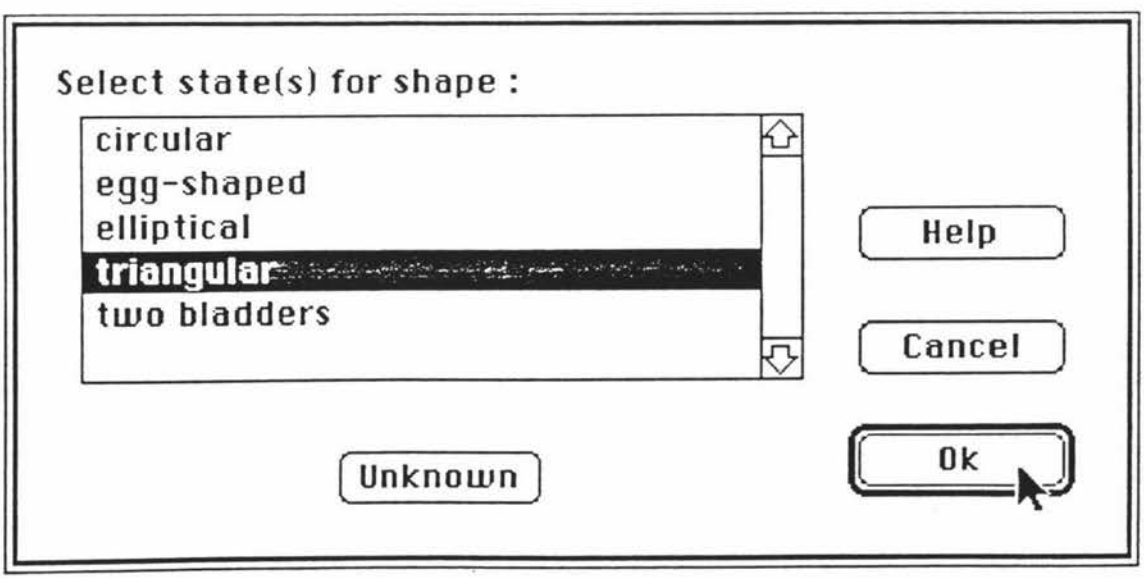


Figure E.15: Shape dialog.

As myrtacea is the only pollen present in the knowledge base which has 'shape' as an essential character, and which matches the state 'triangular', the system is again able to identify the specimen as a myrtacea.

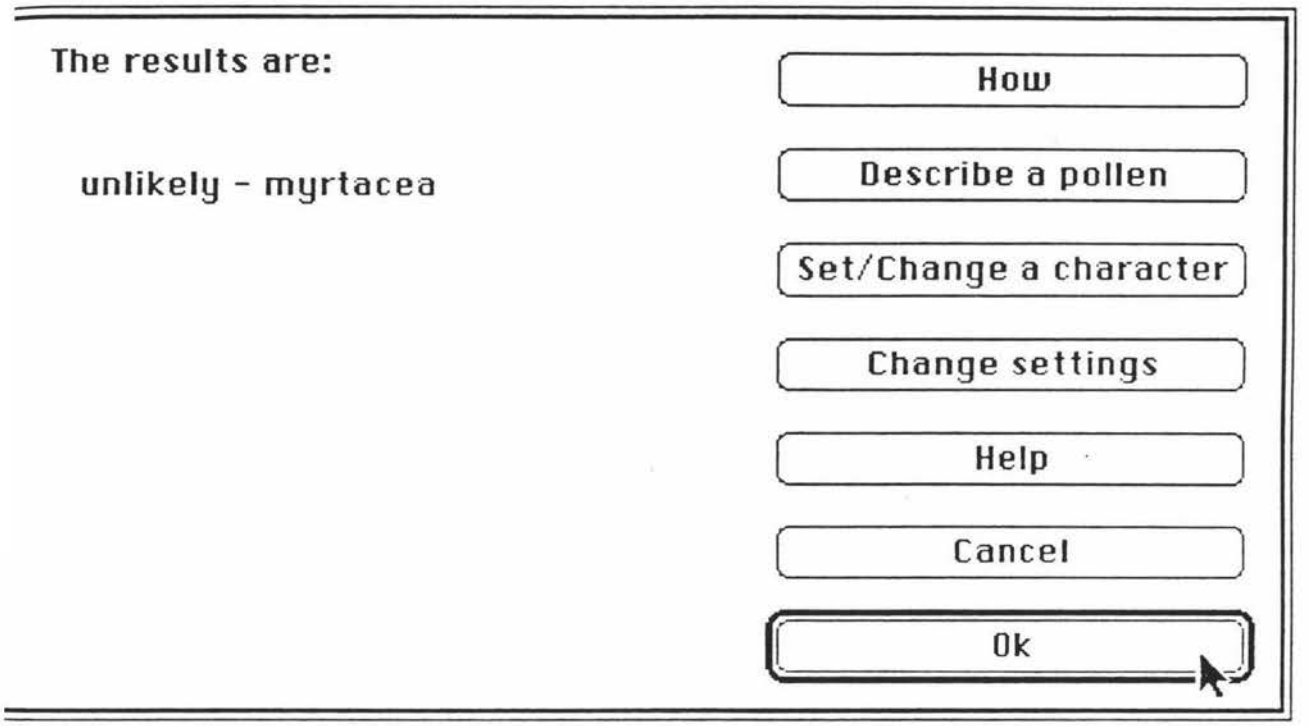


Figure E.16: Result dialog.

Selecting 'How' in the result dialog gives the same result as in Section E.2.1.

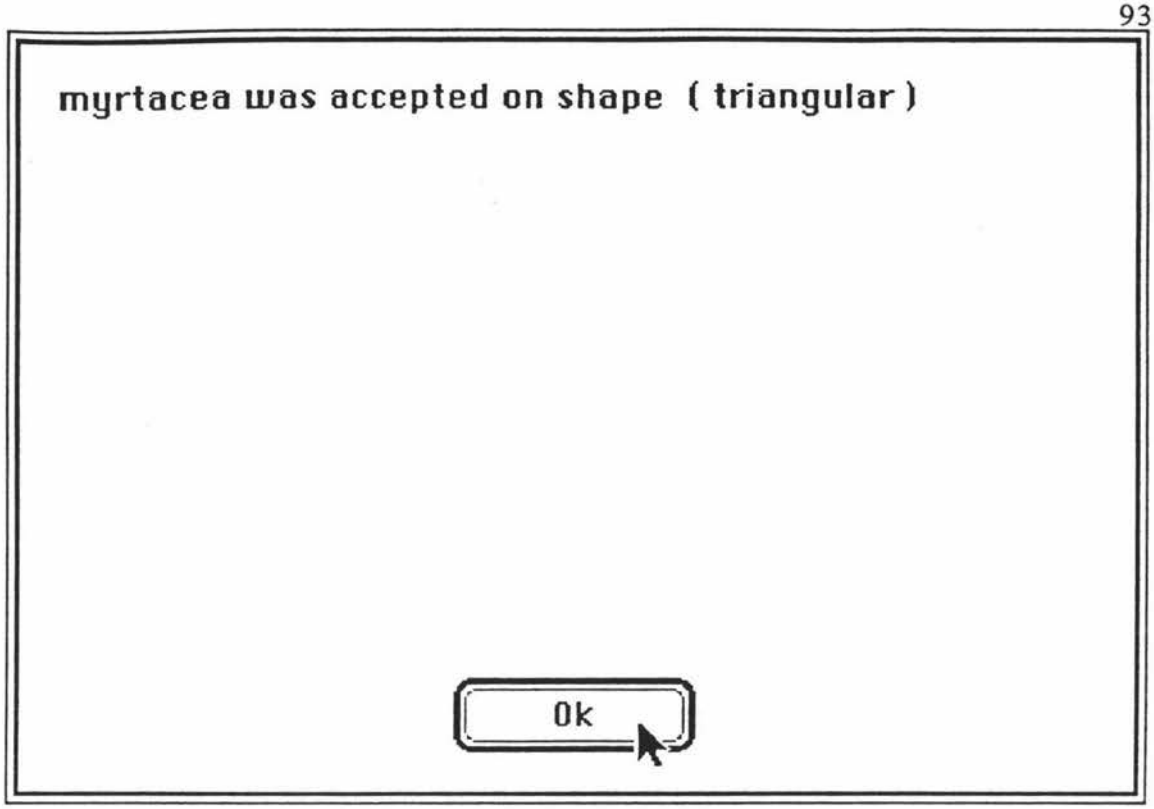


Figure E.17: Explanation dialog.

### E.2.3 Both single-access and multi-access monothetic modes

The multi-access monothetic mode can be used to narrow down an identification, then the single-access monothetic mode used to complete the identification, as in the following example. The user has selected 'number of pores' for description.

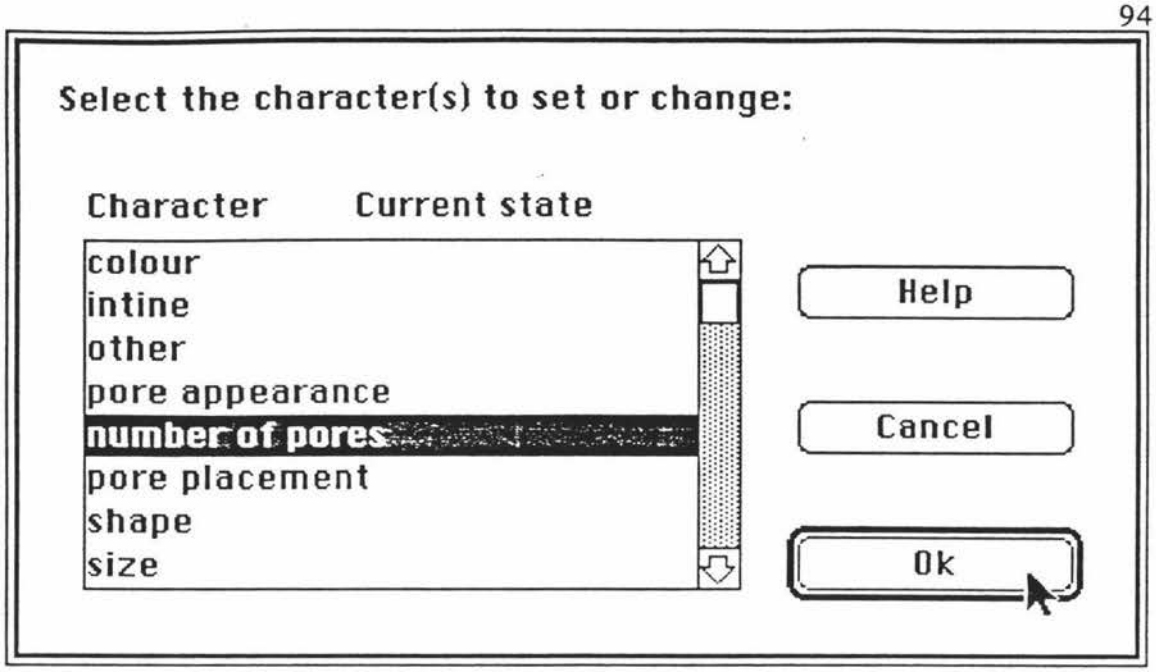


Figure E.18: Character selection dialog.

The system queries the user about the number of pores on the specimen: 4 in this example.

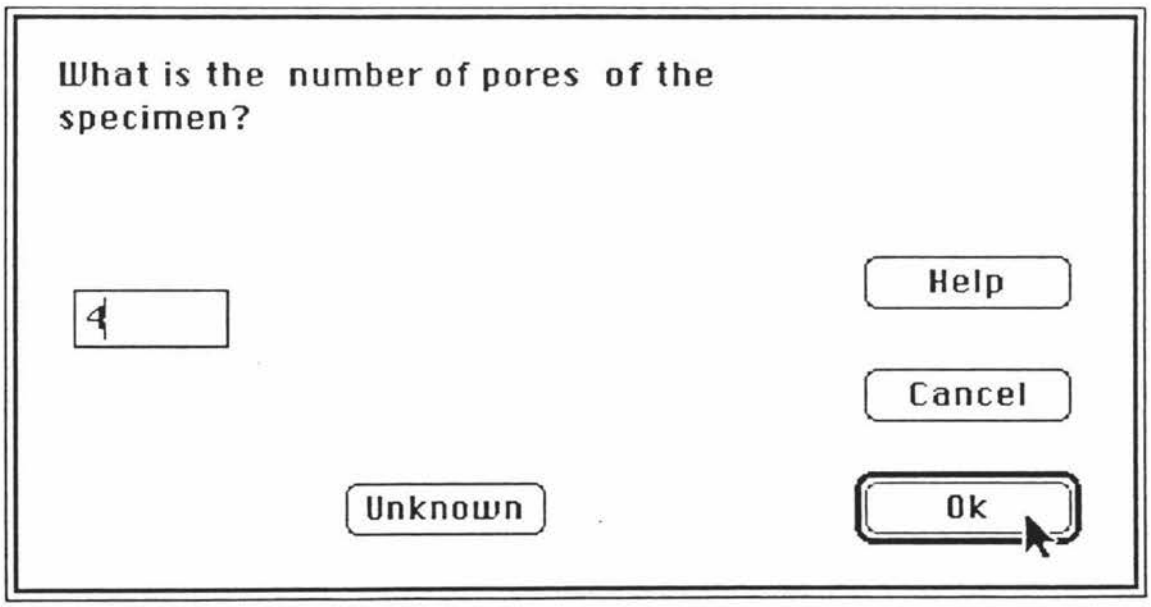


Figure E.19: Pore number dialog.

The system then displays the pollens remaining, and allows the user to select 'Set/Change a character' to continue in multi-access monothetic mode, or to select 'Continue' in the single-access monothetic mode as shown here.

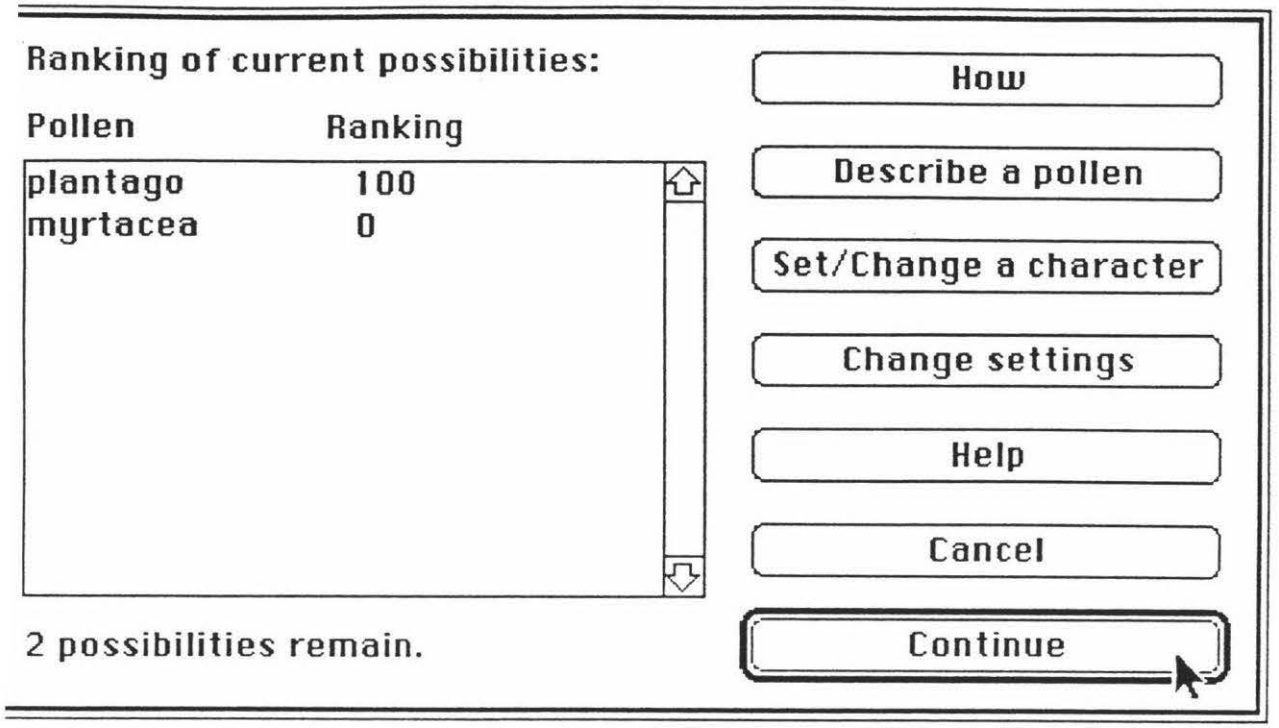


Figure E.20: Pollen ranking dialog.

The system examines the descriptions of the remaining pollens and determines the most important character for the most likely pollen. The user is then queried about that character. In this case, 'pore placement' is the most important character; the user has selected 'At corners'.

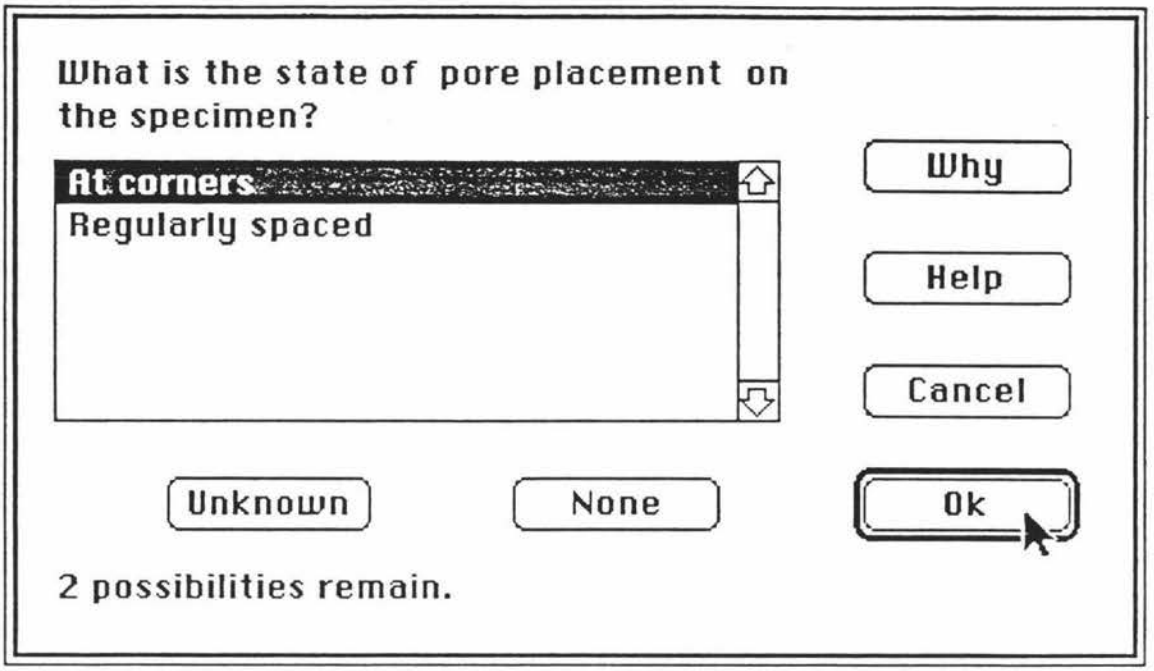


Figure E.21: Pore placement dialog.

Selecting 'Why' in the previous dialog displays the pollen taxa which will be accepted or rejected by the state given for the character.

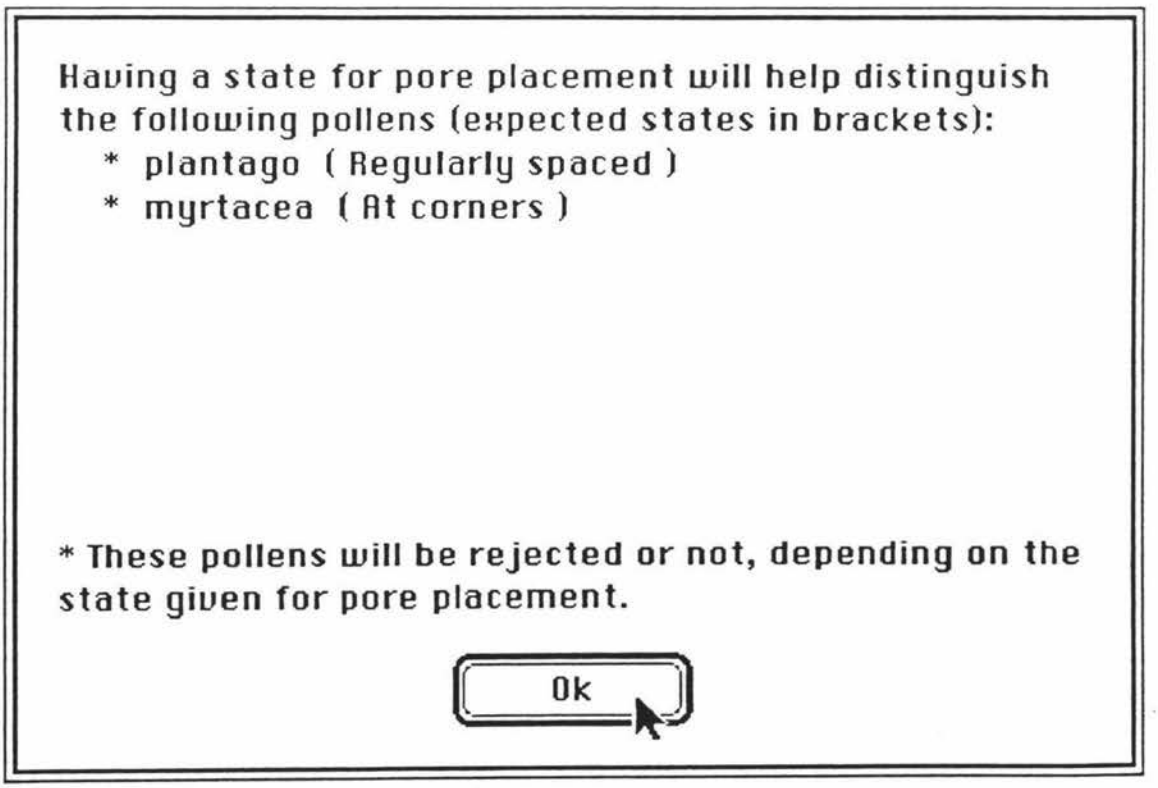


Figure E.22: 'Why' dialog.

Selecting 'At corners' for the character 'pore placement' gives the result myrtacea, as it is the only pollen remaining with the matching character state.

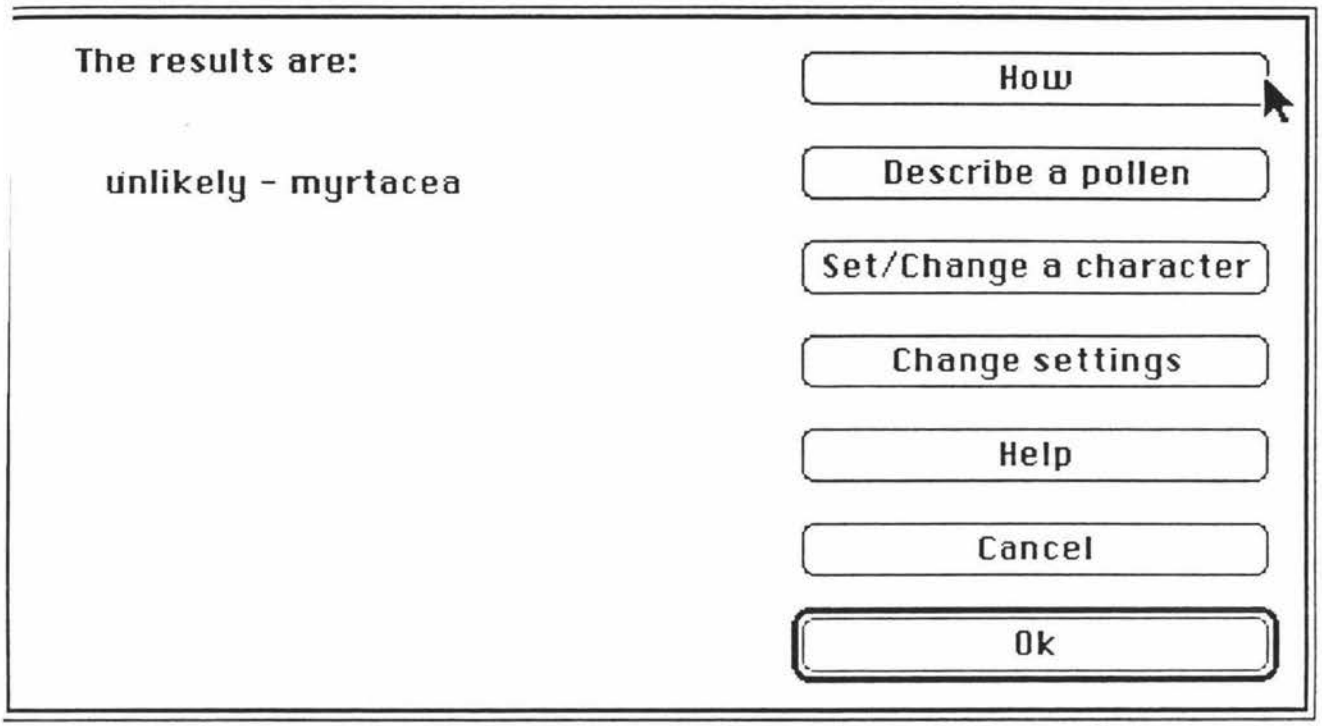


Figure E.23: Result dialog.

Selecting 'How' from the results dialog shows the method used for the identification.

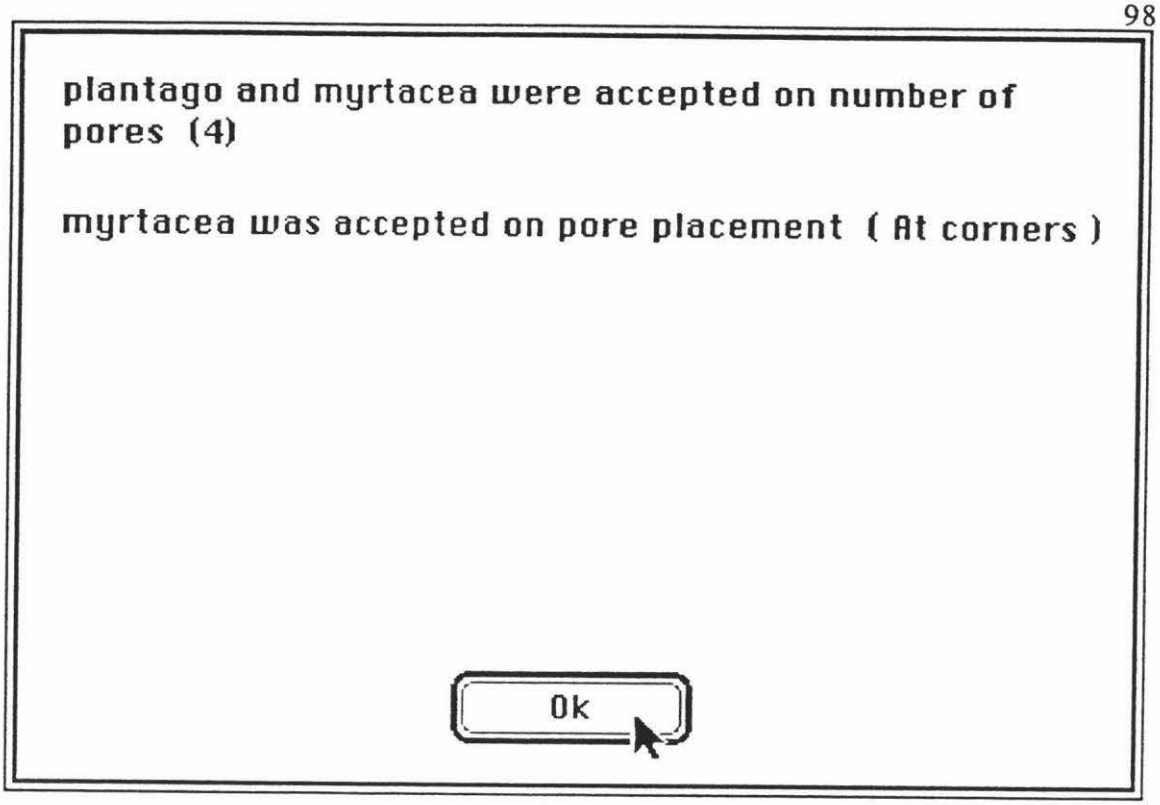


Figure E.24: Explanation dialog.

Selecting 'Describe a pollen' from the results dialog allows the user to request descriptions of one or more pollens. Myrtacea and plantago have been selected in this case.

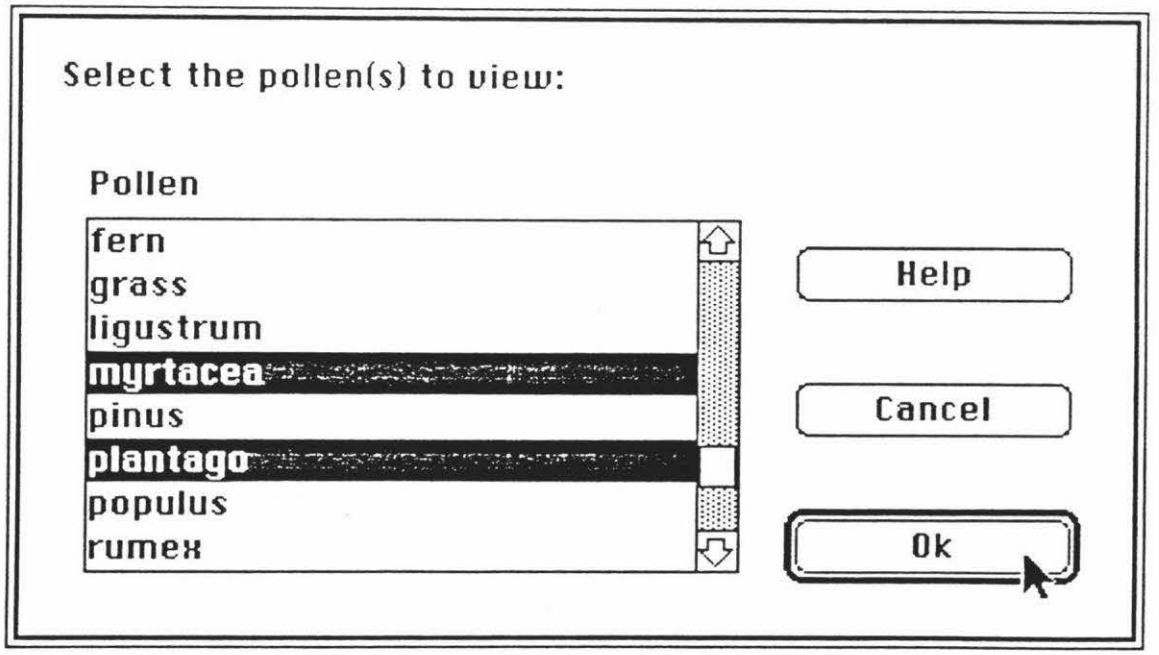


Figure E.25: Pollen selection dialog.

Note the '+', on 'pore placement' and 'number of pores', indicating state matches with the description of the specimen.

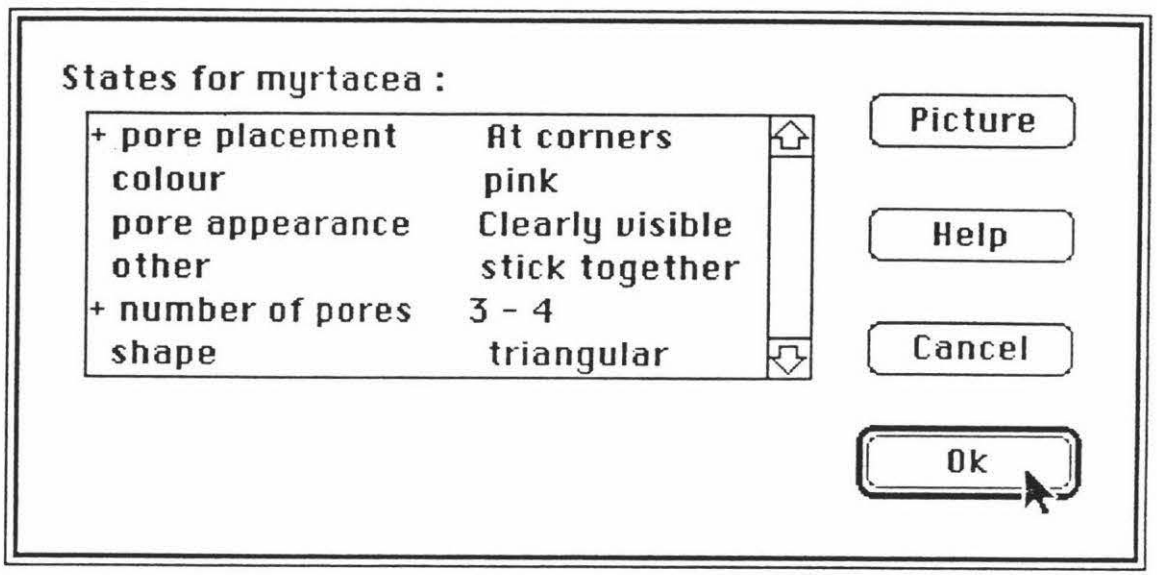


Figure E.26: Pollen description dialog.

Note the '-' on 'pore placement', indicating the state which does not match that of the specimen description.

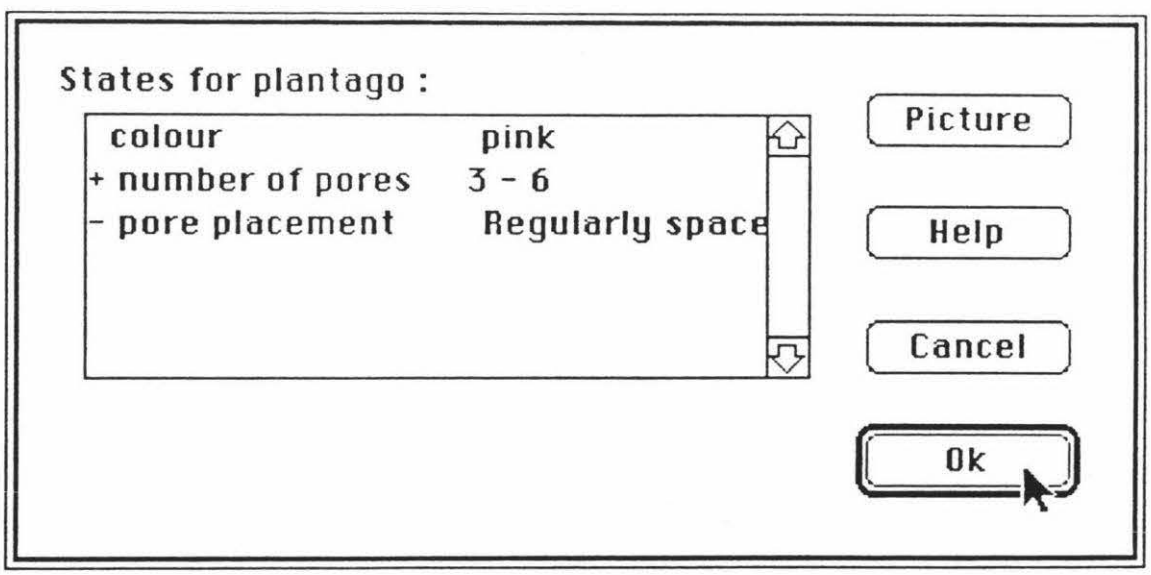


Figure E.27: Pollen description dialog.

# Appendix F

## LPA Prolog for the Apple Macintosh

LPA Prolog for the Apple Macintosh is based on the Edinburgh Prolog syntax, to which has been added a large number of primitives allowing the programmer high-level access to the Macintosh interface. This access allows the user to generate dialogs, manipulate menus and windows, and draw and manipulate graphical objects.

LPA Prolog allows users to implement dialog boxes which can include buttons, check boxes, edit fields, scrolling menus, icons, text and pictures imported from other Macintosh applications. Dynamic dialogues which allow the dialog configuration to be manipulated under the control of a Prolog program are also available. Examples of the use of LPA Prolog dialog boxes can be seen in Appendices A, B, D and E.

LPA Prolog allows menus to be installed, extended, removed or altered easily. Selection of a menu item invokes the Prolog program associated with that item.

Window handling routines allow the user to create, display, hide, move or kill windows using Prolog program calls. Access to standard Macintosh facilities such as the clipboard and desk accessories are available using these routines.

Graphical objects can be drawn and manipulated using Prolog primitives. These have been used in the pollen counting section of the Pollen Identification Expert System, which displays pictures of likely pollens and allows the user to count the pollens he/she can see in the microscope by choosing the appropriate picture using the mouse.

A powerful symbolic debugger is provided with the LPA Prolog system. This allows the user to interact with a program by, for example, setting breakpoints and stepping or jumping through program clauses.

# References

Abbott, L. A., Bisby, F. A. & Rogers, D. J. (1985). Taxonomic Analysis in Biology: Computers, Models, and Databases. New York: Columbia University Press.

Agrawala, A. (1977). Machine recognition of patterns. New York: IEEE Press.

Allkin, R. & Bisby, F. A. (1984) Databases in Systematics. London: Academic Press.

Allkin, R. & Bisby, F. A. (1988). The structure of monographic databases. Taxon, 37, 3, 756-763.

Apple Computer, Inc. (1987). Human interface guidelines: The apple desktop interface. Reading, Massachusetts: Addison-Wesley.

Atkinson, W. D. & Gammerman, A. (1987). An application of expert systems technology to biological identification. Taxon, 36, 4, 705-714.

Ballard, D. H., Brown, C. M., & Feldman, J. A. (1977). An approach to knowledge-directed image analysis. In Pao, Y. & Ernst, G. W. (Eds.) (1982). Tutorial: Context-directed pattern recognition and machine intelligence techniques for information processing. Maryland: IEEE Computer Society.

Baroff, J., Simon, R., Gilman, F., & Shneiderman, B. (1986). Direct manipulation user interfaces for expert systems. Research report CS-TR-1745, University of Maryland.

Baum, B. R. (1986). Computer methods in infraspecific taxonomy of wild and cultivated plants. In Styles, B. T. (Ed.) Infraspecific Classification of Wild and Cultivated Plants. Oxford: Clarendon Press.

Beaman, J. H. & Regalado, J. C. (1989). Development and management of a specimen-oriented database for the flora of Mount Kinabalu. Taxon, 38, 1, 27-42.

Bisby, F. A. (1984). Information services in taxonomy. In Allkin, R. & Bisby, F. A. (Eds.) Databases in Systematics. London: Academic Press.

Bisby, F. A., White, R. J., Macfarlane, T. D. & Babac, M. T. (1983). The Vicieae database project: experimental uses of a monographic taxonomic database for species of pea and vetch. In Felsenstein, J. (Ed.) (1983). Numerical Taxonomy. New York: Springer-Verlag.

Bolc, L. & Coombs, M. J. (1988). Expert System Applications. New York: Springer-Verlag.

Bossert, W. (1969). Computer techniques in systematics. Cited in Morse, L. E. (1975). Recent advances in the theory and practice of biological specimen identification. In Pankhurst, R. J. (Ed.) Biological identification with computers. London: Academic Press.

Brachman, R. J. (1985). What IS-A is and isn't: An analysis of taxonomic links in semantic networks. IEEE Computer, 16, 10, 30-36.

Brachman, R. J. & Levesque, H. J. (Eds.) (1985). Readings in Knowledge Representation. California: Morgan Kaufmann.

Brady, M. (1982). Artificial intelligence approaches to image understanding. In Kittler, J., Fu, K. S. & Pau, L. F. (Ed.) (1982). Pattern recognition theory and applications. Holland: D. Reidel.

Bramer, M. A. (Ed.) (1984). Research and Development in Expert Systems. Cambridge: Cambridge University Press.

Bramer, M. A. (Ed.) (1986). Research and development in expert systems III. Cambridge: Cambridge University Press.

Brandenburg, W. A. (1986). Objectives in classification of cultivated plants. In Styles, B. T. (Ed.) Infraspecific Classification of Wild and Cultivated Plants. Oxford: Clarendon Press.

Bratko, I. (1986). Prolog programming for artificial intelligence. Wokingham, England: Addison-Wesley.

Buchanan, B. G. & Duda, R. O. (1983). Principles of rule-based expert systems. Advances in Computers, 22, 163-216.

Chien, Y. (1978). Interactive pattern recognition. New York: Marcel Dekker.

Clancey, W. J. (1984). Classification problem solving. Proceedings of the National Conference on Artificial Intelligence. California: William Kauffman, Inc.

Clancey, W. J. (1985). Heuristic Classification. Artificial Intelligence, 27, 289-350.

Clancey, W. J. & Bock, C. (1988). Representing control knowledge as abstract tasks and metarules. In Bolc, L. & Coombs, M. J. (Eds.) (1988). Expert System Applications. New York: Springer-Verlag.

Coombs, M. & Alty, J. (1984). Expert systems: an alternative paradigm. Developments in Expert Systems. London: Academic Press.

Cornford, C. A., Fountain, D. W., Burr, R. G. & O'Leary, L. J. (1988). Hayfever in University Students. New Zealand Medical Journal, 101: 520.

Dadd, M. N. & Kelly, M. C. (1984). A concept for a machine-readable taxonomic reference file. In Allkin, R. & Bisby, F. A. (Eds.) Databases in Systematics. London: Academic Press.

Dallwitz, M. J. (1980). A general system for coding taxonomic descriptions. Taxon, 29, 41-46. Cited in Dallwitz, M. J. (1984). Automatic typesetting of computer-generated keys and descriptions. In Allkin, R. & Bisby, F. A. (Eds.) Databases in Systematics. London: Academic Press.

Dallwitz, M. J. (1984). Automatic typesetting of computer-generated keys and descriptions. In Allkin, R. & Bisby, F. A. (Eds.) Databases in Systematics. London: Academic Press.

Dextre Clarke, S. G. (1988). The uses and future of bibliographic database systems. In Hawksworth, D. L. (Ed.) Prospects in Systematics. Oxford: Clarendon Press.

Dodson, D. C. & Rector, A. L. (1984). Importance-driven distributed control of diagnostic inference. In Bramer, M. A. (1984). Research and Development in Expert Systems. Cambridge: Cambridge University Press.

Dodson, J. R. (1972). Computer programs for the pollen analyst. Pollen et spores, 14, 4, 455-465.

Duda, R. O. & Gaschnig, J. G. (1981). Knowledge-based expert systems come of age. BYTE, 9, 238-274.

Ecroyd, C. E. (1986). A key to the identification of 92 species of Eucalyptus found in New Zealand. Forest Research Institute Bulletin, No. 125.

Ellam, S. V. & Maisey, M. N. (1986). A knowledge based system to assist in medical image interpretation: design and evaluation methodology. In Bramer, M. A. (Ed.) (1986). Research and development in expert systems III. Cambridge: Cambridge University Press.

Faegri, K. & Iversen, J. (1975). Textbook of pollen analysis. Oxford: Blackwell Scientific.

Farrell, R. G., Anderson, J. R. & Reiser, B. J. (1984). An interactive computer-based tutor for LISP. Proceedings of the National Conference on Artificial Intelligence. California: William Kauffman, Inc.

Felsenstein, J. (1983). Computers in systematics: One perspective. In Felsenstein, J. (Ed.) (1983). Numerical Taxonomy. New York: Springer-Verlag.

Felsenstein, J. (Ed.) (1983). Numerical Taxonomy. New York: Springer-Verlag.

Fikes, R. & Kehler, T. (1985). The role of frame-based representation in reasoning. Communications of the ACM, 28, 9, 904-920.

Foley, J. D. & van Dam, A. (1984). Fundamentals of Interactive Computer Graphics. Reading, Massachusetts: Addison-Wesley.

Forget, P. R., Lebbe, J., Puig, H., Vignes, R. & Hideux, M. (1986). Microcomputer-aided identification: an application to trees from French Guiana. Botanical Journal of the Linnean Society, 93, 205-223.

Funk, V. A. (1983). The value of natural classification. In Felsenstein, J. (Ed.) (1983). Numerical Taxonomy. New York: Springer-Verlag.

Germeraad, J. N. & Muller, J. (1971). A proposal for a computer based numerical coding system for the description of pollen grains and spores. International Conference on Palynology, 3, 77-80.

Gibbs Russell, G. E. & Arnold, T. H. (1989). Fifteen years with the computer: assessment of the "PRECIS" taxonomic system. Taxon, 38, 2, 178-195.

Gomez-Pompa, A., Moreno, N. P., Gama, L., Sosa, V. & Allkin, R. (1984). Flora of Veracruz: progress and prospects. In Allkin, R. & Bisby, F. A. (Eds.) Databases in Systematics. London: Academic Press.

Gould, J. D. & Lewis, C. (1983). Designing for usability - key principles and what designers think. In Janda, A. (Ed.) (1983). Human Factors in Computing Systems. Amsterdam: North-Holland.

Gunn, C. R. & LaSota, L. (1977). Automated identification of true and surrogate seeds. In Romberger, J. A. (Ed.) Beltsville Symposia in Agricultural Research, [2] Biosystematics in Agriculture. New York: Wiley.

Guppy, J., Milne, P., Glikson, M. & Moore, H. (1973). Further developments in computer assistance to pollen identification. Special publication of the geological society of Australia, 4, 201-206.

Gyllenberg, H. G. & Niemela, T. K. (1975). New approaches to automatic identification of micro-organisms. In Pankhurst, R. J. (Ed.) Biological identification with computers. London: Academic Press.

Hall, A. V. (1975). A system for automatic key-forming. In Pankhurst, R. J. (Ed.) Biological identification with computers. London: Academic Press.

Hawksworth, D. L. (Ed.) (1988) Prospects in Systematics. Oxford: Clarendon Press.

Hayes, P. J. (1979). The logic of frames. In Metzing, D., (1979). Frame conceptions and text understanding. Berlin: Walter de Gruyter and Co. Cited in Brachman, R. J. & Levesque, H. J. (Ed.) (1985). Readings in Knowledge Representation. California: Morgan Kaufmann.

Hayes-Roth, F. (1985). Rule-based systems. Communications of the ACM, 28, 9,921-932.

Hayes-Roth, F., Waterman, D. A. & Lenat, D. B. (1983). An overview of expert systems. In Hayes-Roth, F., Waterman, D. A. & Lenat, D. B. (Eds.) (1983). Building Expert Systems. Massachusetts: Addison-Wesley.

Hayes-Roth, F., Waterman, D. A. & Lenat, D. B. (Eds.) (1983). Building Expert Systems. Massachusetts: Addison-Wesley.

Heywood, V. H., Moore, D. M., Derrick, L. N., Mitchell, K. A. & van Scheepen, J. (1984). The European taxonomic, floristic and biosystematic documentation system - an introduction. In Allkin, R. & Bisby, F. A. (Eds.) Databases in Systematics. London: Academic Press.

Jackson, P. & Lefrere, P. (1984). On the application of rule-based techniques to the design of advice-giving systems. Developments in Expert Systems. London: Academic Press.

Janda, A. (Ed.) (1983). Human Factors in Computing Systems. Amsterdam: North-Holland.

Kanal, L. (1974). Patterns in pattern recognition: 1968-1974. In Agrawala, A. (Ed.) (1977). Machine recognition of patterns. New York: IEEE Press.

Kane, B. & Rucker, D. W. (1988). AI in medicine. AI Expert, 11, 48-55.

Kapp, R. O. (1969). How to know pollen and spores. Dubuque, Iowa: Wm. C. Brown.

Kemp, R. & Boorman, A. (1987). Using WIMPS to beat the expert system blues. Proceedings of the Second New Zealand Conference on Expert Systems.

Kemp, R., Greenwood, J., Tse, A. & Eagle, C. (1988). Maintaining flexibility in expert system design. Proceedings of the Third New Zealand Conference on Expert Systems.

Kendrick, B. (1972). Computer graphics in fungal identification. Canadian journal of botany, 50, 2171-2175.

King, L. (1976). Pollenanalyse und computer: erfahrungen mit palyno, programme zur berehnung und darstellung pollenanalytischer daten. Pollen et spores, 18, 1, 93-104.

Kittler, J., Fu, K. S. & Pau, L. F. (Eds.) (1982). Pattern recognition theory and applications. Holland: D. Reidel.

Lamarck, J. B. A. P. M. de (1778). Flore francoise. Cited in Baum, B. R. (1986). Computer methods in infraspecific taxonomy of wild and cultivated plants. In Styles, B. T. (Ed.) Infraspecific Classification of Wild and Cultivated Plants. Oxford: Clarendon Press.

Lebbe, J., Nilsson, S., Praglowski, J., Vignes, R. & Hideux, M. (1987). A microcomputer-aided method for identification of airborne pollen grains and spores. Grana, 26, 223-229.

Magee, M. & Nathan, M. (1985). A rule based system for pattern recognition that exploits topological constraints. Proceedings IEEE Computer Society conference on Computer vision and Pattern recognition. Holland: IEEE Computer Society.

Mantei, M. & Orbeton, P. (Eds.) (1986). Human factors in computing systems - II. Amsterdam: North-Holland.

Margot, P., Farquhar, G. & Watling, R. (1984). Identification of toxic mushrooms and toadstools (agarics) - an on-line identification program. In Allkin, R. & Bisby, F. A. (Eds.) Databases in Systematics. London: Academic Press.

Melief, A. B. M. & Wijmstra, T. A. (1984). A microcomputer-program for handling palynological data. Pollen et spores, 26, 3-4, 577-586.

Michaelsen, R. H., Michie, D. & Boulanger, A. (1983). The technology of expert systems. BYTE, 4, 303-312.

Moore, J. D. & Swartout, W. D. (1988). Explanation in expert systems: a survey. Research Report ISI/RR-88-228, University of Southern California/Information Sciences Institute.

Moore, P. D & Webb, J. A. (1978). An illustrated guide to pollen analysis. London: Hodder and Stoughton.

Morse, L. E. (1975). Recent advances in the theory and practice of biological specimen identification. In Pankhurst, R. J. (Ed.) Biological identification with computers. London: Academic Press.

Morse, L. E., Pankhurst, R. J., Rypka, E. W. (1973). A glossary of computer-assisted biological specimen identification. In Pankhurst, R. J. (Ed.) Biological identification with computers. London: Academic Press.

Oddy, R. N., Robertson, S. E., van Rijsbergen, C. J. & Williams, P. W. (Eds.) (1981). Information retrieval research. London: Butterworths.

Ogawa, H., Kurioka, S., Kitahashi, T. and Tanaka, K. (1980). An application of knowledge base for image analysis. In Pao, Y. & Ernst, G. W. (Eds.) (1982). Tutorial: Context-directed pattern recognition and machine intelligence techniques for information processing. Maryland: IEEE Computer Society.

Pankhurst, R. J. (1975). Biological identification with computers. London: Academic Press.

Pankhurst, R. J. (1975). Identification by matching. In Pankhurst, R. J. (Ed.) Biological identification with computers. London: Academic Press.

Pankhurst, R. J. (1978). Biological identification: the principles and practice of identification methods in biology. London: Edward Arnold.

Pankhurst, R. J. (1983). An improved algorithm for finding diagnostic taxonomic descriptions. Mathematical Biosciences, 65, 209-218.

Pankhurst, R. J. (1988a). Database design for monographs and floras. Taxon, 37, 3, 733-746.

Pankhurst, R. J. (1988b). An interactive program for the construction of identification keys. Taxon, 37, 3, 747-755.

Pankhurst, R. J. & Aitchison, R. R. (1975). An on-line identification program. In Pankhurst, R. J. (Ed.) Biological identification with computers. London: Academic Press.

Pao, Y. & Ernst, G. W. (Eds.) (1982). Tutorial: Context-directed pattern recognition and machine intelligence techniques for information processing. Maryland: IEEE Computer Society.

Payne, R. W. (1975). Genkey: a program for constructing diagnostic keys. In Pankhurst, R. J. (Ed.) Biological identification with computers. London: Academic Press.

Peters, J. A. (1969). Computer techniques in systematics: discussion. Cited in Morse, L. E. (1975). Recent advances in the theory and practice of biological specimen identification. In Pankhurst, R. J. (Ed.) Biological identification with computers. London: Academic Press.

Poltrock, S. E., Steiner, D. D. & Tarlton, P. N. (1986). Graphic interfaces for knowledge-based systems development. In Mantei, M. & Orbeton, P. (Eds.) (1986). Human factors in computing systems - II. Amsterdam: North-Holland.

Poo, C. D. & Lu, H. (1989). Towards a multi-domain expert system. Research report TRB7/89, Department of Information Sciences and Computer Science, National University of Singapore.

Potter, A. (1988). Direct manipulation interfaces. AI Expert, 10, 28-35.

Ramsey, C. L., Reggia, J. A., Nau, D. S. & Ferrentino, A. (1986). A comparative analysis of methods for expert systems. International Journal of Man-Machine Studies, 24, 475-499.

Rohlf, F. J. & Ferson, S. (1983). Image analysis. In Felsenstein, J. (Ed.) (1983). Numerical Taxonomy. New York: Springer-Verlag.

Romberger, J. A. (Ed.) Beltsville Symposia in Agricultural Research, [2] Biosystematics in Agriculture. New York: Wiley.

Sackin, M. J. (1987). Computer programs for classification and identification. Methods in Microbiology, 19, 459-494.

Sawyer, R. (1981). Pollen identification for beekeepers. Cardiff: University College Cardiff Press.

Shneiderman, B. (1983). Direct manipulation: A step beyond programming languages. IEEE Computer, 16, 8, 57-69. Cited in Baroff, J., Simon, R., Gilman, F., & Shneiderman, B. (1986). Direct manipulation user interfaces for expert systems. Research report CS-TR-1745, University of Maryland.

Smith, E. G. (1984). Sampling and identifying allergenic pollens and molds. San Antonia, Texas: Blewstone Press.

Sneath, P. H. A. (1983). Philosophy and method in biological classification. In Felsenstein, J. (Ed.) (1983). Numerical Taxonomy. New York: Springer-Verlag.

Styles, B. T. (Ed.) (1986) Infraspecific Classification of Wild and Cultivated Plants. Oxford: Clarendon Press.

Thonnat, M., Granger, C. & Berthod, M. (1985). Design of an expert system for object classification through an application to the classification of galaxies. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Holland: IEEE Computer Society.

Tschudi, F. (1988). Matrix representation of expert systems. AI Expert, 10, 44-53.

van Rijsbergen, C. J. (1979). Information retrieval. (Second edition). London: Butterworths.

Walker, D., Milne, P., Guppy, J., & Williams, J. (1968). The computer assisted storage and retrieval of pollen morphological data. Pollen et spores, 10, 251-262. Cited in Morse, L. E. (1975). Recent advances in the theory and practice of biological specimen identification. In Pankhurst, R. J. (Ed.) Biological identification with computers. London: Academic Press.

Watson, L., Dallwitz, M. J., Gibbs, A. J., & Pankhurst, R. J. (1988). Automated taxonomic descriptions. In Hawksworth, D. L. (Ed.) Prospects in Systematics. Oxford: Clarendon Press.

Weiss, S. F. (1981). A probabilistic algorithm for nearest neighbour searching. In Oddy, R. N., Robertson, S. E., van Rijsbergen, C. J. & Williams, P. W. (Eds.) (1981). Information retrieval research. London: Butterworths.

Westfall, R. H., Glen, H. F., & Panagos, M. D. (1986). A new identification aid combining features of a polyclave and an analytical key. Botanical Journal of the Linnean Society, 92, 65-73.

Wolfgram, D. D., Dear, T. J. & Galbraith, C. S. (1987). Expert systems for the technical professional. New York: John Wiley & Sons.

Woolley, J. B. & Stone, N. D. (1987). Application of artificial intelligence to systematics: SYSTEX - a prototype expert system for species identification. Systematic Zoology, 36, 3, 248-267.