

**Generating intelligent tutoring systems
for teaching reading: combining
phonological awareness and thematic
approaches**

Marco Aurelio de Carvalho

Doctor of Philosophy
University of Edinburgh
2000



Agradecimientos

This thesis is dedicated to my wife Deborah for her enormous generosity, patience and support, without which this dream would never come true. The thesis is also dedicated to my father Antero, my mother Isalina, and our children Fernanda, Filipe, Maira and Pedro, who however apart from us most of the time, always made me feel their encouraging presence.

Abstract

The objective of this thesis is to investigate the use of computers with artificial intelligence methods for the teaching of basic literacy skills to be applied eventually to the teaching of illiterate adults in Brazil.

In its development many issues related to adult education have been considered, and two very significant approaches to the teaching of reading were focused on in detail: Phonological Awareness (PA) and Generative Themes. After being excluded from literacy curricula for a long time during the ascendancy of the "Whole Word" approaches, activities for the development of phonological awareness are currently being accepted as fundamental for teaching reading, and are being incorporated in most English literacy programmes. Generative Themes, in turn, were first introduced in Brazil in a massive programme for teaching reading to adults, and have since then been used successfully in a number of developing countries for the same purpose. However, these two approaches are apparently conflicting in their principles and emphasis, for the first (PA) is generally centred on the technical aspects of phonology, based on well controlled experiments and research, whereas the second is socially inspired and focused mainly on meaning and social relationships.

The main question addressed in this research, consequently, is whether these two apparently conflicting approaches could be combined to create a method that would be technically PA oriented but at the same time could concentrate on meaning by using thematic vocabularies as stimuli for teaching. Would it be possible to find words to illustrate all the phonological features with which a PA method deals using a thematic vocabulary?

To answer this question diverse concepts, languages and tools have been developed as part of this research, in order to allow the selection of thematic vocabularies, the description of PA curricula, the distribution of thematic words across PA curricula, the description of teaching activities and the definition of the teaching strategy rules to orient the teaching sequence.

The resultant vocabularies have been evaluated and the outcomes of the research have been assessed by literacy experts. A prototype system for delivering experimental teaching activities through the Internet has also been developed and demonstrated.

Acknowledgements

Many people and organisations contributed for the realisation of this thesis.

Firstly, I would like to express my gratitude to Dr. Helen Pain, my first supervisor, and Dr. Richard Cox, my second supervisor, for their significant participation in the direction of this research.

I thank Dr. Jon Oberlander and Dr. Lyn Pemberton for agreeing to examine this thesis and for their constructive comments on the research and contribution to the thesis text.

I specially thank Barbara Frame for providing material from the literacy programmes Jolly Phonics and THRASS and for participation in the evaluation task of this thesis. Dr. Joyce Watson also deserves my special thanks, for being so generous and encouraging, for coming over to Edinburgh to participate in the evaluation task of this research, bringing with her wise advice and faith in the future.

I am also grateful to my colleagues, members of the AIED group, for forming such a friendly, helpful and respectful research environment, and for helping me in many ways, with comments, suggestions, references, proof-reading, talk attending and participation in surveys. A special thanks to Thomas Segler, also my partner in tennis playing and valuable proof-reader of this thesis, and to Luke Phillips, persistent respondent of the “words survey”. Thanks also to Mark Fraser for his help in proof-reading this thesis. Thanks to Marcio Brandão, my colleague here and in Brazil, for his full-time friendship.

Thanks to colleagues from other research groups such as Maria Lapatta, who helped me with a useful conversation about corpora and thesauri, Ben Curry, who patiently filled in a dozen forms from my “words survey” and rendered endless consultations as the standby native speaker of room F-11. Thanks also to the fifty other colleagues and members of the staff who participated of the words survey included in this thesis.

In the same way, I am grateful for the constant good spirits of the Brazilian community at the University of Edinburgh: Daniela Carbogin, Virginia Brilhante, Alberto Castro, Edjard Mota, Goretti Castro, Herman Gomes, Patricia Machado, Joao Cavalcanti, Wamberto Vasconcelos. Hope to see you back home...

I particularly thank *cumpadre* Paulinho and *cumadre* Helô (Paulo and Heloisa Salles) for pointing me in the direction of the University of Edinburgh.

I thank the support and availability from the University staff, particularly Jean Bunten, Craig Strachan, Neil Brown, John Berry and Olga Franks, the ones I interacted with the most. All the others, however, were always ready to aid me whenever I needed.

I thank Carmen and Geoffrey McGuinness for the productive week during the Phono-Graphix workshop in October 1998 in London, and Reading Reflex for the scholarship awarded to myself to attend that workshop.

I thank Dr. Richihiro Mizoguchi and his team for the opportunity to briefly visit MizLab in Osaka, Japan, and for the interesting discussion and advice on the application and use of ontologies in the construction of tutoring systems.

I thank Cobuild Collins Ltd for authorising the use of the machine readable dictionaries Cobuild Collins and Collins English Dictionary, without which this research would not be possible.

This research was possible thanks to the sponsorship of CAPES - Comissão de Aperfeiçoamento de Ensino Superior, processo número BEX 1238/96-5, from September 1997 to August 2000.

This research was also possible thanks to the allowance of the Department of Computer Science of the University of Brasilia, from which I have been on leave since September 1996 to follow this PhD research.

Finally I would also register the sponsorship of CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico, through the Diretoria de Programas Especiais, Coordenação de Inovação Tecnológica, which, through the individual process n. 280057/96-2 , made it possible to follow this PhD research in parallel with other research in its first year, from September 1996 to August 1997. Particularly I would like to thank Eratóstenes Edson de Araújo for his trust and encouragement, without which this thesis would never have been carried out.

Declaration

I declare that this thesis was composed by myself and that the work contained therein is my own, except where explicitly stated otherwise in the text.

Publications

The following material has been published in connection with this thesis:

Carvalho, M., Pain, H., and Cox, R. (1999). An architecture for a literacy teaching ITS. In Lajoie, S. and Vivet, M., editors, *Proceedings of the 9th International Conference on Artificial Intelligence in Education*, pages 629-631, Le Mans, France.

Carvalho, M., Pain, H., and Cox, R. (1999). An Ontology for a literacy teaching ITS. In Cumming, Murray, T. and Mizoguchi, R., editors, *AIED99 - Proceedings of the Workshop on Ontologies for Intelligent Educational Systems*, pages 93-103, Le Mans, France.

Carvalho, M., Pain, H., and Cox, R. (1999). Automatic vocabulary generation for a literacy teaching ITS. In Cumming, G., Okamoto, T. and Gomez, L., editors, *ICCE99 - Proceedings of the 7th International Congress on Computers in Education*, pages 784-791, Chiba, Japan.

Table of Contents

Chapter 1 Introduction	6
Chapter 2 Literacy Approaches	10
2.1 Writing Systems	10
2.2 Approaches for teaching reading	13
2.3 Phonological Awareness	16
2.3.1 The principles of PA	16
2.3.2 PA teaching practice	18
2.3.3 Experimental evidence in PA	20
2.3.4 PA and Adult Education	22
2.4 Generative Themes	23
2.5 Phonological Awareness versus Thematic Approaches	27
2.6 Related Work	30
Chapter 3 Requirements for a literacy teaching platform	36
Chapter 4 Vocabulary Selection	41
4.1 The source of the words	42
4.2 The Thematic Selection	46
4.2.1 The seed set concept	46
4.2.2 Methods to evaluate target vocabularies	47
4.2.3 The parameters for the selection	48
4.2.4 The definition of the seed set	49
4.2.5 The use of not-words	51
4.2.6 Varying the number of seeds required per entry	53
4.2.7 Allowing the repetition of seeds	54
4.2.8 Selecting with bigger seed sets	55
4.2.9 The use of examples	56
4.2.10 Refining the source of words	58
4.2.11 The use of and-words	59

4.2.12	Discussion and recommendations on seed sets	61
Chapter 5	The Phonological Selection	65
5.1	The description of the phonological features	65
5.2	Introducing the Knowledge Tree concept	69
5.3	Instantiating PA Approaches	72
5.3.1	The instantiation of Phono-Graphix	72
5.3.2	The instantiation of Jolly Phonics	77
5.3.3	The instantiation of THRASS	85
Chapter 6	Analysis of the resultant vocabularies	88
6.1	Setting up the environment	88
6.2	The vocabularies for Phono-Graphix	90
6.3	The vocabularies for Jolly Phonics	94
6.4	The vocabularies for THRASS	98
6.5	Final comments on vocabularies	101
Chapter 7	The ITS Generation	102
7.1	System Architecture	103
7.2	Knowledge Tree	107
7.3	Activities Description	109
7.3.1	Start-Activity Object	112
7.3.2	End-Activity Object	112
7.3.3	Input Object	112
7.3.4	Output Object	112
7.3.5	Show-Scene Object	113
7.3.6	Yes-or-No Object	113
7.3.7	Choice Object	113
7.3.8	If Object	113
7.3.9	Jump Object	114
7.3.10	Set-Var Object	114
7.3.11	Make-Set Object	114
7.3.12	Match Object	114
7.3.13	Making Sense of the Activities Description	114
7.4	Teaching Strategy	115
7.5	Student Model	118
7.6	Observations on the Prototype	122

Chapter 8 Case Studies	124
8.1 Phono-Graphix Applications	124
8.1.1 Pointing out the first sound	124
8.1.2 Word building	130
8.2 Jolly Phonics Applications	134
8.2.1 Missing Sound	134
8.2.2 Adding Sounds	139
8.3 THRASS Applications	142
8.3.1 Identifying Words	142
8.3.2 Alternative spellings	146
8.4 Summary of the Case Studies	150
Chapter 9 Teaching Strategy Application	151
9.1 General Description of Short Phonix	151
9.2 Short Phonix Vocabulary	152
9.3 Short Phonix Knowledge Tree and Activities	152
9.4 Short Phonix Teaching Strategy	156
9.5 Strategy Simulation	158
9.6 Teaching Strategy Summary	162
Chapter 10 Evaluation of Pedagogical Issues	163
10.1 First Part - General Presentation	165
10.2 Second Part - Vocabulary generation	165
10.3 Third Part - Curriculum Description and Vocabulary Distribution	170
10.4 Fourth Part - General Evaluation	178
10.5 Comments on the results of the evaluation	181
Chapter 11 Discussion and Conclusion	182
11.1 Outcomes of the research	182
11.2 Questions and Answers	184
11.3 Sensitive Points	185
11.4 Future Work	186
11.5 Final Evaluation	188
Appendix A Literacy Teaching Dictionary Format and Example	197
Appendix B Survey for Evaluating Relevance of the Target Vocabulary	200
B.1 Survey description	200

B.2	Summary of the Observed Vocabularies	203
B.3	Comparison of Evaluations	207
Appendix C	Festival System Phonemes	209
Appendix D	Knowledge Tree Codes	211
D.1	The Phono-Graphix Knowledge Tree	211
D.2	The Jolly-Phonics Knowledge Tree	214
D.3	The THRASS Knowledge Tree	218
Appendix E	Vocabularies	224
E.1	Food Vocabulary	224
E.2	Sea Vocabulary	228
E.3	Agriculture vocabulary	232
E.4	Freire vocabulary	236
E.5	Common words vocabulary	241
E.6	Phono-Graphix vocabulary	245
E.7	Jolly Phonics vocabulary	248
E.8	Tricky words vocabulary	250
E.9	THRASS vocabulary	251
Appendix F	Grammars	252
F.1	Knowledge Tree Definition Language - KTDL	252
F.2	Teaching Strategy Definition Language - TSDL	254
Appendix G	Object Specifications	256
G.1	Start-Activity Object	256
G.2	End-Activity Object	257
G.3	Input Object	258
G.4	Output Object	260
G.5	Show-Scene Object	260
G.6	Yes-or-No Object	261
G.7	Choice Object	263
G.8	If Object	263
G.9	Jump Object	265
G.10	Set-Var Object	265
G.11	Make-Set Object	267
G.12	Match Object	270

Appendix I Tools

- I.1 Vocabulary Tools 276
- I.2 Authoring Tool 279
- I.3 Lesson Interpretation 280
- I.4 Software Summary 280

Introduction

Russia is a large and very significant country for the world economy. It has plenty of natural resources, forests, minerals, vast arable land and is blessed with a mix of weather systems and topographic features that enable it to be made a strategic auto-reliance in the production of other earth energy resources. It is a rich nation.

However, this resource is not fairly distributed among the population. This is a long term problem which has been taking place for a long time, even before the independence. One of the major reasons is the 90% illiteracy. Today about 20% percent of the Russian adult population are still illiterate due to the lack of teachers, let alone people to attend schools. In some places there is 90% illiteracy to teach them. The young people would not see opportunities and think the government is not taking the necessary steps to solve the problem. Many industries have closed plants in the thick forests, high hills and mountains, but generally only maintaining the stock of the deep protection forest.

The initial objectives of the research was to investigate what kind of need the use of computers and artificial intelligence could bring to the teaching of foreign languages in the classroom practices. This is a very broad question, with many possible answers. Firstly, it is not clear how many non-European people would get into the computers and Internet and how to integrate. Secondly, if they do not use it at all, how could they communicate with computers? Thirdly, which language approach would be used in a big country with plenty of regional cultures and local language groups? And, on top of all that, how would software developed for the English language be transferred to the foreign language system in Brazil?

All these questions led to a research which was the highlight of this research. The history of the writing system of the Latin alphabet had to be studied. The evolution of the alphabets was then shown and explained through language area

Chapter 1

Introduction

Brazil is a huge and very significant country for the world economy. It has plenty of natural resources, plants, minerals, vast unexplored land and is blessed with a mix of weather regimen and topographic features that enables it to be much more than auto-sufficient in the production of clean electric energy from hydraulic sources. It is a rich country.

However, this richness is not yet fairly distributed among the population. This is a long term process which has been taking place for a long time, since before the independence from the Portuguese colonisers in the XIX-th century. Today, around 17 percent of the Brazilian adult population are still illiterate, due to the lack of opportunity for these people to attend schools. In some places there is simply no one to teach them. The young people would try new opportunities and jobs in the bigger centres, often leaving the country-side population depleted. Many initiatives have taken place to tackle this situation, some with good results, but generally only scratching the surface of the deep problem beneath it.

The initial motivation of this research was to investigate which kind of relief the use of computers and artificial intelligence could bring to the teaching of literacy to adults in the context just described. This is a very broad question, with many possible answers. Firstly, it is not obvious how these poor illiterate people would get access to computers and literacy teaching programmes. Secondly, if they are not able to read at all, how could they communicate with computers? Thirdly, which literacy approach should be used in a big country with plenty of regional culture and local language expressions? And, on top of all that, how would an effort developed for the English language be transferred to the Portuguese language spoken in Brazil?

All these questions have increased more and more the interest of this research. The history of the writing system and its basic concepts had to be studied. The evolution of the alphabets and their reuse and adaptation through languages were

also in focus. The approaches for the teaching of literacy were scrutinised and some choices had to be made. The issues related to adult education came up at various times in different focuses: social motivation, psychological adequacy and pedagogical efficiency. The technological questions were also examined to support the choice of the software and hardware necessary to materialise whatever solution should be proposed. Finally, the social and political aspects related to a project of this nature were considered to guarantee that its outcome could be regarded as economically implementable and manageable.

The thesis itself does not tackle all the problems just referred to. It rather concentrates on the aspects related to the generation of the vocabulary to be used as stimulus for the teaching of literacy to adults, and on the organisation of the curriculum and the activities to carry on the teaching task in focus.

As indicated and justified in Chapter 2, preference has been given to *Phonological Awareness* (PA) approaches for teaching reading as the pedagogical line of this thesis. However PA methods are normally aimed at children and people with reading disabilities, and the PA pedagogical material available is generally designed for these groups.

As the target population of this research are adults, considerations were taken on the thematic approaches, and particularly on the *Generative Themes* principle proposed by the Brazilian educationalist Paulo Freire, who had a good deal of positive experience while teaching literacy to Brazilian adults. Freire's methods, however, are socially inspired, based on a mix of life experience, community folklore, and emotion, contrasting with the experimental support behind the orientation of PA approaches. Consequently, the main question of this thesis is: *Is it possible to combine these two apparently conflicting approaches to obtain a productive way for teaching adults to read?*

It might be not straightforward for teachers to identify the appropriate vocabulary in order to adapt existing reading schemes for accomplishing the union of these approaches. Could we use our knowledge of Artificial Intelligence and Education to provide teachers with pedagogical materials for this end? More specifically, in this thesis we address the following questions:

1. Can we devise automatic ways for generating vocabularies for Generative Themes?
2. Can we define the means for structuring a PA curriculum so it can facilitate the use of a thematic vocabulary?
3. Can thematic vocabularies support PA approaches?

4. Can such materials be incorporated in teaching environments to provide the range of activities needed to more fully support reading schemes in which such PA approaches are embedded?
5. What would be the practical implications of the delivery of such an environment?

To address these questions we developed a series of tools to support the selection of vocabulary, the definition of curriculum and the implementation of activities that could run on a computer system as a complementary aid for teaching reading. These activities can be delivered through a computer network compatible with the Internet. A prototype of the proposed system has been implemented. Different literacy methods have been tried with the system, which has been evaluated by volunteer experts in literacy. The unanswered questions left from this work are to be addressed in future research.

The structure of the thesis follows:

- **Chapter 1** poses the problem to be addressed and outlines the related questions.
- **Chapter 2** discusses the basis of the writing systems, the history of literacy teaching and the various approaches for teaching reading, with their advantages and difficulties.
- **Chapter 3** presents the requirements a platform for teaching literacy to adults should attend to and describes briefly the support tools developed during the research.
- **Chapter 4** discusses and presents methods for acquiring and selecting vocabularies according to thematic criteria and describes the vocabularies generated in this research.
- **Chapter 5** describes the concept of a Knowledge Tree for describing the curriculum of PA literacy methods and shows its application for defining the curriculum of three different literacy methods, along with the specifications for distributing the thematic vocabulary across the methods' curriculum items.
- **Chapter 6** analyses the results of the distribution of the vocabularies across the curriculum of the various literacy methods considered and summarises recommendations on this issue.

- **Chapter 7** describes the architecture of a computer environment developed as part of this thesis to support the generation of Intelligent Tutoring Systems for implementing PA literacy programmes, making use of the generative themes vocabularies previously obtained.
- **Chapter 8** presents three case studies of the representation of activities of different methods with the authoring tool developed as part of this research.
- **Chapter 9** describes the application of the scheme for defining the teaching strategy for a reduced literacy teaching method developed to be used as an example.
- **Chapter 10** describes the scheme used to evaluate the work done from the pedagogical point of view and presents the result of the evaluation.
- **Chapter 11** closes the thesis with a general discussion and presents the conclusions.

In addition, nine appendices contain complementary information for the discussion and proposals contained in this thesis.

Chapter 2

Literacy Approaches

2.1 Writing Systems

Most people who can read take it for granted and have even forgotten the details of how they learned. However, learning to read is by no means a simple process. There are many different ways of teaching learners to read, and this area has been in focus for a long time. Before we talk about the methods for teaching reading let us take a look at the principles of writing systems and how they evolved.

Writing systems are ways of recording messages through visual symbols. A major classification of writing systems would be “logographic”, “syllabic” and “alphabetic”.¹ The first known writing system would have been developed by the Sumerians around 3500 B.C., consisting of logograms used to record data about agricultural production and commerce in general in clay tablets. The main principle of logographic systems is the representation of each word through some graphic symbol. As many words do not have any evident representation and as some representations can address or suggest different words, such systems must rely on a number of conventions to avoid ambiguity. The other drawback of this kind of writing system is the limitation of visual symbols that can be recognised. Nevertheless, even today many logographic writing systems are in use, their major representative being the Chinese writing system.²

A thousand years of experience using logographic representations of the language brought the Sumerians the insight that, instead of representing full words, it would be a good idea to represent parts of the words through graphic symbols. A similar development happened around the same time in other writing systems such as the Egyptian hieroglyphic script. Such changes produced the first syllabic

¹Gelb’s discussion of this issue admits this classification whilst presenting the inconsistencies of the model [Gelb, 1952, p.16-17].

²The full Chinese writing system contains as many as 80,000 logograms. A Chinese adult would on average be familiar with about 5,000 logograms [Adams, 1990].

writing systems.

Syllabic systems can be very efficient for graphically representing some languages. The Japanese language, for instance, has only 74 syllables composed either of vowel-only phonemes or consonant-vowel phonemes. The Japanese writing system is rather complex because it has been derived from the Chinese system, and there are several different ways for expressing the words on paper, depending on the context. However, the system includes two different syllabaries which can be used independently to represent any written message: the Katakana and the Hiragana. The Katakana syllabary contains 47 basic graphic symbols to represent the sounds of the language, with some other symbols used to modify the basic sounds. The Hiragana is more complex than the Katakana, and possess one symbol for each syllable of the language [Gelb, 1952].

The English language, on the other hand, has a very complex syllabic structure, admitting agglutination of fifteen different syllabic types: CV, CCV, CCCV, CVC, CCVC, CCCVC, CVCC, CVCCC, CCVCC, CCVCCC, CCCVCCC, VCCC, VCC, VC, V [McGuinness, 1997].³ As a consequence, it is estimated that the English language might have thousands of different syllables in its words [Rozin and Gleitman, 1977]. That leaves us with the idea that a syllabary of thousands of symbols would be needed to represent all English words on paper. This is clearly not the best solution, although it could be much better than representing each English word by one symbol.

The third way for graphically representing a language is through alphabets. The great advantage of the use of alphabets lies in the economy of symbols they allow. An ideal alphabet should have only enough symbols for representing the smallest units of sound of the language: the phonemes. Different languages have different number of phonemes: most people agree that English has 44 phonemes, whereas Portuguese, for instance, is believed to contain 33 phonemes. But none of them have as many symbols as they have phonemes. Generally speaking the alphabets have been adapted to serve new languages since they have been invented and in many of the cases they do not suit perfectly the new languages they represent.

The English alphabet is a descendant of the Phoenician alphabet, originated at Byblos around 1050 B.C. The Phoenician alphabet itself is one of a series of experimental alphabets developed in the Middle East since 1700 B.C. The Phoenicians created a consonantal alphabet in which the reader should guess the vowels not represented relying on the context in order to retrieve the right sense

³C stands for consonant and V stands for vowel.

of the words. The Greeks borrowed and adapted the Phoenician letters to form their alphabet around 900 B.C. In the seventh century B.C. the Etruscans and Romans borrowed the Greek alphabet to create their own. Eventually, because of the Roman conquests and the consequent spread of the Latin language, the Roman alphabet became the basic alphabet of all the languages of western Europe [Diringer, 1947]. Whereas the Greek alphabet was a perfect fit for the Greek language, with a one-to-one relationship between graphemes and phonemes, the English alphabet ended up with 26 letters to represent the 44 sounds of the language. Clearly this is only possible with the use of combinations of letters to represent some of the sounds of the language. Moreover, unfortunately, this is not the only problem. As we know, some sounds in English words also have multiple graphemic representation.

Since the creation of an alphabet for English, the language itself has suffered change and interference from another languages, mainly from French spelling during the Middle Ages. English is today a language spoken more by non-native speakers than by native speakers. Although it possesses a relatively simple grammar which might place it as the *de facto* universal human language, its spelling is complex, which makes it a difficult language to write or speak correctly.

Most of modern English spelling standardisation is due to Samuel Johnson, who in 1755 published "A Dictionary of the English Language" with the aim of providing accurate definitions of words and of standardising the pronunciation of the English language. Although even today no one would say that this has been totally achieved, Johnson contributed significantly by compiling the best works of that time, and producing what has become the basis for current English spelling. Further work such as John Walker's "A Critical Pronouncing Dictionary and Expositor of the English Language" (1771), Noah Webster's "An American Dictionary of the English Language" (1828) and Joseph Worcester's "A Dictionary of English Language" (1860), among others, have amazingly guaranteed that a language spoken in many far away countries could keep the same basic features and allow all these people to talk amongst themselves with little difficulty [Balmuth, 1982, p.133-135].

Despite this enormous effort and brilliant work the result is still a very complex writing system, which has defied linguists and educators when dealing with methodologies for teaching reading.

2.2 Approaches for teaching reading

The evolution of English reading teaching is a long term and cyclic process. Charles Fries [Fries, 1962] reports an amazing succession of discoveries and reforms from 1570 to 1900, in a way much like what has happened in this century. The two main streams that have dominated the last four hundred years are generally called “Phonics” and “Whole Word”, although some people refuse to categorise their particular method as belonging to one or the other.

Although these competing schemes have been active for a long time, the most commonly used practice in teaching reading up to the beginning of the 20th century was based on dull memorisation of letters and drilling on the “sounds of the letters”. This approach is called by many the “Alphabet Method”. It must be observed that most of the teaching of reading in the past was at home rather than in schools. As the matter has become a public concern, “scientific evaluation” of these methods has come to the fore with the “whole word” revolution dominating the arena for almost forty years from the 1920s.

Whole word approaches in a broader sense encompass a number of methods that discourage direct contact with the formal writing system at first, developing instead activities to make the student familiar with the graphic representation of common words as a whole, emphasising meaning and relying on context clues for aiding their recognition. [Fries, 1962, p.11-17] mentions the “Word Method” early in the nineteenth century, with quotations from “The Common School Journal” and from “A New Method of Teaching Children to Read: Founded on Nature and Reason” by J. Russel Webb that illustrate a great many of the basic ideas of the whole word approaches:

...My theory is that words can be more easily remembered than letters...

...I frequently tell little children who know the alphabet to look at the word rather than to spell it over...

...It is desirable that in the first book there should be many repetitions of the same words ... well that is not so arranged as to have some natural sequence ... bird, tree, wings, feathers, bill...

...After learning a few group of words often repeated on a page, let these be combined in short sentences...

...Children of six, who begin to read thus by learning words instead of letters, will be able in three month to read simple stories very easily...

...The child, in this part, is not to be taught a letter, or to spell a word, but is simply to learn the words by their forms ... The child thus reads naturally, by sight ... reads with ease and pleasure...

... The first word which the child is to be taught is the word BOY ... The

next word is GOOD. This, when learned, can be combined with the word boy, making the first phrase, “good boy” in the reading lesson...

Amongst the rationale for whole word approaches in the 20th century, however, there are more serious concerns which are not visible in the quotations above. The contributions of Maria Montessori [Montessori, 1912; Smith, 1912], Froebel [Lilley, 1967] and the influences of Gestalt theories [Diack, 1960; Kohler, 1969] brought new ideas into primary education, that were soon associated with the whole word approaches: the word *learn* began to replace the word *teach*, the study groups became smaller and more informal, contrasting with large and formal classrooms, and an emphasis on the interest and happiness of the children replaced the former concern with discipline [Southgate and Roberts, 1970, p.36-44]. By the fifties, all these reinforcements made the whole word ideology look like the ultimate answer for the complex task of teaching children to read and write English.

In 1955 Rudolph Flesch [Flesch, 1955] passionately invited parents to a discussion he considered no longer to belong to “experts” and educators. In a mixture of report, novel and teaching manual he brilliantly re-inflamed the debate, and motivated a series of experimental works that up to today shows no sign of fading.⁴

Phonics approaches should not be associated with the old Alphabetic Method, but they bring up issues about the concreteness of the writing system, and the necessity of being aware of the sounds contained in the words and the relationship between sounds and their representation on paper. According to [Adams, 1990, p.51-52] “... there exist many, many such programs - each of Robert Aukerman’s books cite over 100... To be sure, a central tenet of each of these programs is that working knowledge of the letter-sound correspondences underlying our system of writing is key to proficient reading. Beyond that they differ greatly...”.

Charles Fries proposed the use of the word *phonics* to represent the various sets of teaching practices that aim to develop the pupil’s ability to “sound out a word” and collected quotations for exemplifying the variety of *phonics* practices from 1925 on. The following excerpts are parts of those quotations used here to illustrate what phonics approaches are about. The complete quotations and references are in [Fries, 1962, p.141-146].

... there are only 26 letters... these letters represent more than 40 sounds. With these few sounds we are able to speak all of the words we use. In order to pronounce words correctly and to learn to pronounce new words,

⁴See [Adams, 1990] and [McGuinness, 1997] for a review of the research.

it is necessary to learn what these sounds are and how they are represented in print ...

... A study of phonetic sounds will be of little assistance at first in word recognition ... The ears of the pupils must first be trained to recognise sounds and their minds must be trained to connect these separate sounds into words...

... At first it will be necessary almost to speak the word before most of the pupils will recognise it ... (with practice) ... the pupils will soon become very skillful at interpreting words when their sounds are given aloud by the teacher... Next ask the pupils to listen to the sounds of a word given by the teacher. Have them repeat the sound heard, and pronounce the word.

...

The pupils are now ready to begin learning the letter symbols. Sound aloud the word *man*. Write it on the board: m a n. Sound it, pointing to the sounds, and have the children sound the word ... In the same manner teach s, c, l, r, f, g and t by sounding then words like sat, cat, lad, rat, fat and glad. Give much practice in sounding the letters from the board ...

... attention to initial and final letters of phonograms, noticing common parts of words - syllables, double letters, and so forth, finding little words in bigger ones ... and contrasting reversal words - *was* and *saw*, for example...

... This little book ... offers rhymes and pictures that make it easy to identify the sounds of speech and letters that represent them...

... This method differs basically from other phonic methods in that it does not start with reading, which is, in phonics, the translating of printed letters or words into the sounds of spoken words ... In fact ... training in the *blending* of the sounds in a syllable is needed for some [children]. When a child has once mastered the phonic tools, he is able to decipher and pronounce and understand any printed word which come within his speaking vocabulary...

... The child breaks down the word into initial, terminal and first vowel sounds. The consonant sounds are whispered or voiced so as to avoid the vowel appendage as much as possible. The vowel is found by breaking down the initial sound plus the vowel. Then he proceeds to initial blends, the vowel r, and diphthongs...

In this last set of quotations we can see clearly the great concern of phonics supporters with the comprehension of the mechanics of the writing system as opposed to whole word school of thought. But it is also clear that there is not only one “phonics” approach, and there is not only one “whole word” approach. In particular, “phonological awareness” approaches which are one of the bases of this thesis, are not generally complete teaching reading programmes, but part of the process. They serve the purpose of increasing the learner’s ability to recognise the sounds of the words in the first stages of the programme, or in remedial interventions for students with learning difficulties.

In this research, efforts are concentrated on two particular approaches to teaching reading, one of which would be undoubtedly placed aside phonics approaches whereas the other would stay closer to whole word approaches, although with some concession to phonic activities. These approaches are detailed in the next two sections.

2.3 Phonological Awareness

2.3.1 The principles of PA

Although further work has revealed the value of phonological awareness in regular and failure-preventive literacy schemes, most of the vast research in the area has been aimed at clinical and remedial schemes for teaching children with learning disabilities to read. This rather closed universe, when compared to research with *able* children, must have given the researchers the opportunity to control more strictly the experiments allowing some more solid conclusions to be drawn.

One of the major contributors in this area, Isabelle Liberman, started her work with learning disabled children in the Children's Hospital in Newington, Connecticut. In 1966 she created a graduate program in learning disabilities in the School of Education at the University of Connecticut, that has trained many researchers who later joined her in her work with her close collaborator Donald Shankweiler [Shankweiler, 1991]. At the same time, at the Learning Research and Development Center at University of Pittsburgh, Jerome Rosner and Dorothy Simon were working in the design of their Auditory Analysis Test (AAT) used to evaluate the ability of children to deal with the removal of sounds of words (for instance "say *cowboy* without the *cow*"). With this test they demonstrated the strong connection between the ability to identify and manipulate the sounds in words and the capacity to read and understand text, with a major impact in the reading community. Other parallel work, such as the Lindamood Test of Auditory Conceptualisation (LAC) developed by Pat and Charles Lindamood also highlighted the same kind of correlation [McGuinness, 1997]. A correlation does not necessarily determine a cause-effect relationship. Although the correlation between the ability of reading and the phonological awareness has been established, a lot of work has been done to try to determine the direction of the implications. According to [Goswami and Bryant, 1990, p.4] "There are two possibilities. One is that children learn how to divide words up into their constituent sounds because they are taught to do so when they learn to read... The second possibility is ... before children learn to read they may build up phonological skills which

then affect how well they learn to read”.

The work of Liberman, Lindamood and other similar early works played an important role in the consolidation of the phonological awareness research community. Before going deeper into the description of particular tasks and tests used in checking and developing phonological awareness, let us state some of its principles.

Broadly speaking, phonological awareness is the awareness of the various sounds or groups of sound on which words are built. This awareness come at different levels: *syllables, onsets and rimes*⁵ and *phonemes*.⁶

The more obvious of these levels is certainly the *syllable* level. Many of the words of a language are monosyllables and most of the simple and “first-words” one learns are monosyllable words. These phonological units do not require much effort from the learners to be identified. Moreover, the English language has too many syllables, and too many different graphic representations for its syllables, to qualify *syllables* as good breaking units for the teaching of reading and writing English [Tunmer and Rohl, 1991, p.5] [Goswami and Bryant, 1990, p.2].

The next level of phonological awareness is on *onsets and rimes*. *Onset* is the name given to the first part of a broken syllable, whereas *rime* is the name of the last part of it. Onsets and rimes are also called intra-syllabic units. For their universality and popularity, rhymes are present in many pre-reading programmes, and indeed play an important role in bringing some degree of phonological awareness to young children [Goswami and Bryant, 1990; Maclean et al., 1987; Treiman, 1991, 1996].

The third level is the *phonemes* level. These are the smallest units of sounds which can change a word in a language. For example, “cat” and “mat” are different words just because they have the first phoneme different from each other. Young children or illiterate adults are generally not aware of phonemes within the words and the bi-causal relationship between phoneme awareness and literacy has been experimentally demonstrated by various researchers such as Bradley and Bryant [1983]; Moraes et al. [1986]; Perfetti et al. [1987]. As mentioned before, a perfect alphabet should represent uniquely the sounds of the language it is designed for. Unfortunately this is not the case in English, and many spelling variations are possible for representing sounds in English written words. This

⁵The word *rime* is an archaic spelling form for *rhyme* that is rather used in this context for defining the last part of a syllable.

⁶As in many others concepts within the literacy area, not everybody agrees with this definition. [Tunmer and Rohl, 1991, p.5] for instance, suggest that “phonological awareness is just that awareness of phonemes: It is not awareness of syllables, awareness of intra-syllabic units, or awareness of words”.

adds a great deal of complexity to the process of learning to read English.

2.3.2 PA teaching practice

Some early experiments have had impressive results and influenced the majority of the methods today used in phonological awareness methods for teaching reading.⁷ The *List of Phonemic Awareness Tasks* elaborated by [Lewkowicz, 1980] describes a very comprehensive set of activities used to test and develop students' phonological awareness:⁸

1 - *Sound-to-word-matching* - the recognition of a previously specified phoneme within a word, for instance: "Does *fish* start with /f/?" (not necessarily only with phonemes in the beginning of the word).

2 - *Word-to-word-matching* - the recognition that a word has the same beginning sound, or same final sound or same intermediate vowel as another word.

3 - *Recognition of rhyme* - the recognition that a word is identical to another one after the stressed vowel (for instance: "does *fish* rhyme with *dish*?").

4 - *Isolation of a beginning, intermediate or final sound* - the pronunciation of the phoneme occupying a designated location in a given word ("What is the first sound of *fish*?" or "What is the last sound of *dog*?").

5 - *Phonemic segmentation* - separately isolating all the sounds of a word in the right order ("What are the three sounds of *fish*?").

6 - *Counting the phonemes* - answering to questions like "how many sounds there are in the word *fish*?" or asking to tap while saying the isolated sounds of a word (e.g. "tap the sounds of *dog*").

⁷The test procedures developed in early PA research are very significant to the present thesis: the tutoring system support associated with this research (described later) offers expert literacy teachers a number of primitive procedures, called *objects*, related to those platforms, to allow the description of the activities of their own literacy schemes to be delivered by a computer system.

⁸[Lewkowicz, 1980] contains the basic references to research and researchers who created and tried the tasks. The author also makes suggestions on how the tasks should be presented, most of the time "either as a yes-or-no decision or multiple-choice". The introduction of new media in basic education in this last decade certainly allows a good amount of variations over these basic forms. Similar descriptions and references can be found in [Stanovich et al., 1984] and in [Yopp, 1988].

7 - *Blending* - assembling a given sequence of isolated sounds to pronounce the word they constitute (e.g. “which word is formed by /f /i /sh?”).⁹

8 - *Deletion of a phoneme* - determining the new word that is obtained by the omission of a phoneme from the originally given word (e.g. “say *meat* without the /m”).

9 - *Specifying which phoneme has been deleted* - comparing two words to determine the difference between them. For instance: “Say *meat* and say *eat* - what is the missing sound in the second word?”.

10 - *Phoneme substitution* - generating a new word by replacing one of the phonemes of a given one (e.g. “say *meat*; now say it again but with /f instead of /m”).¹⁰

From another classification by Adams [Adams, 1990, p.67-80] it is possible to confirm the tasks above and add some slightly different phonological awareness exercises:

11 - *Addition of a phoneme* - producing a new word by inserting a phoneme at some point of a given word for instance “which word can be formed by adding a /k to *money*?” .¹¹

12 - *Syllable splitting* - saying parts of a syllable of a word (often monosyllabic words). For instance: “What is the first sound of *pink*?” or “If we take the /p out of *pink* what is left?”.

13 - *Oddity tasks* - telling out of a set of some few words which of the words is different, or does not belong to the set (e.g. “tell which of these words is different from the others: *pig, hill, pin*” or the same for *nut, get, let*).

Adams’ review concludes by ordering phonological awareness tasks by complexity:

⁹Throughout this work we use a phonemic representation according to the one used in the Festival System of the University of Edinburgh, which generates speech and phonetic patterns for the words in the tutoring system associated with this research [Black et al., 1998]. See Appendix C for detail.

¹⁰This is a tricky example for the generated word could be either *feet* or *feat*, same sounds, different spell and meaning.

¹¹This task is quite the opposite of the number 9 - deletion of phonemes. Adams actually put them together in what she calls *Phoneme Manipulation Tasks*.

- 1 - The most primitive level - that measured by knowledge of nursery rhymes - involves nothing more than an ear for the sound of words.
- 2 - At the next level, the oddity tasks requires the child methodically to compare and contrast the sounds of words for rhyme or alliteration ...
- 3 - The tasks at the third level, blending and syllable-splitting, seem to require that the child have a comfortable familiarity with the notion that words can be subdivided into those small, meaningless sounds corresponding to phonemes and, second, that she or he be comfortably familiar with the way phonemes sound when produced "in isolation" and, better yet, with the act producing them that way by oneself.
- 4 - The phonemic segmentation tasks require not only that the child have a thorough understanding that words can be completely analysed into a series of phonemes but further that she or he be able to so analyse them, completely and on demand.
- 5 - The phoneme manipulation tasks require still further that the child have sufficient proficiency with the phonemic structure of words that she or he be able to add, delete, or move any designated phoneme and regenerate a word (or non-word) from the result.

2.3.3 Experimental evidence in PA

There is plenty of evidence that phonological awareness in each of its levels is highly correlated with the abilities of reading and writing [Bradley and Bryant, 1983; Johnston and Watson, 1997; Mann, 1991; Moraes et al., 1986; Morais et al., 1979; Pennington et al., 1990; Perfetti et al., 1987, 1988; Pratt and Brady, 1988; Tunmer et al., 1988; Vellutino and Scanlon, 1987].¹² The nature of this correlation is certainly a subject of debate: Is phonological awareness a pre-requisite for learning to read or is it the consequence of the development of the ability to read? A large amount of work has been done on this subject and the general feeling now is that it may be both a cause and a consequence of learning to read: basic phonological awareness would be required for starting the process, and during the learning process other skills would be gradually achieved pushing ahead new reading abilities [Tunmer and Rohl, 1991].

However, the phonological awareness tasks described above are essentially auditory processing activities that alone are not sufficient for developing these abilities. The administration of phonologic training (the practice of making explicit links between orthography and phonology) in conjunction with phonological awareness developing tasks has been revealed to be much more efficient for this in interventional training in reading [Bradley and Bryant, 1991; Hatcher et al.,

¹²More references and terminology can be found in [Liberman and Shankweiler, 1991; Tunmer and Rohl, 1991].

1994; Snowling, 1996]. Due to the complexity of English language spelling much care must be taken in the design of exercises and choice of the words to be taught. Isabelle Liberman has earlier rejected the simple explanation that children would fail to learn to read *because* of the complexity of English spelling, for failures would be observed even when the very simple and regular codes are taught [Shankweiler, 1991]. Put another way, students would benefit from phonological awareness starting with simple stimuli, evolving for more complex ones latterly.

The literacy community has always been careful about the introduction and order of presentation of symbols that represent the sounds of the language. The attempts to reform the English orthography from 1949 to 1952 and the creation of the I.T.A. (Initial Teaching Alphabet) by the National Foundation for Educational Research and the University of London in 1960 are strong evidence of this concern. The I.T.A. is an alternative alphabet of 44 graphemes, based on the Roman alphabet extended with 14 new symbols, that was successfully tested in Britain and in the United States in the sixties [Downing, 1979, p.79-98]. The use of I.T.A. in teaching requires a further transfer of the reading skills obtained with it to the traditional orthography (T.O.), which can be carried out easily, and the benefits of the overall scheme have been proved to persist after the transferring process. However, the relative lack of teaching material in I.T.A.; environmental conflicts such as the difficulties of parents while helping children at home; concerns about transfer, and the children's lack of ability in reading T.O. material before the transferring process is completed have raised doubts about the pedagogical advantages of the scheme. Although both the orthographic reform and the I.T.A. movements are alive today, teaching T.O. is effectively the primary target of all schemes.

In the face of this complex scenery much care is taken in phonological training. The basic issue is to link the phonological awareness activities with written material. It can be done by showing the graphic representation of the sounds, simple words, and allowing the students to play with letters - solid letters, pencil and paper or any other media available. Different methods do it in different fashions. Each one advocates its own sequencing in the presentation of the phoneme-grapheme correspondence. Most of them use the official English alphabet to represent the sounds and avoid serious spelling conflicts at first, leaving the *advanced code* to be taught once the alphabetic principle is solidly understood by the student [Adams, 1990] (Chapter 11). Some methods, however, use alternate symbols and spellings to avoid the English code complexity, dealing with the

transfer of the skills acquired when appropriate.¹³ Therefore, the approaches to phonological training associated with phonological awareness tasks are not very different from the proposal of Rudolph Flesch in *Why Johnny Can't Read* [Flesch, 1955]. The main difference between Flesch's and these new approaches lies in the recognition that the task of teaching to read is not as simple and logical as Flesch initially suggested and that phonological awareness plays a very important role in this process.

2.3.4 PA and Adult Education

We have so far examined literature on the teaching of reading for children and people with learning disabilities. Let us examine now how phonological awareness has been regarded in relation to adult education. Until two decades ago not much research had been conducted on adult literacy. Apart from some countries of southern Europe, such as Spain, Portugal, Italy and Greece, most of the concerns with adult literacy in Europe and United States have been centred on meaning and comprehension, which corresponds to more advanced issues of literacy [Limage, 1990].

Although common sense seems to suggest that teaching adults to read is a similar task to teaching children to read, there are some important differences that must be observed, mainly in what relates to motivation, and social and psychological aspects. The illiterate adult must be strongly motivated to learn to read. If he or she has not learned reading up to adulthood for some reason, either disability or lack of opportunity, then there must be some concrete motivation that makes him or her dedicate energy to this endeavour. Other social aspects operate as a motivation factor, since the new ability would probably allow him or her to progress socially. It must be observed that with children these concerns are not an issue. Children would go to school because all other children go to school, without thinking too much about it. Psychologically the differences are even greater. Most illiterate adults tend to hide their lack of this skill. A distinction must be made between illiterate adults who had a chance to learn reading and failed, and adults who have never had any opportunity to learn. For the learners on the second chance the fear of failing again can make things worse. Moreover, illiterate adults may *know* many words that they have come across during their adult lives or in their jobs, although they do not quite understand how to break the writing code. This would suggest much care either in choosing the themes

¹³The Phono-Graphix method [McGuinness and McGuinness, 1998], for instance, uses "a-e" as initial symbol for the sound /ei in words like *make* or *ache*. The other extreme of this kind of approach is represented by the I.T.A. experiment.

and vocabulary to be taught or in evaluating the learning progress [Paris and Parecki, 1993; Perfetti and Marron, 1995].

In relation to basic skills of literacy, however, there is evidence that illiterate children and adults share the same kind of phonological deficits [Bell and Perfetti, 1994], and that phonological awareness and phonological training is quite effective in teaching adults to read [Moraes et al., 1986; Perfetti and Marron, 1995; Pratt and Brady, 1988]. Therefore the phonological awareness tasks and phonic training exercises mentioned earlier in this section would be suitable for teaching adults to read, provided that proper adaptations on themes, vocabulary and ways of presenting are carried out.

2.4 Generative Themes

The work of the Brazilian philosopher and educator Paulo Freire inspired a whole generation of critical educators all over the world. He suffered the horror of hunger in his childhood, and he dedicated his life to fighting the causes and consequences of human poverty and suffering, pursuing with obstinacy the development of an ethical pedagogy for social change. After a brief professional career as a lawyer, Freire started to teach Portuguese in Recife, in the Brazilian North-East, and it did not take him too long to gain international recognition for his work in literacy teaching to adults.

To better understand Freire we have to consider the environment he was in when he developed his acclaimed method. In the sixties Brazil had about 80 million people of whom around 40 percent were illiterate. Literacy was a requirement for voting. In the presidential elections of 1962, won by Janio Quadros and Joao Goulart, only about 11.6 million people voted. Freire wanted not only to teach people to read, but to make them aware of the social and political questions in Brazil and qualify them for voting. He devised a social and cultural revolution in Brazil through literacy. The National Programme of Literacy, under his direction, aimed to politicise through literacy five million people during the Goulart government. The elected president Quadros had started a daring agenda of administrative reforms in his government. After a brief period of seven months he was forced to resign under the pressure of economic and political segments of the society. Goulart, the vice-president assumed the post, and installed a popular-democratic government, but was eventually dismissed by the military coup of March 1964. The programme was extinguished fifteen days after the dismissal of Goulart. Freire was considered a dangerous person, and forced into a fifteen-

year exile. During this period he worked for a while in Bolivia, then in Chile for five years, then taught at Harvard in 1969, and worked for a decade in Geneva before returning to Brazil in 1979. His book "Pedagogy of the Oppressed" has been translated into 18 languages, and most of his work is available in English, Spanish, French and German.

Freire's work is indeed much broader than just a literacy method: it is a whole philosophy which integrates education, psychology and sociology, in theory and practice. He fought domination, aggression and violence, and overall, oppression, in all areas: race, sex, religious beliefs, political affiliation, national origin, age, and physical and intellectual handicap. He considered education as the only way to improve the human condition and defeat the psychology of oppression.

From this commitment to the cause of oppressed peoples, Freire left us many pedagogies¹⁴ all aimed at educating simple people with a view to their liberation. As a consequence, his methods would use simple material and direct dialogue with simple men and women. The *generative themes* concept emerged from this simplicity principle. Before teaching in a particular community, a team of expert teachers would try to find out which subjects and terms would be interesting to the people living there. This is a very clever idea in relation to motivation issues, since the students, besides being not afraid of unknown words, would be able to see quick progress in reading and writing words related to their life experience.

These generative themes should be around fifteen to twenty words with high potential to generate other semantically and phonologically related words. Semantically speaking, they should be rich enough to bring up economic and social issues to be explored later. For instance, in the sixties in Brasilia,¹⁵ where there was a huge population of bricklayers, the word *brick* would be the first of the generative themes. From the word *brick* the teacher would discuss related themes like work, house, home, family, always exploring the social meaning of these themes.

In phonological terms the *theme* would again be explored and unfolded in its particular features. The word *brick* in Portuguese is *tijolo*, with very well defined syllables: ti-jo-lo. The "t-family" would be examined (ta-te-ti-to-tu), and the vowels would be extracted and stressed. Then the other *families* ("j-family" and "l-family") would be examined also. The various syllables would be put on to cards and combined to generate different words from the language. Some of the

¹⁴ *Pedagogy of the Oppressed, Pedagogy of Hope, Pedagogy of the City, A Pedagogy for Liberation* are some titles of Freire's books.

¹⁵ Brasilia is the capital of the country, which was totally planned for this end, and was built in the virtually uninhabited heart of the country in the late fifties and early sixties. Around Brasilia many "satellite" towns have been created to accommodate the huge migrant population who came to work on the building of the town.

combinations would produce nonsense words, but many others would represent valid words, from the student's known universe. It must be noted that Portuguese is much more regular than English, although not always regular, and that syllabic approaches are quite successfully used in Portuguese literacy.¹⁶

The semantic and phonological unfolding of the word would be integrated, and many words related to that generative theme would come up as the student progressed. The next generative themes would be studied, one by one. By the end of this study the students would be able to read and write most of the words related to the surrounding objects and subjects, with a critical vision about them. By this time, Freire would say, they have become literate. The next methodological step would be the post-literacy phase, when the acquired skills would be consolidated and extended.

For example, in recent years in Brasilia's satellite town named Ceilandia, at the "Educational Centre Paulo Freire", these are the generative themes: *lote* (plot), *comida* (food), *jogo* (game), *secura* (dryness), *lixo* (rubbish), *chuva* (rain), *eleição* (election), *barraco* (shack), *religião* (religion), *trabalho* (work), *televisão* (television), *água* (water), *associação* (society, union), *passagem* (ticket for urban buses) and *filho* (son). These words represent the concerns and problems of those very poor people: their basic needs, as food, plot, shack (a place to live), the basic community needs as water, collection of rubbish, the basic social organisation (election, religion), their daily conversations, about weather (dryness, rain), their amusement activities (game, television) and so on.

What is the empirical basis for the success of Freire's work? The literature about Freire does not give much attention to this particular issue. In a biography by his widow [Freire, 1997] she reports that in one experiment, after 21 hours of literacy work, the students would be able to read simple newspaper notes and write short sentences. After 30 hours, one hour a day, five days in a week, three of the five participants were able to read newspapers and write letters. It seems that Freire's laboratory was real life and real people. His statistics come from the National Literacy Programme in Goulart's government, then from the results of the application of his method and ideas in Cabo Verde, Guine Bissau, Chile, India, Nicaragua, and in many other places in Brazil and all over the world.

What makes Freire's work so special to deserve so much attention and recognition from intellectuals all over the world?¹⁷ The first reason might be that he

¹⁶The use of phonetic approaches for teaching reading, has not been significantly considered in the teaching of literacy in Portuguese, as far the author is aware. The impact to the use of these approaches in Portuguese seems to be an interesting follow-up of the present research.

¹⁷Paulo Freire received more than forty awards and titles in recognition for his work: Award

launched a new focus on the subject highlighting the link between literacy and power - "the learners are empowered by the knowledge that they are learners" [Freire and Macedo, 1987] (Foreword by Ann E. Berthoff, page xviii). For him, the meaning of literacy is much broader than the simple mechanical reading of the "back to basic" movements regarded as "unacceptable historical unreality" [Freire and Macedo, 1987] (Foreword to the edition by Margaret Meek, pages xii to x). Literacy for Freire is an instrument of awareness and transformation. In the "Pedagogy of the Oppressed" the frequently used Portuguese word *conscientização*, which in a free translation means "consciousness raising" is never translated into English - what really matters in Freire's method is this transformation process: the *conscientização*. Since Freire, these questions (surprisingly obscure up to then) have been irreversibly exposed, and literacy can no longer be referred to just in technical terms.

He also focused on adult literacy without the usual paternalism of the superior literate classes when conceding the dispossessed the right of learning to read. What he proposed was an equal relationship, a process in which everyone, teachers and learner would be working together to solve a common problem. The teacher would have to learn and adapt their skills to the knowledge of the learner. It was a new vision and a new challenge for educators [Freire, 1970]. He considered many aspects of adult education in his method, with his natural sensitivity for what really matters in adult education. He knew, from his live experience, how those illiterate adults would react in face of the challenges of learning. He knew their fears, their embarrassment, their difficulties, and more than that, he knew their real reasons for that - the feeling of inferiority, the feeling of being oppressed. Besides that, he had a superb talent for designing and testing methods of teaching.

William Rainey Haper, from The Religious Education Association of the US and Canada; Brotherhood Recognition from the City of Los Angeles, US; Commend of the National Order of the Educational Merit, from the Education and Culture Ministry of Brazil; Award Merit of Peace, from the Association of Research and Specialisation in Latin-American Themes, Spain; Diploma of International Merit from the International Reading Association, Stockholm, Sweden; Medal Jam Amos Comenius, from the government of the Czechoslovakia Republic; The Paulo Freire Awards, from the International Consortium of Experimental Learning; Award Mohammad Reza Pahlevi, from UNESCO; Award King Balduino for the Development, from the Belgian government; Award UNESCO for Education for the Peace; Award Andres Bello, from the Organisation of American States (OAS), as the Educator of the Continent. Paulo Freire has received the title of "Doutor Honoris Causa" from the following institutions: Open University of London; University Catholic of Louvain, Belgium; University of Michigan, US, University of Geneve, Switzerland; New Hampshire College, US; University of San Simon, Bolivia; University of Santa Maria, Brazil; University of Barcelona, Spain; and a list of recognitions from twenty other educational institutions. Freire died of a heart attack in May of 1997, at the age of 76, when he was about to receive another title: "Doutor Honoris Causa" of the University of Habana, Cuba.

The above analysis of Freire's work would place it much closer to the whole word approaches than to the phonological approaches. He was concerned with context and with the psychological atmosphere of the teaching, as whole word supporters normally are, and these concerns were given very high priority in his method. Although highly thematically driven, Freire's method is also phonologically oriented. It has been used to teach in countries with Portuguese and Spanish language, but has not been given emphasis on phonemic issues, as these languages apparently do not require it as much as English does. Nevertheless, it examines and stresses the sounds and their associated representations, centred in syllables, which has been proven to work for these languages.

Section 2.3 gave an idea about how controlled the vocabulary must be to match the curriculum items in PA approaches. Finding out words to fulfill all the curriculum needs is not a simple task, which sometimes results in a dull and stilted vocabulary. Now we get to the central question in this work: how compatible is Freire's method with phonological awareness approaches? Put another way, is it possible to determine a vocabulary that will satisfy both the thematic and phonological constraints of the two approaches together? Can we combine the virtues of PA with the social thematic approach of Paulo Freire to produce a valuable method for teaching adults to start getting basic reading skills? How can it be done with the help of computer systems and artificial intelligence?

2.5 Phonological Awareness versus Thematic Approaches

A neutral observer would recognise the valuable arguments and virtues of both approaches just described and discussed. Both require very sophisticated training to be applied. In both cases, designers of the methods (the expert teachers or pedagogical experts) should be aware of the psychological, sociological and cognitive factors involved and the teachers should be very well prepared to apply the methods. In this section, rather than joining the debate and making an option for either side, we want firstly to consider the advantages and difficulties of applying principles and features of both approaches in combination, and then to discuss what the requirements are for such a combined method.

The very first question to be asked is: how do these approaches relate to each other? More specifically this question could be broken into several smaller ones: does either one of them encompass the other? Are they incompatible? Are they complementary? Of course we do not have simple answers to these questions.

The focus and the methodology of these approaches are quite different. While PA approaches principles are extremely technical and well grounded experimentally, Freire's method is spontaneous, emotional, and based in everyday life experience. Freire's roots are universal truth extracted from humanistic philosophers such as Marx, Engels, Lenin, Kant, Freud, Jung, Erich Fromm, Sartre, Simone de Beauvoir, André Malraux and many others, extensively cited in his books.

As a consequence, Freire's method is humanistic, emotional and warm, as opposed to the rather technical, rational and cold reports on experiments with the application of onset and rimes, syllables or phoneme sequences with young learners. However, if we do not let ourselves be influenced by either arguments, we will be able to see that the former can humanise the latter, whereas the latter can be used to better inform the former. They essentially focus on different aspects of the same question. Freire's method is also concerned with the phonological issues of the words, beyond the thematic focus. It has been highlighted that the phonological features of Portuguese or Spanish are simpler than the English ones, but the concerns are present in Freire's method. What he proposes with the generative themes unfolding is precisely the investigation of the basic phonological features of the language and their representation on printed media.

A question of priority may arise at this point. The PA approaches are normally based on a sequence of presentation of the words to the students, so that interesting features would turn up at the right time in the learning process. Would we be able, in a combined approach, to present words in a thematically interesting sequence and still attain the phonological requirements of PA? We have done some experiments on this particular issue and the analysis indicates that this is possible (See chapters 4 and 5).

Another point that must be discussed is the amount and the nature of work to be carried out automatically by the machine and the work left to human instructors. Although in Freire's method audio-visual devices have been welcome,¹⁸ there are some procedures and tasks that Freire believed should never be done automatically. Some of these tasks were evidently related to the definition of the generative themes and to their unfolding in thematic and phonological branches. Besides being indeed a very complex task, the themes discovery process itself is part of the methodology, and presupposes the participation of representatives of the community. Consequently, to comply with Freire's method, the computer should give support to the discussion, but not do all the work itself.

¹⁸Freire frequently refers to slide projectors, tape recorders and posters, to be used with conventional chalk and blackboard, pencil and paper.

Another aspect to be carefully regarded is the interaction with the student. Freire used to consider the most important thing in teaching to be the relationship and contact with the student. Would he accept a computer as a surrogate teacher in his method? This last question is however independent of method: no literacy method so far has been designed for being administered by computers, and most of them count on a great deal of co-operation, involving teachers, colleagues and parents. In the short term, it is unlikely that a computer tutor alone would be able to supply all the intelligence and skills necessary to teach someone to read from beginning to end. In this same way, PA methods require a vivid and precise interface with the student, for the sounding of words in their component phonemes is crucial for their success. Consequently, for both approaches much attention must be paid to the user interface.

Computers and Adult Education

Probably the first significant commercial program to teach basic skills to adults was *Planit* (Personalized Logic for Adult Teaching Operations), developed by Control Data Corporation, in cooperation with the University of Illinois and the National Science Foundation, in 1976. It was the main experimental project carried out in research on the application of this kind of technology for the 1970s. Like most of the early CAI systems, *Planit* was designed to be used in a laboratory or workshop setting by the individual student, rather than by the teacher. The total cost of *Planit* was about 70,000 dollars, of which 30,000 dollars, chosen by the user, would cover the cost of the production of the program (Turner, 1983).

2.6 Related Work

Computers and Education

The application of computers in education, and particularly in literacy, is not a new endeavour, and tends to increase, as it is increasing in virtually all human activity. In the military environment in the United States, software for the development of basic skills has been developed since the beginning of the seventies, although unavailable for the general public. Since then the term CAI (Computer Assisted Instruction) has been coined and largely used up to the present. Other common designations for instructional computerised tools are ICAI (Intelligent Computer Assisted Instruction), ITS (Intelligent Tutoring System) and ILE (Interactive Learning Environment), the first two used alternatively to claim that Artificial Intelligence (AI) techniques are in use, and the latter to characterise that more than being simply a tool, the product encompasses a whole interactive intelligent environment for teaching and learning.

Conventional CAI tools have undeniable advantages over the usual classroom routines for teaching drill and practice material - the students can get instant feedback on their answers and the computer can be programmed to be infinitely patient, giving no evidence of frustration or boredom. However, for more complex tasks, their ability to respond to learners' needs is limited. The introduction of AI techniques for reasoning simulation and expertise codification empowers the tutoring systems to execute some task quite independently of human help, working as companions, assistants, tutors or learners [Eurich, 1990; Turner, 1993; Wenger, 1987].

Computers and Adult Education

Probably the first significant commercial product to teach basic skills to adults was Plato (Programmed Logic for Automated Teaching Operations), developed by Control Data Corporation, in cooperation with the University of Illinois and the National Science Foundation, in 1975. Its use in many experimental projects contributed to research on the application of this kind of technology for teaching adults. Like most of the early CAI systems, Plato was founded in fixed content to be learned in variable time for the students to satisfactorily perform on it. By 1979 the yearly leasing cost of Plato was about 30,000 dollars, almost as expensive as hiring a human teacher. By that time doubts were raised on whether CAI systems would ever be viable [Turner, 1993].

With the explosion of the production and sales of micro-computers in the early

eighties, things changed significantly. Besides being affordable, flexible, portable, connection independent,¹⁹ and architecturally open, the micro-computers were starting to prove reliable enough to carry out some serious tasks.

Computers and Literacy

In the United States the official support for the development of technology for adult education using micro-computers dates from 1981, provided by the National Advisory Council on Adult Education, committed to spending at least 10 percent of their funds for innovative efforts in the area. In 1984 17 US states sponsored computer-related projects with these funds. In 1985 the Gannett Foundation sponsored an invitation for 50 literacy professionals interested in the use of computers in literacy. The result of the meeting was the establishment of the Adult Literacy and Technology Project, at the Institute for the Study of Adult Literacy at Pennsylvania State University. In 1991 the US Department of Education funded a major research centre for adult literacy, the National Center on Adult Literacy (NCAL) which is responsible for conducting research and producing policy-oriented papers to assist decision-makers at the state and federal level [Turner, 1993].

In Britain, efforts have been made on literacy teaching with microcomputers since the beginning of the eighties. In a position book published by the Council for Educational Technology in association with the National Association for the Teaching of English, on exploring English with microcomputers [Chandler, 1983, p.7], the Publications Officer of the National Association for the teaching of English, Mike Torbe, says:

“In 1981 a local branch of the National Association for the Teaching of English organised an evening meeting for its members on Microtechnology and its implications for English teaching... 18 months later, a concerted demand came from heads of English departments, and from primary teachers, for a full conference to guide them in the way they spent money on - and used - the computers becoming available.”

This same book discusses many different ways of using microcomputers in the teaching of literacy, explains why English teachers should use computers, teaches how to use a database, and describes “Poem”, “Walter”, “Gram 3” and “Fantasy” as part of a suite of programs developed by [Sharples, 1983] at the

¹⁹For the high cost, complexity of operation and low speed of on-line equipments, being independent of connection was a great advantage two decades ago.

Artificial Intelligence Department of the University of Edinburgh, for supporting creative writing, working on generative grammar and sentence combination. It is quite interesting to feel in [Coupland, 1983] the atmosphere of excitement that popular programs like STORY, ANIMAL or ADVENTURE²⁰ would bring to language researchers at that time.

The value of using computers for disabled literacy learners has also been investigated in academia. At the Australian National University, [Macleod and Lally, 1981] used an electronic pen to improve the handwriting skills of handicapped children. In the University of Edinburgh, [Howe, 1981] developed a computer-based system for teaching word attack to handicapped children through a model involving sounds and images controlled by a computer system, addressing the phoneme-grapheme relationship. At the same university, [Pain, 1981, 1985] developed a computer tool for helping children with learning difficulties in spelling. At the University of Bristol, [Lees and Chapman, 1986] devised and implemented schemes for the teaching of reading to the hearing-impaired. At the School of Education of the University of Newcastle-upon-Tyne, [Moseley, 1991] investigated the use of talking computers to improve the learning of reading by children and pupils with specific learning difficulties.

In France, at the University of Clermont-II, [Bruneau et al., 1991] proposed an interesting architecture for an intelligent computer-assisted learning environment for helping the teaching of very basic skills of reading, aimed at non-impaired children, at the SIAAL Project. The architecture proposed reflects the concerns to represent the knowledge about the subject domain, the pedagogical knowledge, the didactic situations and the student modelling in a similar way to this research. It was designed to teach French, and was focused on syllables as the main phonological awareness units.

These studies happened mostly at a time when computers did not have all these multimedia devices they have today. For many of those works, beyond the cognitive and pedagogical issues, the connection and interface of the computer system with the devices would be included as part of the research.

²⁰STORY was a program that would help the student to build a story starting with a list of adjectives, adverbs, nouns, names and locations. ANIMAL was a program that would allow the computer to guess what animal were you thinking about by asking questions like “is it a mammal?”, “can it fly?”, and so on. ADVENTURE was a more sophisticated program in which different scenerios and stories would be assembled by the computer as response to simple questions posed to the user.

Commercial Products

Now, two decades since their advent, micro-computers have become much more affordable and acquired capabilities unpredictable at that time. The human-machine interfaces have evolved amazingly, supporting speech synthesis and recognition to fairly acceptable levels, along with high quality graphic images on screen. The application of optical technology to the storage of data on CD-ROM and recordable CDs has contributed significantly to the distribution of software in off-line mode. Another breakthrough, the Internet, is now available in one third of British houses and in many schools and offices, with high speed links and a variety of open communication protocols able to support not only conventional data exchange, but voice, music and high quality image and television, in interactive mode.

A great number of products to help the teaching of literacy to children are available on the shelves these days for prices ranging from 10 to some hundred pounds sterling. Although many of these can be very helpful, it is really a difficult task to understand what the offers are and make a choice. It is a new and growing market, open to a public not always aware of the terminology and needs of the literacy area, and there is a high risk that many of the products on sale never get installed and tried. There are so many of them, that, as previewed in [Hawkrige, 1990], not all would ever get reviewed properly. In this way, the user would have to trust in simple descriptions of the products²¹ and try hit-and-miss approaches. The Educational Software Yearbook 2000 of the R-E-M²² comes with 206 different titles on literacy, each one with a short description of its capabilities, ranging from letter naming, development of phonic skills and spelling, grammar, writing, up to user friendly dictionaries. The magazine has invited actual teachers to evaluate only a couple of products.

The British Dyslexia Association has published a helpful booklet on software for literacy that they have analysed and recommended at different stages of the learning process [Hutchins, 1998]. These are mostly commercial software designed for teaching children, which can be helpful for teaching dyslexic people as well. Most of these products are focused on particular issues of literacy teaching, such as letter naming, spelling, sentence formation, dictation, story listening, and so on. Only one of the software publishers examined in the booklet appeared to cover

²¹Hawkrige defines reviews of educational software progressively as *description*, *analysis*, *critique* or *evaluation* depending on the effort spent on the assessment of the product.

²²The Rickitt Educational Media LTD, based in Somerset, specialises in selling educational material to teachers and schools through mail order and through the Internet, and publishes annually a catalogue of educational software for all areas.

a wide range of aspects in literacy with many different middle priced (between 30 and 60 sterling pounds) products, although it is not clear whether all of them belong to the same literacy method or approach.

It is clearly a market with strong commercial appeal, probably with many good products which can be helpful for the teaching of reading at different stages. Even though we have read some detailed reviews and tried ourselves some of the products available it is difficult to establish a common profile for them. A two-hour trial assessed three different full products: *Starting to Read*, produced by Europress, *Sesame Street Elmo's Reading*, from the Learning Company, and *Reader Rabbit's Reading Ages 6-8*, also produced by the Learning Company. They are all middle priced products (around 20 sterling pounds each), available in software shops all over Britain, developed after 1998, distributed through CD-ROM (all single CDs except *Elmo's Reading* with two CDs) for both PC and Macintosh computers. Technically the three products tested are well constructed and presented: enticing cartoon graphics, good animations, clear pre-recorded dialogues. The resultant environment would rather be described as an interactive cartoon. The mouse and the keyboard are the preferred input devices. *Reader Rabbit's Reading* uses the microphone for recording speech from the student, although it does not make any analysis of the recorded track. The focus of the products is limited to some particular features of the language, so there is not much room for guidance, that is, the exercises seem to be presented in a sequence judged as convenient by the designers. There seems to be little or no effort to use student models or strategic rules for teaching in these products. There are no standards in terms of fonts used for the letters or in the representation of the phonemes. In *Starting to Read*, for instance, the designers decided that the uppercase letters would represent the names of the letters, whereas lowercase letters would represent the sounds of these same letters. The sounds of the consonants in *Starting to Read* are presented with the help of the vowels /a/ or /u/, so the phoneme /g/, for instance, would sound something between the /ga/ or /gu/. Although the ways of presentation are different in each product, the conception of the exercises is quite similar: the first letter and the first phonemes are worked out, then some rhyme comes into focus, and stories are told, highlighting the words while they are being spoken. The exercises with sentences emphasise the suppression of words to be latterly indicated by the students with the mouse. None of the products analysed seems to be consistent with any particular literacy approach. As with many other areas, it is very difficult to judge the benefits of these products - there is no record of evaluation or assessment. Apparently this

technology advances according to market rules, at such a speed that there is no room or time for evaluations to take place. It seems that at least some of the products are beneficial for young learners, but how beneficial they are and how different products compare to each other is something that would demand field evaluation.

As far as the author is aware there is no commercial product or other ongoing research dealing with the combined approach proposed in this work, offering the building blocks for an expert teacher to generate an ITS for helping to teach literacy to adults on this basis.

Chapter 3

Requirements for a literacy teaching platform

As a consequence of all the considerations in Chapter 2, we understand that the teaching of literacy is a complex task with many different activities that acting together are able to lead the student to a condition where he or she is able to read. Phonological awareness is today universally accepted as a requirement for good reading, and its exercise and development is the basis of most clinical and remedial reading schemes. Most reading methods use phonological awareness as standard practice among their activities at some point, and there are some methods that are almost exclusively based on phonological awareness procedures. Once the student is able to deal with the code of the writing system, other reading practices take over the teaching, and phonological awareness can continue to be developed in a less explicit way.

In order to inform the answer to the question addressed in this research, whether it is possible to support the teaching of reading for adults combining the principles of PA approaches with the Generative Themes approach, using computer systems and AI, we have proposed a literacy teaching platform attending to some basic requirements. This teaching platform is a set of computer tools for fostering phonological awareness in adults, to be thematically adjusted and applied to different communities, to be used in the first stages of the reading process. Its focus is mainly on words, considering phonological features, as PA approaches do, and meaning, as recommended by Paulo Freire in his method. The teaching platform is intended to act as a complementary resource to help students improve their phonological awareness, according to their own learning rhythm, and without the direct supervision of a teacher. Consequently, the computer should record and use the student's learning history in each interaction to provide the individual attention required.

The platform proposed should conform to the requirements described below:

- It should be based on successful methods for phonological awareness development: after so much energy has been applied in research and experience in literacy it does not make sense to experiment with unprincipled technology.
- As the aim of the platform is to teach illiterate people, high quality speech generation capability is a strong requirement. Speech recognition is desirable as well and its adoption is inevitable in the near future. Other devices, such as the mouse, the touch screen, and the keyboard, and writing-pads should be used by the student for interacting.¹
- Image processing is a fundamental facility for ITSs in general, and for this kind of ITS in particular. Images of objects, photos of people, and appropriate icons would play an important role in the teaching process.
- Animation is another important requirement, mainly when combined with virtual personages in order to establish the dialogue with the student and help the student in the recognition of the speech fragments with facial expressions. The computers hosting the ITS must be able to process image and sound (speech) with reasonable response time. This should not represent a problem since most of the PCs on sale today exceed the basic requirements in terms of speed and memory.
- Networking is another desirable facility, for it allows the automatic update of software and the central administration of the system, relieving problems with local training and software installation. In this research we are using Java as the programming language to attain to the principle that all the system software would be distributed through the Internet. The use of Java Applets will make the ITS runnable on computers with no special software, but conventional Internet browsers.
- In order to get some operational independence from the teacher it is fundamental that the ITS would be able to recognise the students individually, to keep track of their progress in learning and to have a way to decide which steps should be taken after one activity is finished. In other words, the computer must deal with the *student model*, and cope with the *strategy rules* while interacting with the student. For this purpose, the system must

¹As will become clear in the description of the implemented prototype for the platform, devices for speech and hand-writing recognition and touch screen have not been dealt with yet. However, the models and the tools are open for these future developments.

be assisted by specific Artificial Intelligence capabilities, involving knowledge about the curriculum, knowledge about the subject domain data (word meanings and phonological features) and knowledge about the pedagogies in use.

As it is clear from the review in Chapter 2 and discussion that there is no simple solution to the problem of providing an ITS for helping teaching literacy to adults using principles from PA approaches and Generative Themes, we proposed and implemented in this research a set of tools that act as an ITS generator. In order to describe the facilities and features of the ITS generator some of its technical terms are introduced here. In this thesis, the word *activity* refers to the teaching activity, which can be either a *lesson*, an *example*, an *exercise* or a *test*; the *activities* are declared as sequences of *steps*, which can be specified with twelve different types of *objects*; the *expert teacher* is an expert, or a group of experts, in literacy, who provides the details of the literacy programme in use, and who will define the teaching activities and the strategy rules to be used by the computer system. The complete terminology and details on architecture and components of the system are provided in Chapter 7.

These are the facilities and features of the ITS generator:

- Different ITSs can be generated by the system, each one corresponding to different literacy programmes, using an authoring tool and other utilities provided as tools.
- The curriculum of each ITS is variable and independent. It should be based on a successful phonological awareness scheme, and can be expressed through a *Knowledge Tree*, as described in Chapter 5.
- The system has utilities for the selection of words according to thematic and phonological criteria, generating specific vocabularies to be taught to each ITS generated. Chapter 4 describes the principles and implementation of this facility.
- The authoring tool allows the *expert teacher* to define *activities* to be taken by the students while learning. These activities are described on a step by step fashion, using *objects* as the activity building blocks, which render the following services:
 - Start and finish the *activity*.

- Input data from the student, from any of the available formats and devices.
- Output data to the student through any of the available formats and devices.
- Assemble and show the student pictures with people, scenery and dialogue.
- Ask a yes-or-no question to the student and allow her to answer it.
- Pose a question for which the answer is a representation of a sound (phoneme or sequence of phonemes), letter or word, and present an array of icons representing the alternatives for the student to choose from.
- Let the program evaluate a question and take an action depending on the result.
- Let the program branch unconditionally to a particular point in the *activity*.
- Let the program define working variables and attribute values to them.
- Let the program define sets of elements attending some particular properties.
- Present the student with two lists of icons representing sounds and letters for her to match them.

The aim of these *objects* is to allow the *expert teacher* to compose different *activities* with tasks similar to the ones described in Section 2.3 and afterwards relate them to the educational programme in use. It might be necessary to include new kinds of *objects* to make the tool more generic and more able to deal with different *activities* of particular methods. The *objects* developed so far are detailed in Chapter 7, Section 7.3.

- The authoring tool must allow an expert teacher to express a set of teaching strategies as rules to be interpreted by the system every time it has to define the next *activity* to be run. The scheme is detailed in Section 7.4.

These tools and facilities have been prototyped in computers using the Java language, and are able to process most of the functionalities mentioned in this chapter and detailed in the next ones. In order to evaluate the validity of the proposed concepts and tools, and to confirm the assumption that it would be possible

to apply PA schemes using thematically selected vocabularies, we selected three different successful PA methods to represent with the system: Phono-Graphix, developed by Read America Inc [McGuinness and McGuinness, 1998]; THRASS (Teaching Handwriting Reading and Spelling Skills), developed by THRASS UK Ltd [Davies and Ritchie, 1996]; and Jolly Phonics, developed by Jolly Learning Ltd [Lloyd, 1998]. These are PA methods that organise the curriculum and the teaching activities in different ways, covering most of the traditional activities recommended in the literature to improve phonological awareness. Besides that they have been developed in diverse English speaking countries (US, Australia and UK), and are being extensively used in these three countries.

Consequently, we believe that the implementation of parts of these programmes constitutes a significant test for the concepts and tools developed in this research. The selected methods and their instantiations with the tools developed in this research are described in Section 5.3.

Chapter 4

Vocabulary Selection

The selection of the vocabulary to be taught is crucial to any literacy scheme. Which vocabulary should be used in an approach that combines Generative Themes and PA? Would it be satisfactory simply to adopt vocabularies already defined in the original approaches or available in other literacy methods?

Chapter 2 has already stressed the importance of the vocabulary selection for adult education. Regardless of the approach in use, it is sensible to choose the words carefully so they do not add complexity to the learning process by provoking in the adult learner undesirable sentiments of low self-esteem. In Freire's approach, specifically, the vocabulary to be taught is highly variable, chosen accordingly to the local values and interests of the target population in that particular moment. As a consequence, the method itself requires tools and procedures to allow the dynamic selection of vocabulary to be taught.

Phonological Awareness approaches had been used mainly for remedial interventions aimed mostly at children, and only recently have those been considered as an option for regular teaching and adult education. The priority in the choice of PA vocabularies is the connection with the curriculum, in order to explore the phonologic features of the language in an appropriate sequence. The vocabulary themes in existing PA methods appear to be both designed for children and not focused on any specific area or context.

These considerations indicate the necessity for determining specific vocabularies for the combined approach. Along with this conclusion, however, there comes the major question addressed in this research:

Is it possible to obtain thematic vocabularies with enough words to cover all the desirable phonologic features to be explored in PA approaches?

In order to answer this question, the possible sources of words have been investigated and two criteria for selecting the words to be taught have been considered: the *thematic selection* and the *phonological selection*. The *thematic*

selection intends to make it more probable that the words to be taught belong to the student universe and will serve the purpose of motivating debates, consolidating ideas, bringing immediately to the student the feeling of usefulness in the learning activity. The *phonological selection* aims to choose words that cover all the relevant phonological features of the language and of its representation in print, in a balanced way, according to well-known PA literacy schemes.

As part of this thesis, these criteria have been implemented in computer tools and have been tried in relation to different PA literacy schemes. The vocabularies obtained have been compared to the vocabularies used originally in the methods.

In this chapter we describe the source of words, the concepts and the tools developed for selecting the words according to themes and contexts. An early application of these methods for selecting words for three different themes has been described in [Carvalho et al., 1999].

4.1 The source of the words

The first concern while defining the selection scheme for the vocabulary to be taught as required in this thesis was to find an appropriate source of words in machine readable format. In order to select enough theme related words to encompass the different phonological features of the language, it is necessary to have a word data base that is not only abundant but also rich in meaning. Three traditional types of language resources have been considered for this purpose: corpora, thesauri and dictionaries.

Corpora

Written language corpora are collections of text in electronic form, destined to support research and applications in natural language processing. They have been used in areas such as the development of spelling and grammar correction routines, the automatic lexical acquisition for dictionaries, the building and consistency validation of glossaries and specific terminologies, the automatic translation of text and speech, the study of word use and association (collocation), the language usage (typical, poetic, erroneous use) and many others. Their presentations range from very raw material (pure text) to finished components, with tags and other information added to the text. Currently there are a number of corpora available for public consultation in various sites and in many languages. The British National Corpus (BNC) with its hundred million words database is perhaps the most significant representative of the public corpora in use today.

Likewise, efforts have been made to standardise the storage and use of these resources. The Text Encoding Initiative [Ide and Sperberg-McQueen, 1995] is an example of such efforts [Ejerhed and Church, 1995].

The direct processing of a corpus may constitute a huge computational task involving the efficient parsing of the text and the solution of a significant number of linguistic issues. Although specialised corpora may be rich in meaningful words for a particular theme, they are not commonly marked for meaning [Ejerhed and Church, 1995, p.448], so their words can not be automatically recognised as relevant for that theme. There is a significant amount of work already done on the automatic extraction of semantic information from corpora to improve the information in dictionaries and thesauri. From this point of view, it seems much more reasonable to look at the condensed information already existing in these last two kinds of resources, dictionaries and thesauri, instead of doing all the work from scratch.

Thesauri

Thesauri are organised data structures, where the words are disposed in networks according to their semantic and grammatical values. Two main resources have been considered as possible sources of words for this project: Roget's Thesaurus [Roget, 1994; Sparck Jones, 1991] and Wordnet [Miller et al., 1993].

The Roget's Thesaurus, first published in 1852 by the retired physician Dr. Peter Mark Roget, is probably the most remarkable reference in the area. However, its "Tabular Synopsis of Categories", the structure around which the words network is organised, has changed little since the first edition, even in the computerised version, and many current very common words and expressions are not covered in it. Roget's inspiration when building his thesaurus was the offering of a tool to aid thinking and writing, but not a complete linked dictionary. This seems to be still the inspiration of thesauri researchers and builders in the present. [Townley and Gee, 1980, p.21-22] highlights the "five principles of thesaurus construction which should be engraved on the heart of any thesaurus compiler":

1. No very rare concepts should be included in the thesaurus since they could not be expected to produce many matches between documents and search requests.
2. Very common high-frequency terms should also be excluded from the thesaurus since they produce too many matches for effective retrieval ...
3. Non-significant words should be studied carefully before they are included in the list of words to be eliminated.

4. Ambiguous terms should be included only for the senses that are likely to be present in the document collections to be treated...
5. Each concept class should include only terms of roughly equivalent frequency so that the matching characteristics are approximately the same for each term within a category.

As becomes clear from the rules above, vocabulary completeness is not the main concern of thesauri compilers. In practical terms, for the application related to this thesis, the thesauri lack the desirable completeness in different domains. Many of the simple words needed for teaching beginner readers seem to be either absent or not sufficiently marked in the observed thesauri.

Wordnet is a more recent project that has been developed since 1985, conceived to unite computer technology with modern psycholinguistic principles and to cover a bigger range of word categories. The version of Wordnet examined has around 95,600 entries, including simple words and collocation, converging to around 70,100 word meanings, which is much more than most pocket dictionaries available in the book-shops. Even so, Wordnet still does not satisfy the needs of the present research. In Wordnet, for instance, there were plenty of words related to the theme *food* (around 2000), which was one of the themes targeted in the experiments described later in this chapter. However there are too few words related to the other two themes also focused on in these studies (*sea* and *agriculture*).

Dictionaries

These considerations led to the examination of the third option: machine readable dictionaries. This option turned out to be the one to be chosen as the primary source of words for this work. The main advantage of using dictionaries is that they virtually cover all kind of subjects and contain most of the words of a language in a relatively small volume. Three different machine readable dictionaries were used to assemble a prototype dictionary to be used as the source of words in this work:

- Collins English Dictionary - a formal and concise dictionary
- Collins Cobuild II¹ - a dictionary with vocabulary and examples of use of the words in a form suitable for beginners and learners for whom English is a second language

¹Thanks to Cobuild Ltd for allowing the use of both Collins English Dictionary and Collins Cobuild II Dictionary in this project.

- MRC II (Medical Research Council) Psycho-Linguistic Database - vocabulary with complementary information on word usage, category and age of acquisition

The Festival System from the Centre for Speech Technology Research of the University of Edinburgh [Black et al., 1998] is also used to provide the phonological coding for the words in the prototype dictionary. The format of the prototype dictionary accompanied by an example is shown in Appendix A.

The resulting prototype dictionary (from now on referred to simply as the “dictionary”) has about 60,000 entries and contains most of the information needed for thematic and phonological selection as needed in this research. In order to obtain a more manageable database and reduce further the discarding of words after the selection process, the dictionary was divided into three subsets: the high frequency words dictionary, the low frequency words dictionary and the regular frequency words dictionary.

As one of the principles of the word selection process in this research we assumed that words that are very frequently used in the language are words that should be considered early in the teaching process. The reason for that is because if they are frequent words it is likely that the student would come across these words and need to deal with them early in the literacy process. It might be the case that some frequent words are not the most appropriate ones to start with, in which case the focus on these words should be postponed. However, this principle was adopted as a general orientation. Conversely, words with low frequency of use were given low or no priority to be used as stimuli for teaching.

Consequently, regardless of the theme of the target vocabulary (the set of words to be used as stimulus in the teaching), the frequency of occurrence of the words played a crucial role on the selection process. Among these words there would certainly be functional words such as the articles, the pronouns, the conjunctions and some of the most common adverbs, which are fundamental for the construction of even basic sentences.

The general selection policy adopted is:

- All selections apply to the regular frequency words dictionary.
- The low frequency words dictionary is ignored unless it becomes impossible to obtain a reasonable amount of words for a particular theme.
- The high frequency words, after being examined in relation to pedagogical convenience from the point of view of literacy, can be added to the target vocabulary at the end of the selection process.

The data on word frequency were obtained by counting the occurrences of words in the text of the dictionary itself and in a number of public corpora on general subjects, books and public broadcast.² The words that occurred more than 600 times in these documents were considered as high frequency words, whereas the words with fewer than 6 occurrences were classified as low frequency words. After the frequency counting the dictionary was divided into three parts: the low frequency dictionary with 40,625 words, the regular frequency dictionary, with 19,748 words, and the high frequency dictionary, with 841 words. This configuration was established after a number of trials, considering the volume of information to be dealt with in the selections and the results obtained experimentally. It is not claimed that this is the optimum configuration for the dictionary. However it was considered appropriate for testing the model and methods proposed in this thesis.

4.2 The Thematic Selection

4.2.1 The seed set concept

In Lexical Acquisition research, words that are likely to co-occur in specialised documents are said to belong to the same *semantic domain* [Manning and Schutze, 1999, p.296]. The thematic selection method here proposed is based on the assumption that the words related to a theme would tend to be described in their dictionary entries with the help of a relatively small set of words which act as *keywords* for that domain. This set of words was named the *seed set*. From this point of view, each entry in a dictionary would correspond to a very specialised document with strict reference to the word it describes.

As an illustration of this concept, consider the search for words to be taught for workers in the *food* industry (e.g. cooks, waiters and so on). A possible seed set for *food* could contain words like *food* itself, *cereal*, *cheese*, *cook*, *edible*, *fish*, *kitchen*, *meat*, *milk*, *vegetables*, amongst others. Figure 4.1 illustrates the idea with the entry correspondent to the word *haddock*.

In the example of Figure 4.1, the words *edible* and *fish*, which belong to

²The complete list of the documents used for counting word frequency is: 1 - The Thesis Proposal for this research; 2 - The text of the dictionary of the research itself; 3 - The Voice of America - transcription of various broadcast sessions in a total of around 37,278 words; 4 - Alice Adams, novel by Booth Tankington (this book and the ones that follow it were available in a public domain site in the Internet); 5 - The Ambassadors, by Henry James; 6 - American Notes, by Rudyard Kipling; 7 - Anna Karenina, by Leon Tolstoy; 8 - The Holy Bible; 9 - Rob Roy, by Sir Walter Scott; 10 - Tales of Terror and Mystery, by Sir Conan Doyle; 11 - The Lost World, Sir Conan Doyle; 12 - Through the Magic Door, Sir Conan Doyle

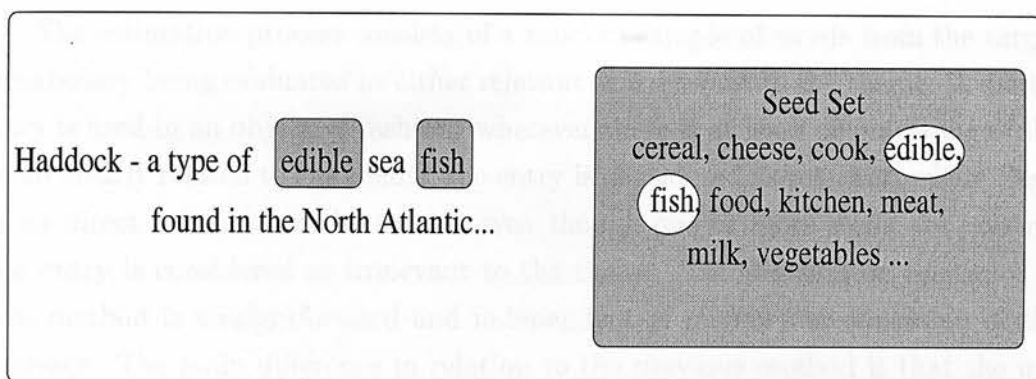


Figure 4.1: Dictionary entry with two seeds

the candidate seed set, are present in the description of the entry for *haddock*, which makes it an eligible word for the target vocabulary. However, due to polysemy (one word with more than one meaning) and to figures of speech (such as hyperbole, metaphor and so on), not all words obtained this way are really related to the theme in focus. Indeed, one can be given *food for thought* or be charged with *cooking the company books*,³ which are undesirable meanings for the proposed context.

Although the idea looks simple, its implementation may involve subtle variations, which are explored in the next sub-sections.

4.2.2 Methods to evaluate target vocabularies

How good is a target vocabulary for the purpose of this research? The first attempts to estimate the relevance of the target vocabularies obtained in experiments were carried out through surveys of students. In these the words of a significant sample would be classified either as relevant or irrelevant for the context, or unknown by the assessor. An Internet application has been developed and 20 different target vocabularies have been evaluated through this method. A description of the method and the summary of its results are given in Appendix B.

Due to the necessity of evaluating a great number of experimental outcomes, and because of the relatively slow process of gathering information through the method just mentioned, a simpler and quicker method has been adopted. It has been compared to the previous one, and shows results somewhat equivalent for the purposes of the research.

³Both examples are from the definitions of the words *food* and *cook* in the Longman Dictionary of Contemporary English, Third Edition.

The estimation process consists of a random sample of words from the target vocabulary being evaluated as either relevant or irrelevant to the theme. A dictionary is used in an objective fashion: wherever there is at least one meaning of the word clearly related to the theme, the entry is considered as relevant; where there is no direct reference to the theme, even though one or more seeds are present, the entry is considered as irrelevant to the theme. The decision on relevance in this method is straightforward and independent of memory or education of the assessor. The main difference in relation to the previous method is that the unknown category is not used. Though there is no information about whether the word might be interesting to the public in general, as the word frequency has been used to filter the words that are not commonly used, the lack of this information is not thought to be critical. This simplified process allowed the evaluation of all the vocabularies in this research.

4.2.3 The parameters for the selection

Ideally, the selection process should obtain large vocabularies with high percentage of relevance for the theme in focus. Even for very successful automatic selection, human validation and some manual discarding of undesirable words from the selection outcome is necessary. A low percentage of relevance may require a costly discarding process. Hence, the next question is: "How can a high percentage of relevance in the target vocabulary be achieved?"

Undoubtedly the quality of the seed set is the most important factor for obtaining a good target vocabulary. The first step is to find out a reasonable set of words to be the seeds for the selection process. Besides that, some selection parameters have been defined, in order to give the process more flexibility and to allow the comparison of different selection approaches. The parameters of the selection are:

1. The amount of seeds required - this is a numeric parameter which defines the number of seeds that should be present in the description of the entry being examined before including it in the target vocabulary.
2. Repetition - this is a yes/no parameter which defines whether repetitions of the same seed are to be considered in the counting of parameter 1.
3. Examples - this is a yes/no parameter that defines whether the search for seeds is to be carried out on the sentences provided as examples of use of the word contained in the entry description.⁴

⁴As described in Appendix A, the research's dictionary entries have also sentences to exemplify the use of the word along with the word's description.

4. Not-words - this is a yes/no parameter to define whether “not-words” are to be considered in the search or not. Each seed can have up to three not-words associated with it, and when they are also present in an entry with the correspondent seed, that entry is ignored.
5. And-words - this is a yes/no parameter to define whether “and-words” are to be considered in the search or not. Each seed can have up to five and-words associated with it. When and-words are considered, the word will figure in the target vocabulary only if there is at least one and-word also present in the description (or in the examples) of the entry.

Further examples and considerations of seed set definition, and of the use of the parameters above, are given in the following sub-sections.

4.2.4 The definition of the seed set

Three different themes have been explored to test the selection methods:

- *food* - words to be used as stimulus for teaching workers in the *food* industry (cooks, waiters).
- *sea* - words for teaching people involved with the sea world (fishermen, sailors).
- *agriculture* - words to be used for workers of the land - people who plant, harvest, breed animals and so on.

The simplest seed set would be composed of one word only. As an example, consider a seed set for the theme *food*, containing only the most obvious seed for the case: *food*. The experiment with this single seed set retrieved 827 words from the regular frequency words dictionary.⁵ Another experiment with the runner-up seed *cook** (words starting with *cook*) selected 313 words for the target vocabulary. Finally, a third experiment with both seeds (either one and/or the other) retrieved 1024 words. This latter result is not the simple addition of the previous ones because some words contain both seeds in their description.

The evaluation of the vocabularies obtained with these experiments revealed that only 53 percent of the words selected with the seed *food* were really related to the theme, compared to 59 percent when selecting with the two seeds (*food* and *cook**). This means that the seed *food* alone would yield around 438 relevant words (discarding 389) and that the seed set with both seeds would produce around 604 relevant words (discarding 420 words).

The seed set should be designed to capture as many words related to a particular theme as possible. An estimate over the whole regular frequency words

⁵Unless otherwise stated, all the experiments in this and in the next sub-sections have used the regular frequency words dictionary, with its 19,748 entries, as the words source.

dictionary using the estimation approach described in the last sub-section suggested that 6.12 percent of its entries might be related to the theme food. This means that some 1,280 entries are expected to be related to this theme in that dictionary.

The definition of the seed set is an intuitive process which can be improved with some interaction with the system in order to examine partial results, to eventually get to an acceptable target vocabulary. The preliminary tests showed that each new seed added to the seed set might contribute in different ways to the resultant vocabulary.

For the theme *food*, for instance, it seems clear that words like *food*, *meal*, *cook*, *eat*, and so on would constitute good seeds, for they are rich in meaning and very focused on the theme. However, some other words may be in some way related to the theme, but they are either too generic, involving other domains (e.g. *salt*, *sugar*) or too specific, representing a particular kind of element (e.g. *pasta*, *pizza*,). These words may or may not work as good seeds. Finally, functional words, such as articles, pronouns, adverbs, and conjunctions are unlikely to constitute good seeds for any theme, for they are not tied to any specific meaning.

In order to start the process of definition of the seed sets for the chosen themes, a list of the main nouns and verbs related to them was produced and tested with the system. The initial seed set for *food*, shown in Table 4.1,⁶ has 44 seeds. Although there is room for many improvements it performed well. By examining Table 4.1 one can see that, on the one hand, the supposedly promising seed “starve” selected only 2 words. This does not imply that it exists only in the description of two words in the whole dictionary, for some other words might have been selected by other seeds before this one got examined, but it certainly implies that it might not be a good seed. On the other hand, the seed “fish” selected 152 words, some of them possibly not related to food, for not every fish is edible.

The seed set of Table 4.1, requiring the presence of at least one seed per entry produced a target vocabulary of 2992 entries. An interesting issue arises when we try to fine-tune the theme domain with a question like “should *drink* be considered as a seed for *food*?”. An altered version of the seed set of Table 4.1, without the seed *drink*, was tried, producing a target vocabulary of 2889 entries, 103 entries less than the original seed set. As *drink* had selected 241 words, we could expect that the new set would be that much smaller. This was not the case

⁶A seed ending with an asterisk requires matches only up to the character preceding the asterisk. The amount beside each seed is the number of entries selected by that particular seed. These are values obtained in selections requiring the presence of only one seed in the description.

Seed	Entries	Seed	Entries	Seed	Entries	Seed	Entries
aliment*	12	dessert	9	flesh	19	nutri*	8
beef	40	diet	88	fried	2	oven	13
bird	185	dinner	65	fruit*	136	pasta	6
breakfast	33	dish	23	juic*	20	plate	63
cake	53	drink	241	kitchen	69	salad	6
candy	7	eat*	219	knife	33	spic*	22
cheese	73	edible	44	lunch*	37	starve	2
cook*	273	fish	152	meal	24	sweet	62
cow	100	fat	66	meat	38	table	115
dairy	7	food	410	milk	25	tast*	87
delicious	12	fork	12	nourish*	7	veg*	74

Table 4.1: Seed Set for Food

because the scanning process is sequential and the search on the entry ends as soon as a seed is encountered. Consequently, other seeds must have occurred in 138 of the entries previously selected by the excluded seed, which caused those entries to figure in the target vocabulary anyway. For these 138 entries it is clear that there *is* at least another seed beyond the seed *drink*, which may indicate that these entries are really likely to be related to the theme in focus.

Seed	Entries	Seed	Entries	Seed	Entries	Seed	Entries
aboard*	15	captain	30	island	113	shark	2
admiral	14	cargo	31	isle	3	shell	53
alga*	15	cetacean	2	marine*	36	ship*	170
aquatic	34	coast*	179	maritime	2	shore	9
archipel*	11	coral	9	navigat*	18	submarine	5
armada	3	crew	41	ocean*	32	swim	15
ashore	12	crustacea*	11	pacific	18	travel	114
atlantic	66	ferry	10	pirate	5	transport*	75
bay	77	fish*	267	port	78	water*	577
beach	46	gulf	64	salt	70	wave*	131
blower	1	harbour	15	sand*	130	whale*	18
boat	177	hook	23	sea	216		

Table 4.2: Seed Set for Sea

The seed sets used for sea and agriculture, with the results of selection per seed, are presented in tables 4.2 and 4.3 as a matter of illustration.

4.2.5 The use of not-words

The realisation that some interesting seeds might belong to more than one context inspired the introduction of the *not-words* feature into the selection process. The



Seed	Entries	Seed	Entries	Seed	Entries	Seed	Entries
agricul*	84	fertili*	59	measure*	258	shovel	1
bean*	37	flour	40	month	99	soil	39
breed*	104	flower	80	oat	2	tool*	87
bush	55	fruit	175	plant*	542	tractor	4
buy*	260	fuel	81	plough	5	trade	96
cattle	74	goods	204	produce	175	transport*	52
cereal	22	govern*	897	region	202	tree*	165
commerc*	144	grain*	61	rice	21	vegeta*	64
corn	31	grass	69	root	20	water	403
cow	24	harvest	5	season*	58	weather	111
crop*	81	herb	11	seed	14	weed*	6
dig	22	land	345	sell	51	wheat	5
farm	53	market	113	sheep	28	year	281

Table 4.3: Seed Set for Agriculture

implementation of this facility is through the declaration of a set of words, named not-words, which are connected to each seed in a way that whenever an entry matches a seed *and* any of the not-words, that match is ignored. An example of a not-word is “song” for the seed “bird”, for song-birds normally are not eaten. In the current implementation of the selection process up to three not-words can be declared for each seed. After the selection, the system reports how many words have been selected by each seed and how many words have been discarded because of each not-word.

Another experiment has been carried out to evaluate the effectiveness of this mechanism. In this experiment not-words have been added to some of the seeds of Table 4.1 and the selection has been run again. Table 4.4 gives only the seeds that have been modified, followed by the corresponding not-words. The amounts in brackets after the seeds tell the number of words that have been selected by the seeds. The amounts after the not-words count the entries that have been avoided because of that not-word. The seeds that appear in Table 4.1 and not in Table 4.4 were in the new seed set without any not-word.

The target vocabulary obtained with this new seed set with not-words was 289 entries shorter than with the previous one, without not-words, counting 2,703 entries. Amongst the discarded words were the words *cuirass* (“a piece of armour of leather or metal, consisting of shell, plate, or scales.”), *curlew* and *dodo* (both “large birds”), *finch* (“a small bird”), *ivy* (“a poison fruit”), and many others.

All the discarded words have been examined and 41 of them *were* related to the theme food. However, the vast majority of them, corresponding to 85.8 percent of the total, were discarded for a good reason. This rate is much better than the estimate of 58.5 percent of irrelevant words for the whole vocabulary,

Seed	Not-word 1	Not-word 2	Not-word 3
bird (85)	small (53)	game* (17)	song* (16)
drink (157)	alcoho* (84)		
eat (215)	corro* (4)		
fish (122)	small (29)	huge (3)	
fat (54)	large (9)	weight (2)	heavy (2)
flesh (17)	human (2)		
fruit* (134)	poison* (2)		
meat (36)	human (2)		
milk (19)	mother (3)	breast (4)	contamin* (1)
plate (46)	metal* (14)	game* (4)	
table (85)	game (11)	billiard (8)	work (12)

Table 4.4: Seeds with not-words for Food

indicating that the not-word scheme may contribute positively to the selection process.

4.2.6 Varying the number of seeds required per entry

How relevant to a certain theme is a vocabulary selected by the presence of just *one* seed in the description of the entries? It certainly depends primarily on the semantic strength of the particular seed. The word *food*, for instance, *is* a good seed for the food theme. In the previous selection with the seed set of Table 4.1, this seed was present in at least 410 entries, out of 2992 selected entries in total. How adherent to the theme were those words? A selection with only the seed *food* was made in order to answer this question. As mentioned before, the seed *food* running alone selected 827 entries from the dictionary, from which only 53 percent were relevant, resulting in a target vocabulary of 438 relevant words.

With a balanced seed set it is reasonable to expect that the presence of multiple seeds on the description of a dictionary entry would increase the chance for that entry to be relevant for the theme. Indeed, an experiment requiring the presence of the both seeds *food* and *cook** in the description for the selection of the entry produced a vocabulary of only 116 words, but with a high degree of relevance (94.8 percent). In this case, only six words should be discarded to obtain a hundred percent of relevance to the theme.

Although this approach enhances the percentage of relevance of the words it fails in relation to the size of the resultant set. Of course there is a great number of words which were refused in the selection even though they are relevant to the theme. The next step is to try a bigger seed set, as the one in the table 4.1, with a selection that admits *any* two seeds present as condition for accepting the

entry. Running this way, the selection produced 765 words with an estimated 78 percent of relevance to the theme, which would produce a vocabulary of 596 words useful for the teaching process. Keeping in mind that the ideal selection would obtain around 1280 words, this is not a bad mark, although there is still room for improvement.

What happens if the presence of more seeds is required for the selection of words? A selection has been carried out using the seed set in table 4.1 and requiring the presence of *three* seeds either in the description of the word or in the examples of use. It produced a vocabulary of 314 words with an index of 94 percent of relevance. The selection with *four* seeds produced 149 words a hundred percent relevant to the theme. The next section explores a slight variation of the multiple seeds approach.

4.2.7 Allowing the repetition of seeds

In the experiments described so far the system looked for the occurrence of *different* seeds in the description of the words in the source dictionary. In the following ones this concept has been relaxed so the multiple occurrence of the same seed would be counted as if they were different seeds. Always using the seed set of Table 4.1, the selection requiring any two seeds in the description produced 1021 words, with 71 percent of relevance. In this case we could expect that 725 useful words would come out of the process, which is considerably more than the 596 words expected without the repetition of seeds.

A selection requiring the presence of three seeds resulted in 533 words in the target vocabulary, found to be 94 percent relevant to the theme; with four seeds, 346 words were obtained, with an index of relevance to the theme of 99 percent. Table 4.5 summarises the results described so far.

Seeds required	Repetition allowed?	Not-words considered?	Selected words	In context percentage	Useful words
0	-	-	19,748	6.12	1280
1	-	no	2992	40.81	1221
1	-	yes	2703	41.41	1119
2	yes	yes	1021	71	724
2	no	yes	765	78	596
3	yes	yes	533	92	490
3	no	yes	314	93.94	295
4	yes	no	346	99	343
4	no	yes	149	100	149

Table 4.5: Comparative study of selections (Food)

4.2.8 Selecting with bigger seed sets

The figures in Table 4.5 were obtained with the initial seed set of 44 seeds, trying different requirements. What happens when the seed set is enlarged with synonyms of the existing seeds and with new words? To investigate the effects of a bigger seed set, a small number of students was invited to critically analyse the initial seed set and suggest deletion and addition of words to act as seeds for the theme food. The seed set compiled from this survey is shown in Table 4.6. The numbers in that table correspond to the performance of the seeds after a selection requiring the presence of at least one seed in the description of the word.

Seed	Entries	Seed	Entries	Seed	Entries	Seed	Entries
aliment*	12	dinner	65	kitchen*	57	rice	14
bake*	69	dish*	129	lamb	5	roast*	2
barbecue	5	drink	138	lunch*	34	salad	5
bean	5	eat*	203	meal*	17	salt*	39
beef	40	edible	38	meat	27	sauce	7
bird	54	egg*	71	milk	12	snack	0
biscuit*	26	fish	106	noodle*	1	sour	1
bitter*	62	fat	47	nourish*	7	soup	5
bread	83	food	354	nut	16	spic*	11
breakfast	31	flavour*	30	nuts	4	spoon	4
butter	43	flesh	17	nutri*	6	stale	6
cake	32	freezer	2	oat	2	sugar*	31
candy	5	fresh	52	oven	9	supermarket	7
cereal	17	fridge	7	pan	11	supper	7
cheese*	58	fried	1	pasta	4	sweet*	69
chicken	28	fries	2	pastry	5	table	75
cook*	228	frozen	17	pepper	3	tast*	67
corn	32	fruit*	113	plate	25	utensil*	2
cream*	80	fry	1	poultry	4	veg*	62
cup	84	gastro*	9	pork	4	wheat	7
dairy	9	grain*	58	potato*	14	wine	49
delicious	9	hen	2	pudding*	1		
dessert	4	herb	3	pulse	7		
diet	86	juic*	16	restaurant	37		

Table 4.6: 93 Seeds for Food

Table 4.7 was obtained after a series of experiments carried out with different selection parameters. When compared to Table 4.5, Table 4.7 suggests that bigger seed sets favour the selection of more entries, but with increasing proportion of non relevant words amongst them.

Seeds required	Repetition allowed?	Not-words considered?	Selected words	In context percentage	Useful words
1	-	yes	3355	35.8	1201
2	yes	yes	1322	61	806
2	no	yes	1012	75.66	765
3	yes	yes	730	82	599
3	no	yes	473	89.89	425

Table 4.7: Selections with the 93 Seeds

4.2.9 The use of examples

The results we obtained in these experiments would probably vary significantly depending on the source of words in use. The three different dictionaries mentioned above were put together in an attempt to increase the universe of available words in order to obtain better results. Although some dictionary developers would prefer to describe the words concisely, the use of sentences with examples to illustrate their application is a common feature in popular dictionaries.

The Collins English Dictionary only describes the semantics of the word in a concise and technical fashion. The Cobuild II dictionary has both approaches, describing the semantics of the words in a friendly way and exemplifying their use, in separated fields. The dictionary prepared in this research inherited all the information from these dictionaries. The examples of use of words are available in 72 percent of its entries. In a selection covering the description and the examples, 69.03 percent of the scanned text come from descriptions and the remainder, 30.97 percent, belong to examples.

As an illustration, examine the texts of the description and of the examples in a typical dictionary entry (the word *enclose*):

Headword: *enclose*

Description:

- (1) To close; hem in; surround; to surround (land) with or as if with a fence; to put in an envelope or wrapper, esp. together with a letter; to contain or hold.
- (2) If a place or object is enclosed by something, the place or object is inside that thing or completely surrounded by it; if you enclose something with a letter, you put it in the same envelope as the letter.

Examples:

- (1) The rules state that samples must be enclosed in two water-tight containers.

- (2) Enclose the pot in a clear polythene bag.
- (3) The surrounding land was enclosed by an eight foot wire fence.
- (4) The enclosed waters of the Baltic.
- (5) I have enclosed a cheque for 10 pounds.
- (6) He tore open the creamy envelope that had been enclosed in the letter.
- (7) The enclosed leaflet shows how Service Care can ease all your worries.

The first paragraph of the description above comes from the Collins English Dictionary. The second paragraph and the examples are from Collins Cobuild II. The differences in style are quite clear. The examples of use can be regarded as an excerpt of a specialised text (or corpora) in the main subjects related to the word.

A closer examination of the examples above reveals that this particular entry would be selected for the theme food (with the 93 seeds seed set and just one seed required) because of the word *creamy* in the sixth sentence. It would be selected for the theme sea as well, because of the word *watertight* in the first sentence (or *water* in the third). And finally, it would be in the target vocabulary for agriculture for it contains the word *land* in the third sentence. As this entry for *enclose* does not seem to be close to any of the three contexts in focus, it is sensible to evaluate whether it is appropriate to scan the sentences provided as examples of use of the words for this purpose. This may also be an indication that the use of corpora as a source of words for this purpose is not as efficient as the use of formal dictionaries is.

Table 4.8 shows the results of five selections with the 93 seeds seed set for food, without scanning the texts of examples. Compare it to Table 4.7.

Seeds required	Repetition allowed?	Not-words considered?	Selected words	In context percentage	Useful words
1	-	yes	2147	52	1116
2	yes	yes	1206	64.92	783
2	no	yes	717	85	609
3	yes	yes	588	87	511
3	no	yes	339	93.16	315

Table 4.8: Selections with 93 Seeds without Examples

The figures in Table 4.8 are compatible with the expectation that the number of selected entries would decrease, but the percentages of relevance would rise, when the sentences with examples of use are not considered.

4.2.10 Refining the source of words

The results in Table 4.8 inspired the realisation of another experiment evaluating the performance of the methods when only the entries originating from the Collins English Dictionary (CED) are used, that is, without the descriptions and the examples from the Collins Cobuild Dictionary. The adapted dictionary with only the words from the CED has 17,454 words, 11.65 percent fewer words than the dictionary developed in the research. This is due to the fact that the missing 2,294 words are only present in Collins Cobuild Dictionary. It is, consequently, a smaller source of words. The results of the selections with the seed set of 93 seeds for the theme food are presented in Table 4.9.

Seeds required	Repetition allowed?	Not-words considered?	Selected words	In context percentage	Useful words
1	-	yes	1634	59.79	976
2	yes	yes	609	88	535
2	no	yes	482	90.22	434

Table 4.9: The 93 Seeds selection over the CED

The good performance of the proposed scheme is dependent on the proper knowledge and use of the characteristics of the elements involved in the process, that is, the source of words, the seed set, and the parameters of selection. The numbers from the experiments shown in Table 4.9 suggest that the conciseness and the regularity of the dictionary can provide better control over the outcomes of the system. The selections in that experiment resulted in smaller amounts of selected entries, probably because the new source of words is both more concise, and has 11.65 percent fewer words than the previous one. However, the entries originated from the Collins English Dictionary seem to be more likely to support a clean selection than the ones from the Collins Cobuild Dictionary. The percentages of relevant words in these selections were greater than the other selections when comparing selections with about the same level of achievement. As with the examples of use of the words, friendly explanations seem to be more verbose at a cost of increasing the need for discarding irrelevant words later in the process. The same effect of obtaining more words can be achieved by careful choice of the seeds, the not-words and the selection parameters, reducing the risk of bringing in undesirable words.

4.2.11 The use of and-words

The seed set approach is easy to understand and reasonably efficient for the purpose of selecting words. Nevertheless, its logical expressibility is limited: a selection with just one seed required is equivalent to a logic expression in the form

seed-1 or seed-2 or seed-3 ...

A selection with two seeds required would be equivalent to a rather complex logic expression, each involving a pair of seeds of the seed set, in the form:

(seed-1 and seed-2) or (seed-1 and seed-3) or ...

The biggest inconveniences of the seed set abstraction as stated so far are that it does not allow flexibility in expressing more sophisticated combinations of seeds, and that all the seeds have the same weight in the evaluation. The and-word feature was introduced in order to allow more complex declarations with different weights for the seeds. With and-words it is also possible to express logical requirements such as:

(seed-1 and (seed-2 or seed-3 or seed-4 or ...)) or
(seed-5 and (seed-6 or seed-7 or seed-8 or ...)) or ...

While this alternative logical device empowers the selection scheme, it does so at a cost of bringing more complexity to the definition of the seed set.

Table 4.10 gives the values of the selection requiring the presence of just one seed either in the description or in the examples of use of the words. With this approach, the way of thinking about the seed set is completely different. For instance, the word *food* in this case is not thought of as a seed, but as an and-word, that should appear as a complement to a seed. In Table 4.10 the word *food* is in fact used as an and-word for various seeds. The seed set in Table 4.10 was based on the 93 words seed set, but ended up with 49 seeds. Not-words were also used in that selection.

After a number of trials and adjustments, the seed set shown in Table 4.10 selected 1096 entries from the regular dictionary, with a relevance of 72.9 percent, producing an estimated 799 words in the target vocabulary, which is equivalent to results obtained with less effort in some of the previous attempts.

Seed	Not-1	Not-2	Not-3	And-1	And-2	And-3
bake*(69)						
bean*(29)	anima*(7)					
beef(40)						
bird(21)	small(52)	game*(8)	song*(4)	food(12)	eat*(6)	prepare*(3)
bitter(26)				tast*(21)	food(5)	
biscuit(26)						
bread(83)						
breakfast(31)						
butter(41)						
cake*(43)						
cereal(6)				food(5)	eat*(1)	
cheese(58)						
chicken(12)				food(8)	cook*(4)	
cream(21)				food(11)	eat*(5)	spread*(5)
cutlery(11)						
dairy(9)						
delicious(6)				food(4)	cand*(1)	suit*(1)
dessert(11)						
diet(34)				food(19)	eat*(7)	health*(8)
dinner(23)				eat*(13)	food(3)	serv*(7)
dish(10)	metal*(3)	game*(2)	tele*(1)	food(0)	eat*(1)	
edible(65)						
egg*(9)				eat*(6)	cook*(1)	boil*(2)
fish(31)	small(40)	sport*(10)	game*(6)	eat*(7)	food(19)	prepar*(5)
fat(10)	large(9)	weight(3)	heavy(1)	food(7)	eat*(2)	cook*(1)
flavour(34)						
flesh(13)	huma*(4)			food(7)	eat*(4)	prepar*(2)
freez*(7)				food(7)		
fresh(13)				food(10)	eat*(2)	cook*(1)
frie*(3)	friend*(351)			food(3)		
frozen(6)				food(6)		
fruit*(15)	poison*(3)			food(9)	eat*(4)	meal*(2)
grain*(10)				food(9)	eat*(1)	
lunch*(9)				food(4)	eat*(4)	prepar*(1)
meat*(16)	human*(3)			eat*(6)	cook*(3)	food(7)
milk(23)	mother*(5)	breast*(2)	contamin*(1)			
nut*(7)	bolt*(3)			food(4)	eat*(3)	
pan(5)						
refreshm*(6)						
restaurant(19)				meal*(6)	food(8)	eat*(5)
salad*(16)						
salt*(6)	formula(1)			food(4)	cook*(1)	eat*(1)
sugar(12)	formula(2)			food(7)	eat*(2)	drink(3)
sweet*(8)				food(7)	eat*(1)	
table(11)	game(13)	billiard(3)	work(13)	food(14)	eat*(3)	serv*(4)
tast*(16)	bad*(10)					
vegetable*(29)						
wine(83)						

Table 4.10: Selection with and-word

4.2.12 Discussion and recommendations on seed sets

These experiments illustrate the relationships between the different choices of seeds and the parameters adopted in the word selection process. The seed sets defined here are not claimed to be the ultimate choices for selecting words for the themes under study. The data on the theme explored the most - *food* - can give valuable insights into this question. Many of the seeds used in Table 4.6 could have been discarded without changing the overall picture. For instance, the seed *noodle**, although occurring 6 times in the regular frequency dictionary brought in just one entry, and *snack*, despite occurring 18 times in the dictionary, did not cause a single word to be selected, for the entries where it appeared had already been selected by other seeds. Although these seeds might play a role in selections where more than one seed is required, they would probably make a modest contribution to the whole process. Forty of the seeds in Table 4.6 selected less than 10 entries each, totalling 177 selected entries all together, out of a vocabulary of 3,355 entries. This corresponds to 5.27 percent of the total of entries selected. A selection without these seeds would probably recall many of these 177 entries, for, as explained before, the scanning process is sequential and stops as soon as any seed is found in the text. This demonstrates that there is still room for improvements in the seed sets.

Care must be taken with respect to the use of the asterisk, as a wild card, in the definition of the seeds. This facility is interesting for capturing inflections of a *head-word*, but sometimes it can be tricky. As examples, the use of *snack** instead of *snack* would have increased the number of occurrences of the seed (and its inflections: *snacks*, *snacking*, *snacked*) in the dictionary from 18 to 29, and possibly given it a more significant participation in the process. However, *bitter** and *dish** brought in words related to *bitterly* and *dishonorable*, for instance, which are totally unrelated to the theme in question.

These observations indicate that conciseness and precision are fundamental requirements of the seed set. Recall that the Generative Themes of Paulo Freire, as described in Chapter 2, used to be from 15 to 20 words long. It is clear that the manual procedures used to determine Freire's teaching vocabulary are more flexible than the ones here implemented, for free intervention of the human intelligence would allow interpretations and derivations that would be hard to work out automatically. Nevertheless, all the words used in his basic teaching scheme would be derived directly from the Generative Themes. This is a confirmation that a small seed set can successfully generate a good target vocabulary.

Another key issue in the selection processes is the compromise between the

amount of relevant entries caught during the selection and the amount of irrelevant entries that must be discarded in order to get to a reasonable teaching vocabulary. According to the estimate already described and expressed in the first line of Table 4.5, there should be a limit of 1280 relevant entries for the theme food in the project dictionary. What is the meaning of this number? Considering that the low frequency dictionary (LFD) with 841 entries will be appended to the selected vocabulary (SV), and that a rise of 84.2 percent of the vocabulary is expected because of the inflections, synonyms and antonyms,⁷ this process can end up achieving a final vocabulary (FV) of 3,923 words to support the teaching activity:

$$\text{FV.Size} = (\text{SV.Size} + \text{LFD.Size}) * 1.842$$

$$\text{FV.Size} = (1289 + 841) * 1.842$$

$$\text{FV.Size} = 3923.$$

This last figure is a statistical upper limit that could be achieved by examining the entire regular dictionary, discarding 93.88 percent of its entries (18,538 words), which would be a huge task. However, if the source of words used is the selection of line 1 in Table 4.7, for instance, 3,355 words would be examined with a discarding expectancy of 64.2 percent (2153 words). This is still a significant amount of manual work, but not as significant as in the case of the whole dictionary. The selection of line 3 of Table 4.7, in its turn, would provide a vocabulary of 765 entries after 247 discards. With this last figure, the final vocabulary would have 2,958 words:

$$\text{FV.Size} = (765 + 841) * 1.842$$

$$\text{FV.Size} = 2958.$$

The above figure seems to be quite acceptable, after a relatively small amount of manual discarding. With a bigger effort applied to the definition of the seed set, and the right choice of the selection parameters, the achievements can be significantly improved. The ultimate test of acceptability of these results will come later in the next chapter, with the distribution of the vocabulary across the various phonological features of the language represented in PA curricula.

As mentioned before, the manual discarding process seems to be somewhat inevitable after the automatic selection. What should be done, at this point, is to support the expert teacher to comfortably execute this task. This support involves the availability of interactive software to allow the browsing of the vocabulary, the pointing out of words that should be discarded or included in the

⁷This is the proportion of between the head-words and the derivative words in the system's dictionary, obtained by actual counting.

final vocabulary, along with other further selection parameters. The proposed system allows the expert teacher to determine the minimum word frequency, the maximum age of acquisition,⁸ the maximum number of letters or the maximum number of phonemes a word should have to figure in the target vocabulary.

It is not clear from the approaches tested whether would be worthwhile considering more generic logic schemes for expressing the selection requirements (e.g. expressions with the connectors *and* and *or*, the negation *not* and priority brackets). From many attempts with some of the more powerful structures, it looks like it might be too difficult to express such requirements in a way that works efficiently. The seed set abstraction with not-words, number of seeds required and repetition factor seems to function quite well for the purposes of this thesis. Besides, it is likely that some of the applications of these methods would be more related to *contextual* vocabulary than to *thematic* vocabulary. In these cases the target vocabularies should be related not to a particular area, such as in the cases studied so far, but to a set of co-related universes, as in Freire's generative themes.

The intuition that the presence of multiple seeds related to a context in a description of a word would be a strong indicator of relevance of that word for the context has been experimentally confirmed. Table 4.11 was obtained after a series of trials with the generative themes for the town of Ceilandia [Freire, 1997], expanded with some synonyms. It performed well, producing 767 words in context, out of 937 entries selected, with a relevance index of 81.91 percent, with the simple requirement of any four seeds present in the dictionary entries for the word to be selected.

A final recommendation in relation to these experiments is that the source of words should rather be concise, instead of verbose, and as abundant as possible. The merging of CED with one or more dictionary with the same characteristics, with some care to not simply duplicate the common entries, but selecting the most appropriate ones, could be worthwhile.

⁸The age of acquisition is available for a small percentage of words, extracted from the MRC II Psycho-Linguistic Database, expressed in terms of average number of weeks of age when the particular word is known.

Seed	Not-word 1	Not-word 2	Not-word 3
plot (8)	plan* (13)	crim* (1)	film* (3)
land (334)	plane (10)	air*(84)	
food (353)			
game (245)			
football (46)			
dry (59)			
weather(77)			
drought (8)			
dust* (28)			
rubbish (14)			
rain* (101)			
election* (101)			
government* (315)			
vote* (92)			
shelter* (45)			
slum* (5)	slumber*(1)	slump*(15)	
house* (221)			
home* (161)			
religion* (56)			
christ* (185)			
church* (166)			
job* (168)			
wage* (18)			
television* (107)			
TV* (15)			
water* (380)			
community* (45)			
union* (48)			
ticket* (19)			
bus* (27)			
buses* (4)			
transport* (41)			
son* (33)			
sons* (2)			
famil* (110)			
father* (64)			
mother* (54)			

Table 4.11: Seed set for the city of Ceilandia

Chapter 5

The Phonological Selection

When the right time comes for reading and writing words, any literacy method would start teaching words with some particular features the designers consider appropriate to start up the learning process, and would evolve to more complex ones as the process goes on. PA methods are critically concerned with the order of introduction of words to their students, and many of them are indeed organised around the phonological issues in question [McGuinness, 1997]. In these methods it must be assured that the initial examples and exercises will contribute to the real comprehension of the writing system. Once the student grasps the basic mechanisms, more complex usage can be explored.

Chapter 4 has described a method for selecting a target vocabulary thematically. This chapter describes a method for distributing the selected vocabulary into the parts of the particular curriculum to be taught. Firstly the general scheme for defining the *phonological features* of the desired words is described and then the distribution of the words into a *knowledge tree* is explained. Finally, the three PA programmes selected as representatives of real PA schemes are introduced and their instantiations using the concepts developed in this thesis are described.

5.1 The description of the phonological features

A scheme has been developed to allow the classification of the words by their phonological features. In Figure 5.1 the words to be taught are to be selected according to the description of the kind of phonemes they contain. In this example, the word should have three phonemes, each one belonging to a particular subset of the English phonemes.

The proposed scheme is primarily phoneme-driven, but it can also be used to describe syllables, onsets and rimes, letters and letter clusterings. The features

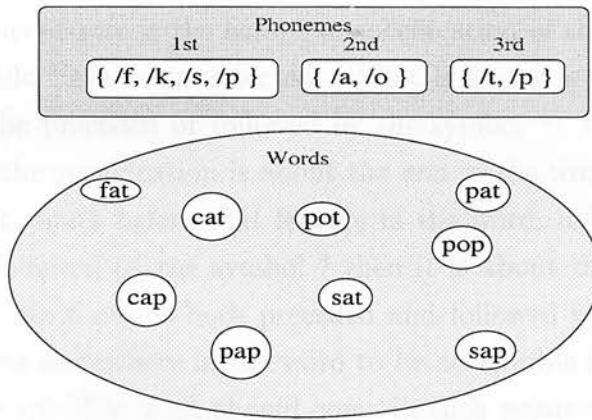


Figure 5.1: Phonological Features Definition

of the words are represented through a sequence of symbols as described below:¹

```
<complete-word-spec> ::= <word-spec> | * <word-spec> |
                        <word-spec> * | * <word-spec> *
```

```
<word-spec> ::= <position-spec> | <word-spec> <position-spec>
```

```
<position-spec> ::= <position-unit> . |
                   <position-spec> <position-unit> .
```

```
<position-unit> ::= <phoneme-spec> | <letter-spec> |
                   [ <set-spec> ] | { <set-spec> }
```

```
<phoneme-spec> ::= X | C | V | /<phoneme>
```

```
<phoneme> ::= /* any phoneme in festival system (appendix C) */
```

```
<letter-spec> ::= $ | <letter>
```

```
<letter> ::= /* any of the 26 lower-case letters of the alphabet */
```

```
<set-spec> ::= <set-element> | <set-spec>, <set-element>
```

```
<set-element> ::= <phoneme-spec> | <letter-spec>
```

¹A slight variation of the Backus-Naur Form (BNF) is used in this specification. In this variation, some informal description can be introduced in the right-hand term of the BNF statement enclosed by the symbols /* and */.

The *complete-word-spec* is the complete specification of the phonological features a word should have to be selected. What it states is that the *word-spec* may or may not be preceded or followed by the symbol *. If it is preceded by this symbol then the specification is about the *end* of the word. In other words, regardless of what comes before that feature in the word, it will be selected. If the *word-spec* is followed by the symbol * then it is about the *beginning* of the word. And if the *word-spec* is both preceded and followed by this symbol then the feature must be somewhere in the word to be acceptable in the selection.

The *word-spec* specifies what should come in each *position* in the word (the positions can be phoneme or letter based). In the *phoneme-spec* description the symbol X stands for “any phoneme in the position”, C stands for any consonant and V stands for any vowel. For instance, the command XVC means “words with any kind of phoneme in the first position, a vowel in the second and a consonant in the third position”. Any phoneme symbol preceded by a slash will require that the word should have that phoneme in that position to be selected (e.g. /kVC would select words like “cat”, “cap”, “kit”, “cut” and so on). Consult Appendix C for the complete list of phonemes.

In the *letter-spec* description the symbol \$ means that any letter can appear in that position, whereas when a letter is specified then the word must have that letter in that position to be selected. The letters must be represented in lower case. For example, “*ion” would capture all words ending with “ion”, as action, innovation, nation, and so on.

Alternative possibilities in one position can be indicated by bracketing: for instance [/b, /c, /f] means “either the phoneme b, c or f in that position”. The use of square brackets means optional, that is, that position can be empty. The use of curly brackets implies that one of the alternatives must be present at that position for the word to be selected, as illustrated in figure 5.1. The examples that follow will facilitate the comprehension of the model and illustrate its use for different purposes.

1. CVC.

The requirements are for words with three phonemes, the first must be a consonant, the second a vowel, and the third a consonant. It will select words such as *bat*, *mug*, *lap*, *shot*, and so on.

2. [/f, /k, /s, /p, /m, /t] { /a, /o, /aa } { /t, /p, /m, /s }.

The word must have two or three phonemes, the first is optional, but if present must be one of those in square brackets. It will catch words such as

fat, cat, sap, mop, pot, at, as, and so on.

3. CVCC.

The word must have four phonemes: consonant, vowel, consonant, consonant. It will select words like: *third, bird, bald, text,* and so forth.

4. {s,c}{h}{\$}{\$}.

Selects four letter words starting either with *sh* or *ch*, like *shot, chat, shut, chap, shed,* etc.

5. \$\$th.

Four letter words ending with *th*, as in *bath, both, with, path,* or *myth*

6. {X}{/ou}{*}.

For words with the sound /ou in the second phonemic position, as in *boat, note, cone,* etc.

7. {X}{/e@}{*}.

For words with the sound /e@ in the second position, as in *her, girl, shirt, burn, bird,*...

8. */sh/@/n.

Looking for words with the termination similar to the one of the word “action” (compare to a previous example with letters: “*ion”).

9. */i/ng.

Looking for words with the termination “ing”.

10. */l/ii.

Looking for possible adverbs (ending with “ly”).

The scheme just described is flexible enough to represent virtually any phonological feature which it might be necessary to discriminate in the literacy teaching methods. The next step is to link this scheme to the curriculum to be taught, so the selected words would come up at the appropriate moment. The definitions of the interesting features will be incorporated later in the knowledge tree, described next.

5.2 Introducing the Knowledge Tree concept

The Knowledge Tree is a hierarchical structure designed to organise the subject domain to be taught and to support the definition and further application of the curriculum. Conceptually it is similar to the content lists of many literacy programmes. As an example, consider the knowledge tree for the Phono-Graphix Method for teaching literacy [McGuinness and McGuinness, 1998] shown in Figure 5.2.

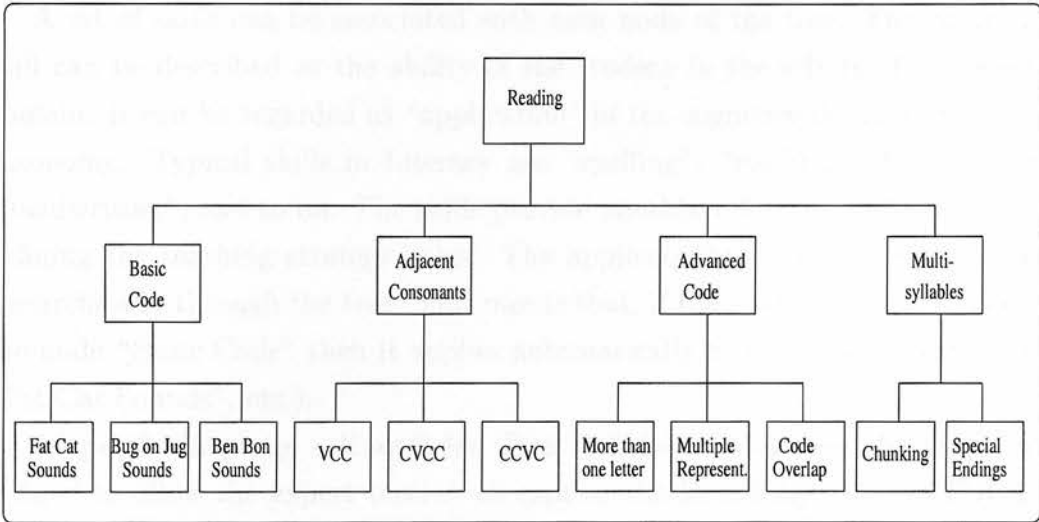


Figure 5.2: Phono-Graphix Knowledge Tree

The Phono-Graphix activities are organised around the structure expressed in Figure 5.2.² The structure is hierarchical but not rigid. It codes what corresponds to the “knowledge” in the cognitive domain in Bloom’s taxonomy. The *root* of the tree is its most comprehensive *node* and represents all the knowledge to be taught. The knowledge can be subdivided into areas of knowledge represented by more specific nodes linked to the root through lines called *edges*. These new nodes can be again subdivided into branches of knowledge in a process of successive refinement, up to the leaves of the tree, with no subdivision. These lower level nodes correspond to the “knowledge of specifics” in Bloom’s taxonomy [Bloom, 1956].

At the beginning of the teaching process the activities tend to be concen-

²This tree is an independent representation of the Phono-Graphix Method. It is not explicitly stated in any of the documents that compose the Phono-Graphix packet. In order to obtain it the author has studied the vocabulary and the sequence of activities recommended in the method and has compiled this structure. The same approach has been carried out with the other literacy methods considered in this research.

trated in the “basic code” region, and its lower level nodes, where the clearest sounds and the less complex graphic representations are placed. As the student begins to master these basic features, incursions into more complex material are attempted. Whenever necessary, the teacher guides the student to return to any of the previously studied material. The same kind of lessons, exercises or tests can be re-used in the different regions of the tree, each time with different complexity and vocabulary. This issue is addressed later in the “teaching strategy rules”, Chapter 7, Section 7.4.

A set of skills can be associated with each node of the tree. The concept of skill can be described as the ability of the student in the affairs of the subject domain. It can be regarded as “application” in the cognitive domain in Bloom’s taxonomy. Typical skills in Literacy are “spelling”, “reading isolated words”, “handwriting”, and so on. The skills provide another reference to be used while defining the teaching strategy rules. The applicability of the skills propagates hierarchically through the tree. This means that, if the skill “spelling” applies to the node “Basic Code” then it applies automatically to the nodes below it (e.g. “Fat Cat Sounds”, etc.).

A specific language - Knowledge Tree Definition Language (KTDL) - was defined to allow the expert teacher to express the knowledge tree. KTDL is a declarative language which can be typed in through an ordinary text editor. The knowledge tree written representation uses hierarchical numbers to convey its structure. In each line there is a hierarchical number followed by the name of the node and optionally by the skills it addresses. If a node is not a leaf, its offsprings’ hierarchical numbers will start with the same sequence as their ancestor node, as illustrated below. The skills are represented by their names, separated by a colon after the node name. All these names can be multiple-word expressions, like “Basic Sounds”, “Adjacent Consonants” or “Auditory Processing”. If there is more than one skill in one command, they are separated by commas. The phonological requirements of the vocabulary to be taught at each branch of the tree are also described in the KTDL.

The formal definition of KTDL is given later in Chapter 7. Part of the code for the Phono-Graphix knowledge tree is shown in Table 5.1 to illustrate the placement of the selected vocabulary into the correspondent points of the curriculum.

The result of the compilation of the knowledge tree is that a phonologic vocabulary set is produced in correspondence to each node of the tree. Each of the vocabulary sets contains all the words with the targeted phonological features described in the node’s definition. As a corollary of this, words from the thematic

```

% Phono-Graphix Knowledge Tree (Partial)
% -----
1 - Reading: Code Knowledge, Segmenting, Blending.
1.1 - Basic Code.
    VC. CVC.
1.1.1 - Fat Cat Sounds.
    [/f,/k,/s,/p,/m,/t] {/a,/o,/aa} {/t,/p,/m,/s}.
1.1.2 - Bug On Jug Sounds.
    [/r,/h,/b,/jh,/d,/m,/t] {/uh,/i,/a} {/g,/d,/m}.
1.1.3 - Ben Bon Sounds.
    [/h,/b,/w,/l,/n,/f,/z,/s]{/e,/o,/i,/uh}{/l,/n,/t,/z,/b,/p}.
1.2 - Adjacent Consonants: Auditory Processing.
1.2.1 - VCC Words.
    VCC.
1.2.2 - CVCC Words.
    CVCC.
1.2.3 - CCVC Words.
    CCVC.
1.3 - Advanced Code: Reading Coded Text, Reading Uncoded Text.
1.3.1 - More Than One Letter Sounds.
1.3.1.1 - Ch And Sh Sounds.
    {s,c}{h}{${}$}. % selects 4 letters words starting with sh/ch
    {s,c}{h}{${}$}{${}$}. % and 5 letters words starting with sh/ch
    ${s,c}{h}{${}$}. % and 4 letters words with sh/ch in the middle
    ${s,c}{h}{${}$}{${}$}. % and so on...
    ${}${s,c}{h}{${}$}.
    ${}${s,c}{h}.
    ${}${${}$}{s,c}{h}.
    ${}${${}$}{${}$}{s,c}{h}.
1.3.1.2 - Th Sounds.
    th$. th$$$. % th and so on...
    $th$. $$th.
    $$th$. $$$th.

```

ooo

Table 5.1: Part of the Phono-Graphix Knowledge Tree

vocabulary may figure in multiple vocabulary sets, satisfying different demands for phonological features. After that, the phonologic vocabulary sets are available to be referred to in the teaching activities. The description of the activities schemata is given in Chapter 7.

5.3 Instantiating PA Approaches

Sections 4.2, 5.1 and 5.2 described the methods for selecting words related to a theme or to a context, and for distributing the words according to an organised description of the subject domain. Both methods impose restrictions on the English vocabulary, in particular on the vocabulary in the dictionary in use in this research. The main question underlying this issue rises again:

“Is it possible to obtain thematic vocabularies that would still allow the coverage of all the desirable phonologic features explored in PA approaches?”

In other words, after all the restrictions imposed on the vocabulary, will there be enough words to cover a typical PA curriculum? The only way to respond to this is through experimental evaluation with real phonological awareness methods. In this section the instantiations of the PA methods Phono-Graphix TM, Jolly Phonics TM and THRASS TM ³ in the system proposed are described. The instantiations comprise the representation of the subject domain and of the criteria for distributing the words through the domain, using the knowledge tree abstraction. After that, the thematic vocabularies obtained in the thematic selections described in Chapter 4 are used in conjunction with the knowledge trees of the methods under study to obtain the vocabulary distribution according to the phonological requirements of each curriculum item of these methods. Eventually these vocabularies will be compared to the vocabularies originally used in those literacy methods,⁴ to decide whether the procedures recommended and implemented here are able to produce enough stimuli for supporting the PA literacy teaching with thematic vocabulary.

5.3.1 The instantiation of Phono-Graphix

The Phono-Graphix literacy method is an almost pure phonological awareness scheme which has been created in a clinical environment and has been adapted to

³Phono-Graphix is a trademark of Read America Inc. Jolly Phonics is a trademark of Jolly Learning Ltd., Chigwell, England. THRASS has been developed by Alan Davies and Denyse Ritchie.

⁴The three literacy methods selected are used for teaching children, with vocabularies focusing on words suitable for youngsters. As the vocabularies generated in this research are thematically and phonetically appropriate, the comparison to take place is purely quantitative.

be used in school-based teaching. It focuses strongly on the phonological issues of the spoken and written language, and its curriculum is clearly organised around phonology. The main focus of Phono-Graphix, likewise the main focus of this thesis, is on the basic reading skills - code knowledge, segmenting, blending, and so on. It does not give much attention to comprehension or composition, at least in this stage of the reading process. Its first aim is to make the student able to decode and code the words, and master the writing system.⁵

The Phono-Graphix knowledge tree has been shown in Figure 5.2. In Phono-Graphix the subject domain *Reading* has been divided into four main nodes namely the *Basic Code*, the *Adjacent Consonants*, the *Advanced Code* and the *Multi-syllable Words*, ordered by ascending degree of complexity. This does not mean that the teaching must follow this order in a rigid way. As described in Chapter 7 there are teaching strategy rules that can modify this sequence. Each of these nodes has also been divided into smaller segments, some of which have again been divided into lower levels, not represented in Figure 5.2. The complete written representation of the Phono-Graphix knowledge tree can be seen in Appendix D.1.

The first node, *Basic Code*, and its descendants are designed to support the introduction of the principles of the writing system, with lessons, examples and exercises on phonologically simple words. They are represented in the following code:⁶

1 - Reading: Code Knowledge, Segmenting, Blending.

1.1 - Basic Code.

VC. CVC. &exclude.

1.1.1 - Fat Cat Sounds. % fat, cat, sap, mat, pot, mop ...

[/f,/k,/s,/p,/m,/t] {/a,/o,/aa} {/t,/p,/m,/s}.

1.1.2 - Bug On Jug Sounds. % rug, rid, ham, jam, dig, tag, ...

[/r,/h,/b,/jh,/d,/m,/t] {/uh,/i,/a} {/g,/d,/m}.

1.1.3 - Ben Bon Sounds. % hip, bip, let, fit, set, son, ...

[/h,/b,/w,/l,/n,/f,/z,/s]{/e,/o,/i,/uh}{/l,/n,/t,/z,/b,/p}.

The vocabulary for the node *Basic Code* will be composed of vowel-consonant and vowel-consonant-vowel words (VC and CVC). The vocabulary for the descendants of *Basic Code* are specialisations of its vocabulary. The vocabulary for *Fat*

⁵There are further extensions of Phono-Graphix for dealing with more advanced skills, but these issues are outside the scope of this thesis.

⁶In each curriculum item, after the comment character %, there are examples of words with different forms of representation for the sound.

Cat Sounds is also a VC and CVC vocabulary, but only with words that start with either /f,/k,/s,/p,/m, or /t (or none of these), with either the vowel /a, /o or /aa, and ending with one of the consonants indicated for the third position. Similarly, *Bug on Jug Sounds* and *Ben Bon Sounds*, select VC and CVC words, with different phonemes in each position. They are not more or less complex than their pairs, but just variations to be exercised. As the result of the code seen above, four vocabularies will be produced. The one for *Basic Code* will contain all the VC and CVC words except the ones selected for its descendents. This is so because of the “&exclude” option in the *Basic Code* entry.

The next node, the *Adjacent Consonants*, is intended to develop *Auditory Processing* skills,⁷ with slightly more complex words:

1.2 - Adjacent Consonants: Auditory Processing.

1.2.1 - VCC Words. % act, alp, and, ant, asp east, elm, ...
VCC.

1.2.2 - CVCC Words. % bend, bulb, camp, castle, child, coast, ...
CVCC.

1.2.3 - CCVC Words. % black, block, blouse, bread, brick, ...
CCVC.

The node *VCC Words* will be associated with a vowel-consonant-consonant vocabulary, with words like *ant*, *alp*, *elk*, *and*, and so on. The node *CVCC Words* will contain words such as *milk*, *desk*, *lamb*, *lamp*, etc., whereas the node *CCVC Words* will deal with words such as *stop*, *flag*, *frog*, *plus*, and so on. Although simple and regular, the words in this block are clearly more complex than the ones in the first one.

The next node, *Advanced Code*, however, encompasses many more features than the previous one, requiring further sub-divisions in order to isolate the vocabulary for future use in tutoring. At this point the student might be mastering basic code and stimuli for new skills are supposed to be introduced (the reading of coded and uncoded text). The specification for *More than one letter sounds* is shown below. Under it there are separated segments for the sounds “ch” and “sh”, the sound “th”, the sound “ck” and the double l. Observe that the system will look for letter patterns instead of phonemes in this case.

1.3 - Advanced Code: Reading Coded Text, Reading Uncoded Text.

1.3.1 - More Than One Letter Sounds.

⁷Besides this new skill, this node inherits the skills of its ancestral node: *Code Knowledge*, *Segmenting*, *Blending*.

- 1.3.1.1 - Ch And Sh Sounds. % beach, bush, chalk, chap, chimp...
 {s,c}{h}{\$}{\$}. {s,c}{h}{\$}{\$}{\$}. {\$}{s,c}{h}{\$}.
 {\$}{s,c}{h}{\$}{\$}. {\$}{\$}{s,c}{h}{\$}.
 {\$}{\$}{s,c}{h}. {\$}{\$}{\$}{s,c}{h}. {\$}{\$}{\$}{\$}{s,c}{h}.
- 1.3.1.2 - Th Sounds. % earth, math, month, them, they,...
 th\$. th\$\$\$. \$th\$. \$\$th. \$\$th\$. \$\$\$th.
- 1.3.1.3 - Ck Sounds. % clock, dock, lick, jack, luck,...
 \$\$ck. \$\$ck\$. \$\$\$ck.
- 1.3.1.4 - Double L Sounds. % ball, bell, chill, hill, shell,...
 \$ll. \$ll\$. \$\$ll. \$\$ll\$. \$\$\$ll. \$\$\$ll\$.

In the next sub-division of *Advanced Code* the system is looking for multiple representation for a sound. In the code below one can see the specification for patterns of phonemes that may be represented in many different ways, such as the sounds /ou, /au, /e@, /r, /ii, and so on. Words with these sounds will be selected and stored in separated groups.

1.3.2 - Multiple Representation For A Sound: Homonymous Processing.

1.3.2.1 - Ou Sounds. % for boat, note, cone, snow, etc.

{/ou}{*}. {X}{/ou}{*}. {X}{X}{/ou}. {X}{X}{/ou}{*}. {*}{/ou}.

1.3.2.2 - Au Sounds. % for how owl, cow, town, house, ground, etc.

{/au}{*}. {X}{/au}{*}. {X}{X}{/au}{*}. {*}{/au}.

1.3.2.3 - Er Sounds. % for her, girl, shirt, burn, bird,...

{/e@}{*}. {X}{/e@}{*}. {X}{X}{/e@}{*}. {*}{/e@}.

1.3.2.4 - R Sounds. % for rat, wreck, ride, wrap, trap, risk...

{/r}{X}. {/r}{X}{X}. {/r}{X}{X}{X}. {X}{/r}{X}.
 {X}{/r}{X}{X}. {X}{X}{/r}{X}.

1.3.2.5 - Ee Sounds. % for see, east, chief, knee, leave...

{/ii}{X}. {/ii}{X}{X}. {/ii}{X}{X}{X}. {X}{/ii}{X}.
 {X}{/ii}{X}{X}. {X}{X}{/ii}{X}. {X}{/ii}.
 {X}{X}{/ii}. {X}{X}{X}{/ii}.

1.3.2.6 - Ei Sounds. % for ate, mate, flakes, etc.

\$a\$e. \$a\$e\$. \$\$a\$e. \$\$a\$e\$.
 {/ei}{*}. {X}{/ei}{X}. {X}{/ei}{X}{X}. {X}{X}{/ei}{X}.
 {X}{X}{/ei}{X}{X}. {*}{/ei}.

1.3.2.7 - Oo Sounds. % for put, cut, etc.

{/u}{*}. {X}{/u}{X}. {X}{/u}{X}{X}. {X}{X}{/u}{*}. {*}{/u}.

1.3.2.8 - U Sounds. % for poor, book, ...

{/uh}{*}. {X}{/uh}{X}. {X}{X}{/uh}{X}. {*}{/uh}.

1.3.2.9 - E Sounds. % for bed, bet, ...

{/e}{X}. {/e}{X}{X}. {/e}{X}{X}{X}. {X}{/e}{X}.

{X}{/e}{X}{X}. {X}{X}{/e}{X}. {*}{/e}.

1.3.2.10 - Ai Sounds. % my, sky, idle, ...

{/ai}{X}. {/ai}{X}{X}. {X}{/ai}. {X}{/ai}{X}.

{X}{X}{/ai}. {X}{X}{/ai}{X}. {*}{/ai}.

1.3.2.11 - S Sounds. % sorry, ask, ...

{/s}{X}. {/s}{X}{X}. {/s}{X}{X}{X}. {X}{X}{/s}{X}. {X}{X}{/s}.

1.3.2.12 - Z Sounds. % zero, ...

{/z}{*}. {X}{/z}{*}. {X}{X}{/z}{*}. {*}{/z}.

1.3.2.13 - L Sounds. % life, floor, ...

{/l}{X}. {/l}{X}{X}. {/l}{X}{X}{X}. {X}{/l}{X}. {X}{/l}{X}{X}.

{X}{X}{/l}. {X}{X}{X}{/l}.

Next, the multiple sounds for a single letter pattern are gathered together, through the code shown below. In “words with *ow*” there would come words such as *show, throw, cow, tow, now, know*, etc. The “words with *a*” part highlights the variations of the letter *a* and the *a-e* issue, as in *fate, made, mate* and so on. Finally, “words with *ea*” brings in words like *eat, meat, bread, earn*, etc.

1.3.3 - Multiple Sounds For One Representation: Nonsense Word Spelling.

1.3.3.1 - Words With Ow. % how, show, owner, allow...

{o}{w}{*}. {\$}{o}{w}{*}. {\$}{\$}{o}{w}{*}. {\$}{\$}{\$}{o}{w}{*}.

1.3.3.2 - Words With A. % at, ate, mat, mate ...

{/a}{X}. {/a}{X}{X}. {/ei}{X}. {/ei}{X}{X}. {X}{/a}{X}.

{X}{/ei}{X}. {X}{/a}{X}{X}. {X}{/ei}{X}{X}.

1.3.3.3 - Words With Ea. % eat, bread, year, early ...

{e}{a}{*}. {\$}{e}{a}{*}. {\$}{\$}{e}{a}{*}.

The last and most advanced group deal with more complex multi-syllable words, and is designed to develop the ability to deal with word splitting and to recognise and process common endings and high frequency non regular words. The code presented below shows how to isolate different endings such as the ones for “ate” verbs (generate, elucidate), adjectives (American, Brazilian, rosaceous, eatable, vibrant), regular simple past tenses (ending with “ed”), words related to “action” (explanation, suspension, migration, invasion, persuasion), gerunds (going, writing), common nouns and adjectives ending with “ant” or “ent” (commitment, investment, instant, ignorant, diligent), nouns ending with “ism” (impe-

rialism, journalism, egoism) and adverbs ending with “ly” (separately, seriously, definitely).

1.4 - Multi-Syllable: Reading Coded Text, Word Splitting.

1.4.1 - Chunking. % pudding, river, coffee, happily, ..

XXX/i/ng. XXX/@/r. XXX/o/n. XXX/ii. XXXX/l/ii.

XXX/p/l. XX/ei/n. XXX/ei/n.

1.4.2 - Special Endings: High Frequency Words Processing.

1.4.2.1 - Ate Verbs.

*/ei/t. % as generate, elucidate...

1.4.2.2 - Adjectives.

*/@/n. % as in American

*/@/n/z. % as in Brazilians

*/ii/@/s. % as in rosaceous

*/@/b/l. % as in eatable

*/@/n/t. % as in vibrant

1.4.2.3 - Past Tense. % packed, scared, worried

*/t/i/d. */r/i/d. */z/d. */ch/t.

\$\$\$ed. \$\$\$\$ed. \$\$\$\$\$ed. \$\$\$\$\$\$ed.

1.4.2.4 - Action. % election, explanation,

*/sh/@/n.

1.4.2.5 - Gerund/Present Continuous.

*/i/ng. % going, pressing

1.4.2.6 - Common Nouns And Adjectives.

*/i/z/@/m. */@/n/t. % absent, accident, nationalism

1.4.2.7 - Ly Adverbs. % happily, simply

*/l/ii.

The assemblage of all these parts form the written representation of the knowledge tree for the Phono-Graphix system. The complete code has been used to distribute the words of the four thematic selections described in Chapter 4 into the segments of the subject domain. The results are shown and compared with the other literacy methods in Chapter 6.

5.3.2 The instantiation of Jolly Phonics

Jolly Phonics is a literacy teaching method that has been developed at Woods Loke Primary School in Lowestoft, Suffolk, England, for twenty five years, and is based on real classroom experience. Today it is a method that strongly relies on

PA for starting up the learning process, in a fashion very similar to the Phono-Graphix method. In its Third Edition published in February 1998 it presents a daring timetable of nine weeks to cover all the basic PA issues. Its presentation documents and tapes report very positive results of research on the effectiveness of the method [Lloyd, 1998].

In spite of some differences in the order of teaching and in the use of the symbols for representing the sounds, Jolly Phonics is in essence very similar to Phono-Graphix. Its designers have divided the subject domain into seven groups of six sounds according to their phonological complexity. These groups are to be systematically studied in a nine week period, after which the students are expected to be able to read most English words. During this period they would develop the five basic skills of the method: *letter recognition*, *letter formation*, *blending*, *identifying sounds in words* and *dealing with tricky words*. After that the students would move on to more advanced skills, such as reading books, story creation and writing, etc.

The seven sound groups are:

1. s, a, t, i, p, n
2. ck, e, h, r, m, d
3. g, o, u, l, f, b
4. ai, j, oa, ie, ee, or
5. z, w, ng, v, little oo, long oo
6. y, x, ch, sh, voiced th, unvoiced th
7. qu, ou, oi, ue, er, ar

More precisely, the *letter recognition* skill involves the recognition of the *sounds* and the association with the graphic symbols used to represent them. The *letter formation* skill is related to the forms and dynamics of handwriting, as well as the handling of the pencil.⁸ The *blending* skill is related to the coding and decoding of the words - the kernel of the cognitive process for reading and writing. The *identifying sounds in words* skill in a way precedes the *blending* skill, for it is related to the phonological awareness needed to be able to read and write. Finally, the *dealing with tricky words* skill is designed to offer contact with alternative ways

⁸The support for the development of this skill is not contemplated in this thesis.

to deal with irregular words, many of which “should be learned as whole words”, as the designers assert.

The sound groups are intended to be studied in the order defined above, although the re-use of the sounds already mastered with the new ones is recommended. The activities for the development of the different skills, however, have no specific order and are present in the study of almost all sound groups. The Jolly Phonics package is accompanied by plenty of material to be stuck on the wall and photo-copied, and by a complete set of suggestions for teaching activities.

The application of the knowledge tree concept in order to distribute the vocabulary through the curriculum and according to the skills to be developed is straightforward. The complete knowledge tree code to support the nine-week “Timetable for first term with Jolly Phonics” is shown in Appendix D.2.⁹ The analysis of the code and of its expected outcomes follows.

The knowledge tree has been divided primarily by the sound groups, most of them used to develop the skills for *letter recognition*, *blending* and *identifying sounds*. This is the specification for the first sound group:¹⁰

% Jolly Phonics Knowledge Tree

1 - Reading.

1.1 - Satipn Sounds: Letter Recognition, Blending, Identifying Sounds.

&G={/s,/a,/t,/i,/p,/n}.

1.1.1 - First Satipn Words.

GG. GGG. % 2 and 3 phonemes only: as, at, sap, pin, ant, etc.

1.1.2 - More Satipn Words. % with 4, and 5 phonemes.

GGGG. % pant, spin, tips...

GGGGG. % pants, spins

As these are the very first sounds to be formally taught to the students it is desirable that the words are formed only by these same sounds. In *1.1.1 - First Satipn Sounds* then would only be words with two or three “satipn” sounds (e.g. *at, it, sat, pan, pin*). The next group *1.1.2 - More Satipn Words* would contain words of four and five “satipn” sounds.

⁹As in the case of Phono-Graphix, this representation of Jolly Phonics is an adaptation of the method for the purposes of this thesis and is not explicitly stated in any of the Jolly Phonics documents.

¹⁰In KTDL (Knowledge Tree Definition Language), along with the standard symbols used to identify vowels (V), consonants (C) and any sounds (X), it is for the user to define the variables G, H, I and J, to represent any set of sounds required. For this first group the definition &G={/s,/a,/t,/i,/p,/n} can be further referred to for representing these sounds. The reference GGG, in this context means “a word of three sounds all of them belonging to the set {/s,/a,/t,/i,/p,/n}” (for example, *sit*). The variables G, H, I and J can be redefined in the same knowledge tree for representing different sound groups for other curriculum items.

The next block deals with the consonants /c, /h, /r, /m and /d, and with the vowel /e. However, the previous sounds are also included in the study:

1.2 - Chrmd Sounds: Letter Recognition, Blending, Identifying Sounds.

1.2.1 - First Chrmd Words. % with the vowel /e

&G={/c,/e,/h,/r,/m,/d}.

GGG. % red, deck, hem

1.2.2 - More Chrmd Words. % with the already known vowels

&G={/c,/e,/h,/r,/m,/d,/a,/i}. % hid, him, kick, mad, ...

GG. GGG. GGGG. GGGGG. GGGGGG. % from 2 to 6 phonemes.

1.2.3 - Satipn Plus Chrmd Sounds. % all the already known sounds.

&G={/c,/e,/h,/r,/m,/d,/s,/a,/t,/i,/p,/n }.

GG. GGG. GGGG. GGGGG. GGGGGG. % damp, crisp, drip, end...

The 1.2.1 - *First Chrmd Words* group would contain only the words with those new sounds. The groups 1.2.2 and 1.2.3 combine also the previous vowels and all the previous sounds, respectively.

The next block is quite similar to the previous one, this time in relation to the sounds /g, /o, /uh, /l, /f, and /b. In the group 1.3.3, as the availability of vowels has increased substantially, the words have been limited to CVC words - observe that the variable &G contains only consonants and the variable &H contains only vowels. Consequently, the requirement of *GHG* words is equivalent to the requirement of CVC words with just these particular consonants and vowels.

1.3 - Goulfb Sounds: Letter Recognition, Blending, Identifying Sounds.

1.3.1 - First Goulfb Words. % log, bulb, fog, of...

&G={/g,/o,/uh,/l,/f,/b}. % Jolly Phonics: g o u l f b

GG. GGG. GGGG. % with two, three and four phonemes only.

1.3.2 - More Goulfb Words. % with all the known vowels

&G={/g,/o,/uh,/l,/f,/b,/e,/a,/i}. % big, fill, bill...

GG. GGG. GGGG. GGGGG. GGGGGG.

1.3.3 - CVC Goulfb Plus Chrmd Plus Satipn Words.

&G={/g,/o,/uh,/l,/f,/b,/s,/t,/p,/n,/c,/h,/r,/m,/d}.

&H={/a,/i,/e,/o,/uh}.

GHG. % cut, come, top, hack, head, sick...

The same approach is adopted in the next block, where VC and CVC words are grouped in 1.4.1 whereas CCVC and CVCC words are gathered together in 1.4.2:

1.4 - Aioa Sounds:Letter Recognition,Blending,Identifying Sounds.
&H={/ei,/j,/ou,/ai,/ii,/oo}.*(in Jolly Ph: ai j oa ie ee or)
&G={/s,/t,/p,/n,/c,/h,/r,/m,/d,/j}.

1.4.1 - VC And CVC Aioa Plus Known Consonants.

HG. GHG. % eat, ice, deep, course, known,...

1.4.2 - CCVC And CVCC Aioa Plus Known Consonants.

GGHG. GHGG. % trade, cream, dates, kind, stone,...

For the block of words with the sounds /z,/w,/ng,/v and the “short and long oo” the approach is similar, but in 1.5.3 the words ending with the consonants and vowels of the group appear. It has already been noted that nothing prevents a word from figuring in different groups.

1.5 - Zwng Sounds.

&G={/z,/w,/ng,/v}. % (in Jolly Pho: z w ng v lit oo long oo)

&H={/u,/uu,/a,/i,/e,/o,/uh,/ei,/ou,/ai,/ii,/oo}.

&I={/u,/uu,/a,/i,/e,/o,/uh,/ei,/ou,/ai,/ii,/oo,
/z,/w,/ng,/v}.

1.5.1 - VC And CVC Zwng Sounds.

HG. GHG. % as, eyes, was,...

1.5.2 - Other Zwng Sounds.

III. IIII. IIIII. IIIIII. IIIIIII. % wives, ways

1.5.3 - Words Ending With Zwng Sounds.

{*}{H}{H}{G}. % always, buying...

{*}{G}{H}{G}. % bruises, causes...

In the next two blocks, presented together, the approach is again similar. In the curriculum items 1.6.3 and 1.7.3 there are specifications for the words that start with the sounds of the respective groups.

1.6 - Yxchshth Sounds.

&G={/y,/x,/ch,/sh,/dh,/th}.*JoPho:y x ch sh voi th unv th

&H={/u,/uu,/a,/i,/e,/o,/uh,/ei,/ou,/ai,/ii,/oo}.

&I={/u,/uu,/a,/i,/e,/o,/uh,/ei,/ou,/ai,/ii,/oo,/y,
/x,/ch,/sh,/dh,/th}.

1.6.1 - VC And CVC And CV Yxchshth Sounds.

HG. GHG. GH. % ash, axe, each, she, thy, you ...

1.6.2 - Words Ending With Yxchshth Sounds.

{*}{I}{G}. % books, british, brush, death ...

1.6.3 - Words Starting With Yxchshth Sounds.

{G}{I}{*}. % think, cheap, cherry, shut, young ...

1.7 - Quou Sounds.

&G={/k,/w,/au,/oi,/y,/uu,/@@,/aa}. %JP: qu ou oi ue er ar

&H={/u,/uu,/a,/i,/e,/o,/uh,/ei,/ou,/ai,/ii,/oo,/y}.

&I={/u,/uu,/a,/i,/e,/o,/uh,/ei,/ou,/ai,/ii,/oo,/y,
/@@,/aa,/oi,/au}.

1.7.1 - VC And CVC And CV Quou Sounds.

HG. GHG. GH. % we, why, wick, you ...

1.7.2 - Words Ending With Quou Sounds.

{*}{I}{G}. % back, aquatic, avenue, value

1.7.3 - Words Starting With Quou Sounds.

{G}{I}{X}{X}. {G}{I}{X}. % worse, young, keep...

The next block is dedicated to the study of alternative spellings. Each group concentrates on a different sound that has more than one graphic representation.

1.8 - Alternative Spellings.

1.8.1 - Long A. % rain, date, day

{C}{/ei}. {C}{/ei}{C}.

1.8.2 - Long E. % seed, read

{C}{/ii}. {C}{/ii}{C}.

1.8.3 - Long I. % pie, pipe, night, my

{C}{/ai}. {C}{/ai}{C}.

1.8.4 - Long O. % boat, bone, snow

{C}{/ou}. {C}{/ou}{C}. {C}{C}{/ou}. {C}{V}{C}{/ou}.

1.8.5 - Long U. % due, cube, few

{C}{/y}{/uu}{*}. {/y}{/uu}{*}.

1.8.6 - Little OO. % book, put

{C}{/u}{C}.

1.8.7 - Long OO. % moon, glue, blew, June

{C}{/uu}{C}.

1.8.8 - ER Sound. % bird, turn

{*}{/@@}. {C}{/@@}{C}.

1.8.9 - OR Sound. % fork, pause, claw, talk

{C}{/oo}. {C}{/oo}{C}.

1.8.10 - OI Sound. % oil, boy

{C}{/oi}. {C}{/oi}{C}.

1.8.11 - OU Sound. % loud, cow

{C}{/ou}. {C}{/ou}{C}.

The following group is specifically aimed at support for blending training, concentrating on blending in the beginning and in the end of the words:

1.9 - Blending.

1.9.1 - Initial Consonant Blending.

1.9.1.1 - Xl Blending. % words beginning with bl cl fl pl sl

{/b,/c,/f,/p,/s}{/l}{*}.

1.9.1.2 - Xr Blending. % br cr dr fr gr pr tr shr thr

{/b,/c,/d,/f,/g,/p,/t,/sh,/th}{/r}{*}.

1.9.1.3 - Sx Blending. % words beginning with st sc sm sn

{/s}{/t,/c,/m,/n}{*}.

1.9.1.4 - TW Blending. % words beginning with tw

{/t}{/w}{*}.

1.9.1.5 - Sxr Blending. % words beginning with scr spr str

{/s}{/c,/p,/t}{/r}{*}.

1.9.2 - Final Consonant Blending.

1.9.2.1 - Lx Blending. % words ending with lb ld lf lm ln lp

{*}{/l}{/b,/d,/f,/m,/n,/p}.

1.9.2.2 - Xt Blending. % words ending with ct ft nt pt xt

{*}{/c,/f,/n,/p,/x}{/t}.

1.9.2.3 - MPandNd Blendings. % words ending with mp and nd

{*}{/m}{/p}. {*}{/n}{/d}.

The next block is to select words to support the phonological awareness training on rhymes (identifying the sounds in words). It contains some of the common terminations used in rhymes.¹¹

1.10 - Rhymes.

1.10.1 - Ink Rhymes.

{*}{/i}{/ng}{/k}. % ink, pink, think ...

1.10.2 - Ice Rhymes.

¹¹It must be observed that not all the rhyming possibilities are covered by the specifications in this block of words. For the application of the methods to a particular theme it is always possible to insert new specifications to address different rhymes, whenever appropriate.

5.3.3 The instantiation of THRASS

The third PA method used to test the proposals of this thesis is the THRASS Special Needs Pack. THRASS stands for “Teaching Handwriting, Reading and Spelling Skills”, and has been developed by Alan Davies and Denyse Ritchie [Davies and Ritchie, 1996] in England and Australia. Although the product has received additions in recent years, including a multi-media computer program (called THRASS-it), the authors do not consider it a complete literacy teaching programme, but a “pack specifically designed for 7-11 years old who have difficulties with handwriting, reading and/or spelling”. They add that “THRASS is a support programme, not *the* programme. It is assumed that the children also take part, for much of the time, in the normal writing and reading experiences”.

The method relies on the ability of the student to master the phonemes of the English language and to understand their relationship to the corresponding graphemes which represent these sounds on paper. The authors have developed a table called the “ThrassChart” (Figure 5.3) where all the 44 English phonemes are organised in two sets (20 vowel phonemes and 24 consonant phonemes) and related with their possible graphic representations, the graphemes. In addition to that they defined the “ThrassWords”, which is a set of 120 words chosen to illustrate the representation of the phonemes on paper.

This referential frame is used in all the activities of this very synthetic method. The activities are designed to develop the main skills related to basic literacy: handwriting, reading and spelling. Additionally there are evaluation activities which can be applied before starting to use the programme, and later on, to assess the student’s progress.

According to the method the activities for developing all the skills should be followed in the order the phonemes appear in the ThrassChart (Figure 5.3).

The description of THRASS curriculum with the methods proposed in this thesis is straightforward: the knowledge tree for THRASS has two basic divisions, one for the consonant phonemes and one for the vowel phonemes, and each one of these divisions contains the specifications for capturing words that are phonologically equivalent to the ThrassWords for each phoneme. The complete code for the knowledge tree for THRASS is presented in Appendix D.3. Parts of the tree with comments on the specifications follow.

The representation of the tree for THRASS in KTDL (Knowledge Tree Definition Language) is extensive and repetitive. The five first consonant phonemes are shown below. Observe the presence of the ThrassWords to the right of some of the specifications, after the comment (%), followed by an asterisk. This represen-

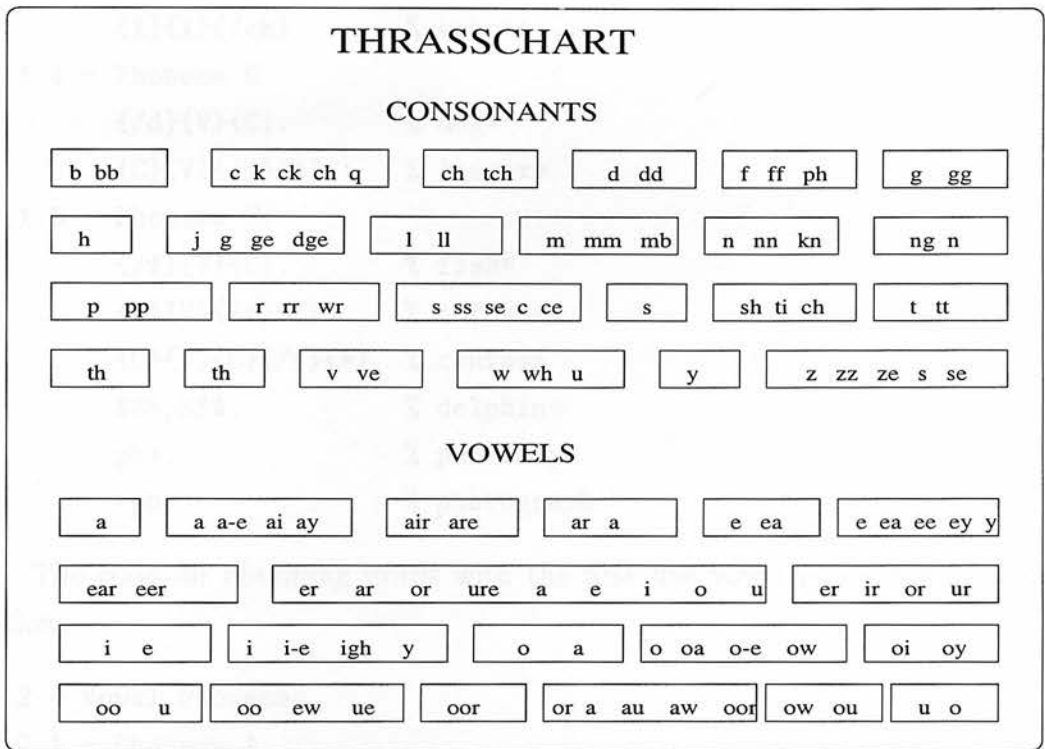


Figure 5.3: The THRASSCHART

tation is designed to capture not only the *thrasswords*, but some equally simple equivalent words using the same phonemes.

% THRASS Knowledge Tree

1 - Literacy Teaching: Handwriting, Reading, Spelling, Assessing.

1.1 - Consonant Phonemes.

1.1.1 - Phoneme B.

```
{/b}{V}{C}.      % bird*
b$$              % bat
$bb{*}.         % abbey
$$bb{*}.        % rabbit*
```

1.1.2 - Phoneme C.

```
{/k}{V}{C}.      % cat*
{/k}{V}{C}{C}.   % kitten*
{C}{V}{/k}.      % duck
{C}{/k}{V}{C}.   % school
{/k}{V}{V}{C}.  % queen
```

1.1.3 - Phoneme Ch.

```
{/ch}{V}{C}.    % chair
```

	{X}{X}{/ch}.	% watch*
1.1.4 - Phoneme D.		
	{/d}{V}{C}.	% dog*
	{C}{V}{/d}{X}.	% ladder*
1.1.5 - Phoneme F.		
	{/f}{V}{C}.	% fish*
	{C}{V}{/f}{C}.	% coffee*
	{C}{V}{C}{/f}{*}.	% confess
	\$\$\$ph\$\$.	% dolphin*
	ph*.	% pharmacy
	*ph.	% photograph

The code for obtaining words with the first five vowel phonemes is shown below.

1.2 - Vowel Phonemes.

1.2.1 - Phoneme A.

{/a}{C}{X}.	% ant*
{C}{/a}{C}.	% cat

1.2.2 - Phoneme Ei.

{C}{/ei}{C}.	% tape*
{C}{/ei}{C}{C}.	% baby*
{C}{C}{/ei}{C}.	% snail*
{C}{C}{/ei}.	% tray*

1.2.3 - Phoneme Air.

{X}{/e@}{*}.	% chair
{*}{/e@}.	% hair*
{*}{/e@}{/r}.	% square*

1.2.4 - Phoneme Aa.

{C}{/aa}{C}.	% car*
{C}{V}{C}{/aa}{*}.	% banana*

1.2.5 - Phoneme E.

{C}{/e}{C}.	% bed*
{C}{C}{/e}{*}.	% bread*

In the same way as with Phono-Graphix and Jolly Phonics, the knowledge tree for THRASS has been used to distribute the words of the thematic selections described in Section 4.2 into the curriculum items. The results for the different methods are discussed in the next chapter.

Chapter 6

Analysis of the resultant vocabularies

6.1 Setting up the environment

The methods designed to select words thematically in this research allow a large range of alternatives to choose from. There is a clear compromise between the size of the vocabulary and the percentage of relevance to the theme in focus. Larger vocabularies would generally contain a larger number of irrelevant words to the target theme. Small vocabularies would be more focused, but sometimes they might just not have words enough to cover all the phonological requirements of some of the curriculum items of the teaching method in use. As already stressed, the refinement of the seed set and the appropriate choice of parameters can improve the quality of the vocabulary.

In order to support the claim that it is possible to obtain thematic vocabularies that can cover the requirements of phonological awareness approaches for teaching to read, ten different thematic vocabularies obtained after selections by the methods described in Chapter 4 were considered in conjunction with the three PA methods examined in this thesis:

1. *Food only* - the selection with 2 seeds without repetition using the seed set of 44 seeds for *food* (line 5 of table 4.5) produced 765 words. The words were examined and 163 words considered out of context were eliminated, resulting in 602 words in context.¹ The resulting set of words was expanded with its derivative words (plural, verb tenses, related nouns or adjectives, antonyms and synonyms), producing a vocabulary of 1179 words related to food. This and the other thematic vocabularies referred to in this chapter

¹This is a very reassuring result for the estimation process for it had been estimated that the set would contain 596 words in context, a very close figure.

can be seen in Appendix E.

2. *Sea only* - the selection with 2 seeds without repetition using the seed set of 47 seeds for *sea* (Table 4.2) has produced 995 supposedly thematically relevant words. After a closer examination 344 were excluded manually, resulting in a base vocabulary of 651 words. The expansion of this vocabulary with its derivative words resulted in a thematic vocabulary of 1435 words (see Appendix E).
3. *Agri only* - the selection with 3 seeds without repetition and without examples, using the seed set of 52 seeds for *agriculture* (Table 4.3) produced 974 words. An individual examination of the words took place and 243 words were excluded manually. The expansion of the base vocabulary with its derivative words resulted in vocabulary of 1516 words related to agriculture (Appendix E).
4. *Freire only* - the selection with 3 seeds without repetition using the seed set of 37 seeds for *The city of Ceilandia* (Table 4.11) produced 826 words. After the manual exclusion of 164 words considered out of context, the base vocabulary was expanded to 1633 words with the derivative words (Appendix E).
5. *Common words* - as explained in Section 4.1, there is an option to exclude the most frequent words from the selection process and add them to the thematic vocabularies whenever needed. This vocabulary of common words was obtained as a compilation of the most frequent words (supposedly the most common words). The common words vocabulary from [Pain, 1985] was also taken in account. The result is a base vocabulary of 768 words which was expanded to 1406 words when the derivations were included. It can be seen in Appendix E.
6. *Food&Common* - This is the product of the union of the *Food only* and the *Common words* vocabularies. It is 2767 words long after the addition of the derived words.
7. *Sea&Common* - The product of the union of the *Sea only* and the *Common words* vocabularies. It is 3030 words long.
8. *Agri&Common* - The product of the union of the *Agri only* and the *Common words* vocabularies. It is 3113 words long.

9. *Freire&Common* - The product of the union of the *Freire only* and the *Common words* vocabularies. It is 3225 words long.
10. *Tricky* - This vocabulary has been specially created to address the THRASS method. It consists of manually chosen words that do not behave as regular words, according to the method. It can be seen in Appendix E.

The PA approaches are generally more concerned with the quality of, rather than with the quantity of, words to be used as stimuli for teaching reading. Among the three methods analysed in this research, THRASS is the one in which this issue is more evident. THRASS basically works with a few samples of each feature or phoneme, trusting that the student will be able to generalise their learning to other words with similar features. Jolly Phonics is also economic with respect to the quantity of words used as stimulus for teaching. Only Phono-Graphix has a vocabulary that could be considered large.

Some phonemic features are abundantly provided with words to be used as examples (e.g. /s, /c, /p), whereas some others have very few words (e.g. /y or /k/s). At the very beginning of the learning process it is likely that neither too few nor too many words should be available as stimulus for teaching in each of the relevant parts of the curriculum. The approach pursued for the allocation of words to the curricula in this research is to try to get enough words from the strictly thematic vocabularies initially, and if this is not possible, then try to get them from the vocabularies extended with the common words. The *Source* command of KTDL allows the indication of the source of the words either globally, to the whole curriculum, or locally, overriding the global option for a particular curriculum item. This facilitates the application of the allocation strategy (the use of the *Source* command can be observed in the KTDL knowledge trees in appendices D.1, D.2 and D.3).

6.2 The vocabularies for Phono-Graphix

Table 6.1 compares the quantities of words obtained with the methods and strategies described in this thesis for the four different themes with the ones originally used in Phono-Graphix. The words are distributed according to the curriculum as modelled in this research. In the table the free figures are the amount of words obtained from the purely thematic vocabularies and the figures between brackets, when present, are the numbers of words obtained with the thematic vocabularies *plus* the common words vocabulary.

Curriculum Item	Original words	Food words	Sea words	Agri words	Freire words
1.1	331	108	160	163	157
1.1.1	20	12 (12)	3 (9)	9	9
1.1.2	31	1 (8)	1 (8)	7	1 (8)
1.1.3	52	7	9	7	12
1.2.1	16	5 (22)	7 (24)	3 (21)	1 (19)
1.2.2	91	115	157	159	165
1.2.3	117	49	98	94	91
1.3.1.1	42	28	36	40	34
1.3.1.2	14	4 (27)	4 (27)	7	4 (27)
1.3.1.3	27	13	23	24	28
1.3.1.4	13	10	13	15	12
1.3.2.1	55	66	64	67	57
1.3.2.2	32	11	31	35	28
1.3.2.3	4	13	3 (21)	17	12
1.3.2.4	107	52	79	91	94
1.3.2.5	71	64	41	60	40
1.3.2.6	67	46	68	71	77
1.3.2.7	14	8	7	8	5 (31)
1.3.2.8	63	30	28	34	35
1.3.2.9	68	16	37	24	45
1.3.2.10	45	16	25	26	26
1.3.2.11	126	57	80	76	89
1.3.2.12	33	123	187	218	180
1.3.2.13	136	43	92	84	79
1.3.3.1	22	10	23	22	27
1.3.3.2	100	45	62	73	79
1.3.3.3	53	60	67	62	38
1.4.1	64	85	95	99	103
1.4.2.1	17	7	11	17	21
1.4.2.2	28	44	52	50	82
1.4.2.3	6	87	114	87	134
1.4.2.4	12	6 (15)	11	14	27
1.4.2.5	5	101	147	123	179
1.4.2.6	11	8	14	6 (11)	26
1.4.2.7	7	8	5 (25)	6 (26)	2 (22)

Table 6.1: Distribution of thematic words in the Phono-Graphix curriculum

The distribution of the thematic words across the Phono-Graphix curriculum resulted in quite acceptable numbers. In 87.9% of the items only the thematic vocabularies have been used, whereas in the other 12.1% the thematic vocabularies have been complemented with the common words vocabulary. With this approach not a single item of the curriculum has got less than 7 words allocated to it. In order to evaluate how acceptable it is, observe that Phono-Graphix curriculum item 1.1.1 - Fat Cat Sounds, for example, counts originally with 20 words not particularly related to any theme, with three of them presented as proper names (Pat, Sam and Tom). Curriculum item 1.3.2.3, for the “Er Sounds” (*girl, burn, bird*) has only 4 representatives in Phono-Graphix vocabulary, whereas item 1.4.2.5, related to words ending with “ing” has just 5 elements. This also indicates that the main requirement in this case is the existence of enough words to illustrate the phonological features in focus, instead of the availability of a big amount of words with those properties.

Another important remark is that the original Phono-Graphix vocabulary (Appendix E) used as reference is a large one, supposedly suitable for children. However, in this vocabulary it is possible to find words that might be inappropriate to be used as stimuli to teach children, such as absolute, advance, aim (6.4),² anxious, bitterly, bountiful, budget (9.2), cashier, chlorine (10.85), containment, debt, dementia, establish, fictitious, gadget, gravity, hesitate (9.6), inertia, militia, momentous, observer (9), obstacle, petite, precocious, publicity, refrigerator, saturate, structure, suspension, symbolic, torment (8.7), uneasy, and variation, just to mention a few. Besides that, Phono-Graphix vocabulary contains 26 words which due to their low frequency of use would not be selected for the final vocabulary according to the criteria adopted in this research.

Table 6.2 shows examples of the distribution of words across the curriculum items of Phono-Graphix for each of the sources used: in the first column there are the words figuring in the original method and the following ones contain the thematic words for *food, sea, agriculture* and the Freire’s Generative Themes for the city of *Ceilandia*.

Table 6.2 is useful to clarify the idea of keeping the common words (that is, the ones that are useful in any context) in an isolated file, to have the option of adding them to the vocabulary whenever the thematic vocabulary does not cover the needs of a particular curriculum item of the programme. Due to the very strict phonological requirements for the items 1.1.1 (Fat Cat Sounds), 1.1.2 (Bug

²For the words where there is a number in brackets in this list, it represents the *Age-of-Acquisition* parameter of the Medical Council Research dictionary (MCR-II), in years (and fraction of year).

Item	Original words	Food words	Sea words	Agri words	Freire words
1.1.1	at cap cat cop cot fat map mat pop pass pat pot sam sop sot tap tom top tot mop	arm art at top carp fat mass part pass pop pot tart	arm art at top calm mass part pass toss	top calm cart farm fat moss palm pot sap	calm cap cart cat farm map mat tap tart
1.1.2	bad bag bid big bud bug dad did dig dim dug gym had hag hid hug hum jam etc.	bad bag big dad did had him	add bag big dad did had him tug	bag bud dam dig dug hid mud	add bad bif dad did had him jam
1.1.3	ate bell ben bet bib bin bit bun but buzz fib fin fun fuzz hill his hit hop hot in is etc ...	bib bun hen lip nut sup whip	bit ebb fin hop hull knot net wet whip	ben bit hull bill knot nut sell	bill bit fit hell net nut ill wet whip win won
1.2.1	act alp amp and ant apple asp eagle east east elf elk elm end inp ink oink...	able act and ant apple arms ask eats eaten iced its oats often oils old oven...	able act aft adds and arms armed auks east eats ebbs even inch isles often oils old oozed...	able act and ant apple arms armed ask axe east eats eaten else even its old...	able act adds and ant arms armed axe eats eaten else even its often oils old...
1.2.2	bend boast box bump camp castle cattle cent certain child chimp coast count damp..	baits bakes baked bards beans beefs bind bound boards bowls bulb buns burns...	baits bank basin bathes belt bilge bitten boards boats bones cable capes coast decks..	barks basin beans beats beaten bent bites bitten box bison bulb bulk bulls burns	band bank bets bible bill births board bond bottle bowls bounce box cakes chance...
1.2.3	black blip block blop blouse brain bread break breeze brick clam clean...	bran broth clam clean cream crop cube curry dairy freeze fresh fries ..	bleak bright bluff brace bridge brit claws crab crack crash drag ferry...	carry berry block bloom bran breed broth bruise brush shrub cherry clone...	block brave breath breeze bruise brush bury claim cloud club tory flood...

Table 6.2: Examples of the distribution of words across Phono-Graphix curriculum items

on Jug Sounds) and 1.1.3 (Ben Bon Sounds) of Phono-Graphix programme, the system was able only to find a small number of words for them.

The addition of the common words to the few words of each of the thematic sets generated automatically resulted in a considerable degree of overlap in the thematic vocabularies for those items, because the majority of the words came from the common words set. It is worth commenting that in Phono-Graphix these first curriculum items are the first ones to be applied, in general, and that the use of well-known common words would represent no harm in terms of a thematic policy.

Moreover, the common words set so far established is experimental and obviously deserves reviews and additions that could reinforce its expressiveness without compromising the thematic purposes. Even many of words used in the Phono-Graphix method could be considered under more careful observation, to be added to the common words set (e.g. “cop”, “tot”, “hug” and so on).

For the other items exemplified in Table 6.2 (1.2.1, 1.2.2 and 1.2.3) there were plenty of words available in the thematic vocabularies and the addition of the common words was not necessary. The examples do not cover the complete range of words selected for each item, for reasons of space. However it can be confirmed that the selected words in each thematic set follow the same phonological patterns present in the original words. As the selection process is a mechanical one, as long as the phonological coding of the words, provided by the Festival system, is trusted³ and the expression of the requirements is correct, there is no doubt that the right words will be in the right positions in the *knowledge tree*.

6.3 The vocabularies for Jolly Phonics

In the same way, Table 6.3 compares the quantities of words obtained with the tools developed in this research for the four different themes with the ones originally used in Jolly Phonics. Table 6.4 contains examples of the original words for that method, compared with the ones generated by the system for the four different themes in focus. The same remarks and conventions for the two tables for Phono-Graphix hold for the tables for Jolly Phonics.

Table 6.3 demonstrates that it is also possible to generate vocabulary automatically to cover all the curriculum items of Jolly Phonics, and only in a few cases did the system generate fewer words than there are in the original vocabulary

³The Festival System has been used for providing the phonologic coding of the words, which it does fairly well and which it uses as the basis for its speech production. Obviously there are regional and accent issues to be observed, but in general, it produces a very acceptable coding.

Curriculum Item	Original words	Food words	Sea words	Agri words	Freire words
1.1.1	15	2 (11)	1 (10)	3 (12)	3 (12)
1.1.2	5	1 (4)	4 (7)	2 (5)	7
1.2.1	4	0 (4)	2 (6)	0 (4)	1 (5)
1.2.2	13	1 (11)	4 (14)	3 (13)	2 (12)
1.2.3	91	39	47	45	75
1.3.1	7	0 (2)	5 (6)	3 (4)	0 (1)1
1.3.2	15	4 (10)	9	6 (12)	3 (8)
1.3.3	76	15	16	20	24
1.4.1	38	18	23	27	21
1.4.2	17	18	17	19	25
1.5.1	4	1 (7)	4 (10)	0 (6)	2 (8)
1.5.2	2	1 (5)	6 (10)	0 (4)	2 (6)
1.5.3	4	20	30	31	49
1.6.1	3	2 (10)	2 (9)	3 (11)	3 (10)
1.6.2	34	33	55	48	45
1.6.3	50	47	36	56	66
1.7.1	8	5 (10)	5 (11)	3 (8)	2 (8)
1.7.2	51	33	49	35	42
1.7.3	64	36	45	45	42
1.8.1	23	11	20	20	20
1.8.2	17	18	11	22	11
1.8.3	21	7	14	13	15
1.8.4	24	17	22	22	18
1.8.5	8	15	7	13	13
1.8.6	12	2 (14)	2 (14)	3 (15)	1 (14)
1.8.7	11	6 (14)	4 (12)	8	13
1.8.8	13	4 (18)	6 (20)	10	6 (20)
1.8.9	23	9	22	14	12
1.8.10	11	2 (5)	2 (5)	1 (4)	1 (3)
1.8.11	20	12	16	13	11
1.9.1.1	44	53	74	76	64
1.9.1.2	64	87	143	171	154
1.9.1.3	41	67	95	85	85
1.9.1.4	3	1 (4)	0 (3)	2 (5)	1 (4)
1.9.1.5	7	19	27	36	30
1.9.2.1	8	10	16	15	18
1.9.2.2	27	33	48	34	60
1.9.2.3	32	14	28	22	26
1.10.1	4	1 (4)	1 (4)	2 (5)	1 (4)
1.10.2	3	5 (7)	0 (2)	3 (5)	2 (4)
1.10.3	3	1 (3)	0 (2)	1 (3)	1 (3)
1.10.4	11	6 (14)	9	7	6 (15)
1.10.5	9	3 (4)	4 (4)	5 (6)	4 (6)
1.11.1	58	58	58	58	58

Table 6.3: Distribution of thematic words in the Jolly Phonics curriculum

Item	Original words	Food words	Sea words	Agri words	Freire words
1.1.1	an ant in it nap nip pan pat pin pip pit sit tan tap tip	an ant at in it its pan sin sit tin	an ant at in it its sin sit tan sin sit	an ant at in it its pan pit sap	an ant at in it its pin sin sit tap tip
1.1.2	pant snap snip spin spit	ants pastas since sits	ants since sits spat spats spin spit	ants saps sapped since sits	ants sins sits snap snaps snapt spin spit taps tips
1.2.1	deck hem read red	dead head read red	dead deck head read red wreck	dead head read red	dead head read red wreck
1.2.2	dad deck did had ham hem hid him kick mad ram read red	add added dad dead did hack had head him read red	add added crack dad dead deck derrick did had head him read red wreck	add added dad dam dead did hack had head hid him read red	add added arid dad dead did had head him read red wreck
1.2.3	an ant camp can cap cat crisp dad damp deck dent did dip drip end...	can add ant assam crisp dipped dress ate excess hand hen sick ..	can act captain crack dead deck derrick dip drip hand head sick...	can act add ant axe axis cement chemist credit dead dip had...	can act add an and ant arid axe cap cat christen credit ...

Table 6.4: Examples of the distribution of words across Jolly Phonics curriculum items

(e.g. items 1.10.2 and 1.10.3 for the theme *sea*).

The distribution of words across the Jolly Phonics curriculum is much more balanced than the Phono-Graphix one and produced numbers and vocabularies equivalent to those originally provided with the programme. There is still some overlap of words, mainly because the themes in study do overlap to some extent. The handling of the *tricky words* might deserve a little more attention in a real implementation of the programme.

6.4 The vocabularies for THRASS

Finally, tables 6.5 and 6.6 contain the numbers and the examples of words from the original vocabulary for THRASS along with the results of the application of the proposed tools.

Due to the conciseness of the THRASS method, the figures obtained by the vocabulary tools in each case have largely surpassed the number of words available in the original vocabulary, and only in a few cases was it necessary to extend the thematic vocabularies with the common words to obtain satisfactory figures. Indeed, the initial vocabulary for THRASS is restricted to 120 words - the so called *Thrasswords*.⁴ This implies an average of less than three words per phoneme, for demonstrating the use of the 44 English phonemes and their possible graphical representations. The vocabulary of 268 words used in this research as the THRASS vocabulary has been obtained by scanning all the teaching stimuli available in the packet provided for the method.

It is probable that a real implementation of THRASS with the tools of this research would require the development of some special facilities to adapt the tools to the THRASS characteristics. At this moment, however, focusing only on the selection of words, the results of the distribution of thematic words across the THRASS curriculum are very encouraging.

⁴Words marked with an asterisk in Table 6.6 are the ones used as examples of the phonological features in the “Thrasswords” Table (see Sub-section 5.3.3 - The Instantiation of THRASS).

Curriculum Item	Original words	Food words	Sea words	Agri words	Freire words
1.1.1	11	19	34	39	20
1.1.2	16	31	41	34	35
1.1.3	5	10	12	12	15
1.1.4	5	12	22	25	19
1.1.5	6	16	23	19	24
1.1.6	8	11	24	21	11
1.1.7	4	14	19	21	21
1.1.8	7	10	22	13	8
1.1.9	7	13	28	23	30
1.1.10	9	17	17	31	22
1.1.11	5	6	6	2 (6)	8
1.1.12	2	7	10	3 (30)	10
1.1.13	2	21	20	26	14
1.1.14	3	23	36	39	42
1.1.15	7	12	17	22	15
1.1.16	2	3	5	11	4
1.1.17	4	10	20	32	42
1.1.18	4	11	10	4 (20)	10
1.1.19	2	7	3	11	8
1.1.20	3	2 (38)	6	10	8
1.1.21	5	20	15	33	35
1.1.22	4	11	25	17	18
1.1.23	4	39	15	36	47
1.1.24	9	18	27	34	31
1.2.1	7	7	14	13	16
1.2.2	18	17	34	33	39
1.2.3	3	14	4	14	15
1.2.4	4	11	13	18	14
1.2.5	13	23	40	25	41
1.2.6	22	60	45	56	44
1.2.7	3	6	9	8	8
1.2.8	17	68	68	82	51
1.2.9	5	4 (18)	6	10	6
1.2.10	8	24	20	26	26
1.2.11	3	8	17	15	15
1.2.12	10	4 (12)	26	18	9
1.2.13	10	22	29	39	28
1.2.14	3	6	5	2 (10)	6
1.2.15	2	17	9	13	9
1.2.16	5	25	12	27	40
1.2.17	2	9	9	9	15
1.2.18	7	9	22	14	12
1.2.19	4	4	8	10	4
1.2.20	13	12	18	24	24

Table 6.5: Distribution of thematic words in the THRASS curriculum

Item	Original words	Food words	Sea words	Agri words	Freire words
1.1.1	rabbit* ball beach bed bell bird* boat book box bull bus	bag bait bake bard base bean beef beet bib board boil booth bowl bun burn cabbage rubbery	bait bar barge base bash bass bay beach beam beat berth bite bit board boat bone...	bag bail balm bare bark barn base bath bean bear bore borne beat beech beet bite...	bar bear bet bill birth bite bit board bowl box burn bus rubbish booth boom..
1.1.2	book cage car cat* coin cold duck* fork king king kite kitten* school shark skate skip skull	bake cake cakes caked carp coat cone cook cooked cows cup curd fork kettle pick pork...	cable calm cape catch cave check coarse coast conch cork deck dock dyke hook jack lake...	bark calf calm cane card cart carton cone corn cost cotton cows kernel lake pack tick...	cake cable calm cap cart cat catch check coach cook cooked cool cup duck folk shack...
1.1.3	beach chair* cheese march watch*	char charred cheese choose poach touch	beach catch chart check ditch hatch inch pitch reach rich...	beech birch chaff cheap chill rich march patch search hitch	catch chase check cheer cheers chill chuck coach match choose.
1.2.1	and ant* cat jam lamb tap that	bag fat hack mash pan tab	bass catch gaff hatch gap jack lash pack pad rang...	axe bag dam ban fat mash pan patch sap pack...	axe cap cat catch hat shack jam lap lash match...
1.2.2	baby* blaze brain cage crane flame gate grain grape plate rain skate snail* snake spade tape* train tray*	bait bake base cake drain glaze grain great crepe plaice raid raise make save shake stale wait	bait base bathe bays brace cape cave freight lake mate navy safe sail save scale shake spade sway..	bail trade base blade cane save dahlia daisy date drain male flail freight gauge grape graze lake maize...	brave cake chase claim drain fade fail frame grace grey hate haze plain mail game great pain plate...
1.2.3	chair hair* square*	dairy fare fares pare paring pared pear pears prepare rare..	everywhere fare fares prepare	bare bear bearing dairy fair fairer pear pears square...	bear bearing declare lair fair fairer parent repair share square.

Table 6.6: Examples of the distribution of words across THRASS curriculum items

6.5 Final comments on vocabularies

Even though many improvements are still possible in the seed sets and in the choice of the parameters for selecting the words, the six tables shown in the last three sections demonstrate the viability of the combination of thematic and phonological awareness approaches for teaching reading in the context of vocabulary issues. Recall that the vocabularies chosen to represent the four different themes studied here have around half of the potentially thematically related words in the dictionary, and that the choice of a bigger set could have resulted in bigger and richer final thematic vocabularies. In addition, more complete dictionaries, and vocabularies specific to particular areas could be added to the project's dictionary in the future in order to increase the availability of words for the selection process.

Nonetheless, for the purpose of demonstrating that it is possible to obtain enough PA stimuli in thematic vocabularies, the results obtained were more than sufficient.

The tools developed in this research were initially designed to be used with the lesson interpreter described in Chapter 7. Nevertheless, they can also be useful for helping human teachers in selecting vocabulary for their classes, to complement the vocabulary provided in original schemes.

The tools developed so far are able to provide an adequate number of words for each curriculum item, to be used in teaching activities delivered by the computer. However, it gives only a little information on the syntax and semantics of the words, and does not link them together, which is one of the strongest requirements for Freire's approach.⁵

In order to organise the thematic vocabulary in a way in which the sequence of stimuli provided makes sense for the student it is necessary to adopt a linked data structure together with an appropriate word allocation policy. So far, the stimuli are allocated either sequentially or in random order at the time they are required for teaching. This issue is not covered in this thesis, although it certainly must be looked at in future research. The thesaurus structure as described in [Sparck Jones, 1991; Townley and Gee, 1980] seems to be suitable for this end. Alternatively, Wordnet [Miller et al., 1993] could also be investigated as an engine to provide information for this purpose.

⁵In the case of the derivative words it is possible to identify some information on the syntax and relationship with the original words. However, with distinct "word families" this information is not available yet.

Chapter 7

The ITS Generation

In chapters 4, 5 and 6 it was demonstrated that it is possible to automatically generate thematic stimuli to be used in PA approaches for teaching reading. A scheme for producing thematic vocabularies and the means for distributing the obtained vocabularies across the curriculum of specific PA methods have been described. This chapter describes the principles and tools developed to allow the use of the thematic vocabularies in Intelligent Tutoring Systems (ITSs) that act as auxiliary resources in teaching.

One of the key issues of this thesis is the acquisition of the expertise of a literacy specialist, herein denominated *the expert teacher*, and the codification of his or her knowledge in order to allow the computer to take decisions in the way specialists do. This knowledge is not built into the system; it should instead be obtained through interaction with the expert teacher. As a result, theoretically, different ITSs could be generated with these tools, each one corresponding to the vision of the particular expert teacher who is defining the literacy programme to the system. Consequently, the establishment of a consistent educational terminology to be used by the expert teacher while interacting with the system is paramount to this research. A correct understanding of what the terms in use mean increases the possibility of real transference of the expertise in the subject domain to the system.

In order to define the terminology to be used in the interaction with the expert teacher, and to facilitate the implementation of the system designed to deliver the teaching related to this thesis, an ontology has been developed with all the main concepts involved in it. Before presenting the system ontology let us provide an overview of its architecture.

7.1 System Architecture

In the general anatomy of an Intelligent Tutoring System (ITS), as seen by [Burns and Capps, 1988], the core of the system is composed of the student model, the tutor module and the expert system that contains all the knowledge about the subject domain. The real student would interact with this system core through the user interface, parameterised by a set of environmental information.

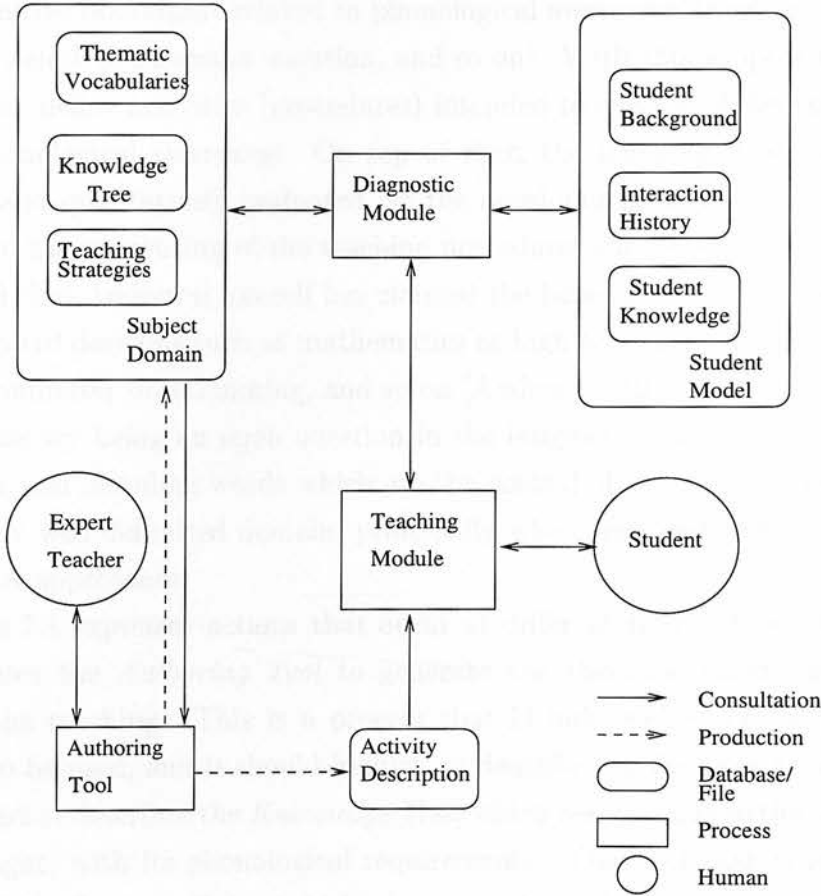


Figure 7.1: ITS Architecture

The architecture of the ITS developed in this research is shown in Figure 7.1. The overall scheme is intended to capture the teacher expertise in structuring the knowledge, in deciding on the teaching strategy and in delivering the teaching activities. The teaching expertise will be present in the vocabulary to be taught, in the structure of the curriculum, in the procedures used to teach and test the student, and in the teaching strategy rules.

Bloom's Taxonomy of Educational Objectives [Bloom, 1956] defines three domains of educational activities: the cognitive domain, the affective domain and

the psycho-motor domain. The focus of this research is mainly on providing the means for expressing and interpreting the facts and conditions at the level of the cognitive domain of literacy teaching. The general modelling approaches the cognitive models proposed by [Anderson, 1988, p.23-37], with features for describing procedural, declarative and qualitative aspects of the teaching process, going beyond the simple *expert system* paradigm. Put another way, the system offers the expert teacher the support for building up a literacy programme departing from primitive operations related to phonological awareness (word construction, phoneme deletion, phoneme insertion, and so on). With this support the expert teacher can define activities (procedures) intended to check and develop the student's phonological awareness. On top of that, the teaching strategies can be declared and qualitatively evaluated, so the development of the student can be traced and the sequencing of the teaching procedures can be sorted out. It could be argued that Anderson himself has claimed the benefits of such models only to well delimited domains such as mathematics at high school, basic sciences, introductory computer programming, and so on [Anderson, 1988, p.50]. Nonetheless, despite literacy being an open question in the language domain, the basic skills for coding and decoding words which are the central object of this thesis constitute a very well delimited domain, principally when regarded from the point of view of PA approaches.

Figure 7.1 expresses actions that occur at different times. First, the expert teacher uses the *Authoring Tool* to generate the thematic vocabularies to be used in the teaching. This is a process that is independent of the particular method to be used, and it should happen as described in Chapter 4. Second the expert teacher describes the *Knowledge Tree*, which corresponds to the curriculum to be taught, with its phonological requirements. The next task is to describe the *Teaching Strategy Rules*, which represent the ordering for the coverage of the curriculum and the plans for dealing with the various pedagogical scenarios. Eventually the expert teacher will describe the *Activities* of the methods, which are the real pieces of interaction with the student. All these descriptions are made with the help of the system's Authoring Tool. At this point, the interaction of the expert teacher with the system ceases, and, supposedly, all his or her expertise has been transferred to it.

Regarding the students, the interaction begins when their profile is first communicated to the system, in the form of the *Student Background* information. From this the *Student Knowledge* is set up, establishing the basis for tracking their progress in further interactions. The *Interaction History* is kept to help the

system in decisions on the sequencing of the activities for each particular student.

The interaction with the students is controlled by the *Teaching Module*, which is assisted by the *Diagnostic Module* with information on the *Subject Domain* and on the *Student Model* databases.

It is a rather complex environment in which the terminology plays an important role, as detailed below.

System Ontology

The word *ontology* is defined in the eighth edition of the Concise Oxford Dictionary as “the branch of metaphysics dealing with the nature of the being”. According to [Mizoguchi et al., 1996b] :

1. Ontology is a term in philosophy meaning “Theory of Existence”;
2. In AI, ontology is an explicit representation of conceptualisation;
3. In Knowledge Base Systems ontology is a system of primitive vocabulary and concepts used for building artificial systems.

The concept of ontology used here corresponds to the third definition above. The use of such a concept is useful not only for defining and structuring the terminology, as “taxonomy” does, but also as a tool for further formalisation and implementation of system components and procedures [Mizoguchi et al., 1996a]. The system ontology is presented as a network of terms used in the scope of the system, in functional or hierarchical relationships.¹

No automatic ontology processor has been used in this research, but the concepts which have been applied for this approach provide the conceptual support for the formal development of the system, mainly in relation to the development of the architecture previously described. In this approach all the terms applied in the system should be either defined using the other terms of the ontology or explicitly defined using ontology primitives. The primitives for describing terms can be simple sentences (e.g. “*student* is an illiterate person who interacts with the system to improve his or her reading skills”) or more complex expressions, making use of Backus-Naur Form (BNF) for defining the meaning of the terms. The first time an ontology term is defined or referred to, it should appear in italics, to highlight the fact that it is being defined. It might happen that some terms are not detailed for reasons of obviousness or lack of specific interest.

¹The functional relationship characterises elements that are directly related for co-operating or interacting functionally (e.g. *arms*, *legs*), whereas the hierarchical relationship characterises situations where one element subsumes others (*hands* versus *fingers*).

The broader concepts of the system ontology are presented below:

Literacy Teaching System Ontology

Hardware

Network

Server

Work Station

Input devices

Output devices

Software

Operating System

Support Software

System Architecture

Databases

Subject Domain

Thematic Vocabularies

Knowledge Tree

Teaching Strategy

Activities Description

Student Model

Student Background

Student History

Student Knowledge

System Agents

Authoring Tool (AT)

Teaching Module

Diagnostic Module

Human Agents

Student

Expert Teacher

Developer

Comments on the ontology top level concepts *Hardware*, *Software* and *System Architecture* follow.

Hardware

The *Hardware* concept refers to the physical environment of the system. Its parts are *Network*, *Server* and *Working Station*. The system is conceived to run on a network environment, with a central server providing the data for local terminals (or working stations) to deliver lessons and other activities for the students. The relevant subdivisions in *Working Station* are the *Input Devices* and *Output Devices* involved in the communication with the people that interact with the system. In the prototypical version of the system only the screen, the sound output device, the keyboard and the mouse are involved in the interaction. This is an open area as new hardware for interfacing computers is likely to be

available in future development. It is not the concern of this thesis, however, to go into this group of concepts in depth.

Software

The concept of *Software* refers to the system's developmental and operational support. It runs on different *Operating Systems*, such as IBM-compatible PCs and Unix work stations. The *Support Software* is the multi-platform multi-user Java environment. There is no need to go further into these definitions in the scope of this thesis.

Databases and Agents

The concept of *System Architecture* is subdivided into *Databases* and *Agents*. The term *agents* here is a more generic representative of the common denomination for "processes" in the computer science world, because it includes human agents. Databases embodies the basic structures for storing the expertise of the expert teacher and for providing other information required for teaching. *Thematic Vocabularies*, *Knowledge Tree*, *Teaching Strategy* and *Activities Description* are related to the programme. The other databases are related to the student model, with the information needed for guiding the interaction.

System Agents comprise the system's computer processes: the *Authoring Tool*, the *Teaching Module* and the *Diagnostic Module*.

Finally, *Human Agents* are the *Expert Teacher*, the *Student* and the *Developer* of the system, who might sometimes be referred to.

All the sub-concepts under the *System Architecture* concept are strongly related and must be anchored to a common ground. In the rest of this chapter most of these concepts will be detailed, not necessarily in the order they appear in the ontology.

7.2 Knowledge Tree

The *knowledge tree* corresponds to the "knowledge" in the cognitive domain in Bloom's taxonomy. It has been organised in a hierarchical tree, although this does not mean that the programme must be delivered in the same order that the topics appear in the tree. The facility for describing the *Teaching Strategy* will give the flexibility for sequencing the activities in more appropriate orders.

The ontological components of the knowledge tree are the *nodes*. In analogy to a real tree, the knowledge tree has a node of the type *root*, which is the most

comprehensive node of the tree and represents all the knowledge to be taught. As already mentioned in Section 5.2, the whole knowledge can be subdivided into more specific nodes, called *branches* and *leaves*, to represent “parts” of the knowledge.

Each node can be related to the development of one or more *skills*, which make use of the knowledge represented by the node. Each node should exercise one or more skills, but the skills are not exclusive to any node. The skills propagate hierarchically through the tree. This means that a skill declared in the root of the tree will hold for all nodes of the tree; a skill that is declared in a branch will be valid for all the sub-branches and leaves of that branch.

Another attributes of the nodes are the *Phonological Requirements* which represent the phonological features that a word must possess in order to belong to the vocabulary related to that particular item of the curriculum.

Therefore, the following ontological expansion applies to the Knowledge Tree:

Knowledge Tree

Node

Node Type

Node-name

Skill-name

Phonological Requirements

The KTDL language (*Knowledge Tree Definition Language*) was defined to allow the expert teacher to represent the Knowledge Tree and its related elements. It is a declarative language that can be *compiled* from the Authoring Tool interface.

Besides having the features for describing the curriculum structure, KTDL also implements the ways for expressing the phonological requirements of the vocabulary related to the nodes. For example, if words from the category “consonant-vowel-consonant” are required in that particular curriculum item then “CVC” should be declared as the phonological requirement for that item. As another example, if the curriculum item is related to three phoneme words starting with the phoneme “/k” or “/f”, containing a vowel phoneme in the second position and a “/t” in the third position, then the phonological requirement for that item should be “{/k,/f} {V} {/t}”. This would produce words like *fat*, *cat*, *fit* and so on.

The Appendices D.1, D.2 and D.3 are examples of complete written knowledge tree representations for the three literacy teaching methods examined in this thesis. The syntax for KTDL is presented in Appendix F.1.

Attention must be paid to the fact that the Knowledge Tree concept allows the definition of the “Application Ontology”, with a variable set of terms to be defined by the expert teacher, and that will be integrated with the terms of the “System Ontology” in the processing of the *Teaching Strategy*, described in Section 7.4. The integration of these two universes is another key issue in this research: as the System Ontology is kept open, it is possible to instantiate different methods for teaching reading and generate different ITSs with the same set of tools, without requiring the developer to make changes in the core of the system. This allows the expert teacher to experiment with different programmes or to develop new releases of a particular one. Put another way, this approach makes this system an ITS generator instead of a fixed ITS for teaching literacy.

7.3 Activities Description

The ontological expansion of the *Activities Description* comprises:

Activities Description

Lessons

Examples

Tests

Exercises

Steps

Objects

The Authoring Tool allows the expert teacher to define the teaching activities. The concept of *activities* refers to the different procedures or actions taken by the system to teach the contents of the curriculum. In [Major, 1995] many of these kinds of activities are referred to and described: presentation, assessment, teaching, testing and summarising are some examples. In this research the activities can be either *lessons*, *examples*, *exercises* or *tests*, possibly covering all the requirements of a literacy ITS. *Lessons* are intended to be used to first present a concept or an element within the subject domain. *Examples* should be used to illustrate the use of a particular element of the subject domain. The *Exercises* are meant to develop the skills of the students by trying their own solutions to the proposed questions. The *tests* play a more formal role in the student evaluation. It must be noted, however, that the processing of the *Student Model*, discussed in Section 7.5, will consider all kinds of activities to assess the student’s knowledge. By definition, one activity is able to exercise influence over one area of knowledge (node) and one skill of the student who performs it. Each activity will be declared as belonging to one of the nodes of the knowledge tree, with different degrees of

action over that area of knowledge (*Complexity Level* - ranging from 1 to 5) and with the skill it addresses. Note that a particular node can address more than one skill, but the same does not apply to activities, that are intended to exercise only one of the skills addressed by the node.

What this organisation model for the teaching proposes is that, in principle, for each skill addressed in a particular node of the knowledge tree there would be *one* lesson, *many* examples, *many* exercises and *one* test. This is not obligatory in all cases, and the system will be able to cope with different arrangements.

Once applied, lessons and examples deliver the same kind of output for the student modelling process of the student: the information that a lesson or example of complexity x within a particular node-skill combination has been visited. It is up to the expert teacher to define how the student model should be updated in face of the new situation after the activity has been executed. This can be done via TSDL, as described in Section 7.4.

The exercises and tests, likewise, deliver similar information for assessing purposes: an exercise or a test of complexity x belonging to a particular node-skill combination has been performed with a percentage y of success. Again, it is up to the expert teacher to define, via TSDL, how the student model should be regarded after this information is updated.

By definition the activities are sequences of *steps*. The expert teacher uses the *objects* available in the system to declare the steps of the activities. In the context of this research an *object* is a primitive concept in the system and the expert teacher is not able to create objects by him or herself alone. They have been created by the developer, after consideration of the pedagogic resources needed by PA methodologies for teaching reading. It might happen that at some point the expert teacher feels the necessity of a new primitive object in order to better express some generic teaching situation. In such a case the developer will work on the creation of this new object. All the objects available in the system already have their intrinsic interface with the assessment procedure of the student modelling process. The steps of the lesson are declared by choosing among various objects related to devices and resources to be used in the communication and work with the student. There are objects designed to print letters on the screen, to play songs or produce speech, to draw pictures, to allow the choice of one picture out of many others, to pose “yes-or-no” questions, to deal with sets of words or letters, and to manipulate variables in the system’s memory.

The lesson design style for the definitive authoring tool is similar to the one used to design automatic speech dialogues on the telephone in the CSLU/Toolkit

of the Center for Spoken Language Understanding of the Oregon Graduate Institute of Science and Technology [Colton et al., 1996]. The steps are described in a graphic screen where the object types are chosen by clicking with the mouse and placed in the design canvas. Once placed, the instances of the objects can be linked to determine the sequence of execution. An illustration of the look of the main screen of the future tool follows.

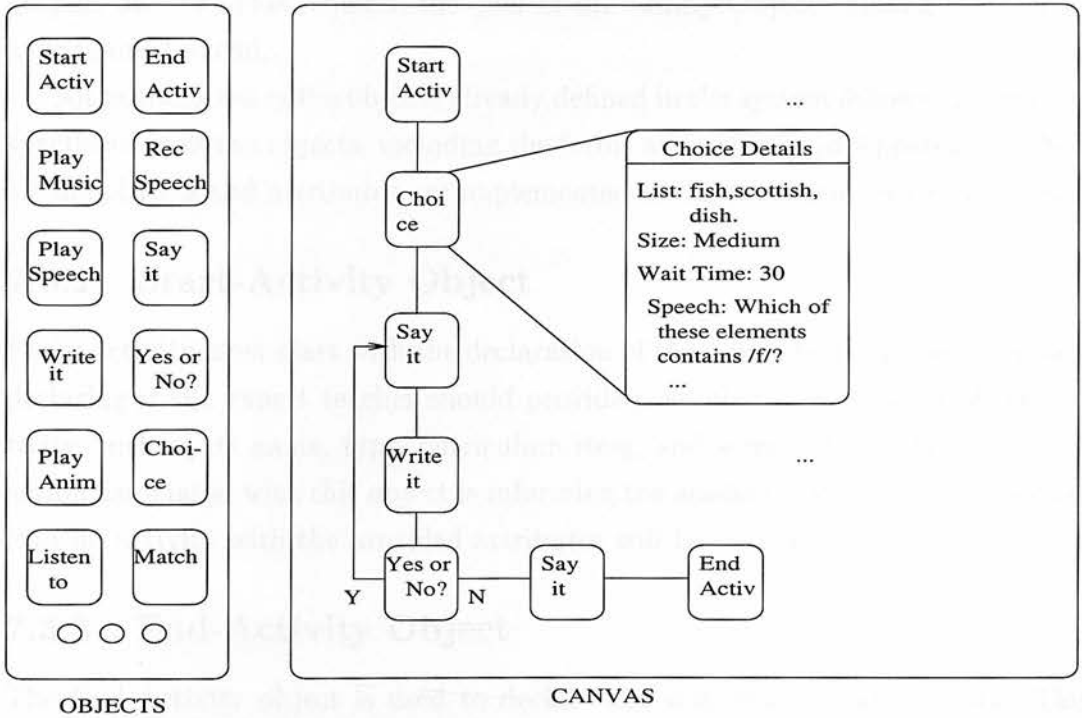


Figure 7.2: Main screen of the Authoring Tool

The ontological unfolding of the Objects group corresponds to the twelve different types of objects:

Objects

- Start-Activity Object*
- End-Activity Object*
- Input Object*
- Output Object*
- Show-Scene Object*
- Yes-or-No Object*
- Choice Object*
- If Object*
- Jump Object*
- Set-Var Object*
- Make-Set Object*
- Match Object*

In the first version of the authoring tool, instead of dragging the objects and placing them on the *canvas*, as suggested in Figure 7.2, the expert teacher is expected to click on one of the twelve objects presented in a list on the screen. Once the expert teacher chooses an object, the computer presents a form for that specific object on the screen to be filled in with the step attributes. In the first version the sequencing of the steps is indicated by sequence numbers instead of by graphic arrows. This required the offer of the “Jump Object” among the others, as explained ahead.

An explanation of the objects already defined in the system follows. A detailed specification of the objects, including the forms and fields, is in Appendix G. Not all the objects and attributes are implemented in the prototype of the system.

7.3.1 Start-Activity Object

Every activity must start with the declaration of the Start-Activity object. When declaring it the expert teacher should provide a number of attributes of the activity, such as its name, type, curriculum item, and some others. The semantic action associated with this object is informing the student and the student model that an activity with the provided attributes will be carried out.

7.3.2 End-Activity Object

The End-Activity object is used to declare the stop point of an activity. The object allows the definition of a *Banner* and a *song* to be shown and played as soon as the activity ends. After its completion the diagnostic module will consult the student model to determine the next activity to be executed.

7.3.3 Input Object

The Input Object is conceived to allow the student to input data either through the keyboard, microphone or hand-writing device. In the prototype of the system, input of data is only possible through the pointing device (mouse). The input data from the mouse are treated as special cases, via the objects Yes-or-No, Choice and Match, as described below. Consequently, the Input Object has not been implemented yet.

7.3.4 Output Object

The Output Object allows output from the system to be produced either as speech, a song to be played, written material or a picture to be displayed on the

computer screen. The speech is produced by the Festival System by synthesising from a written text, which can be dynamically generated. It is possible to produce a combination of these different forms of output in the same step.

7.3.5 Show-Scene Object

The Show Scene Object has not been implemented yet. It is a particular case of output, which, by its complexity, is considered separately from the other forms of output. It is a basic “comics composer”, which will be able to represent either one, two or three *characters* (either a man, a woman, a boy or a girl, with different features) in a *scenery*, each one with an optional *speech balloon* over their head. Its aim is to provide the expert teacher a way for automatically generating comic pictures to convey relevant communication with the student.

7.3.6 Yes-or-No Object

The Yes-or-No Object poses the student a “yes-or-no” question, often related with the sequencing of the activity (e.g. “Would you like to see another word?”). As semantic action it shows the student a screen with two buttons (yes, no), to be chosen via the mouse.

7.3.7 Choice Object

The Choice Object is used to a place set of *buttons* on the computer screen, from which the student is required to choose one (multiple choice). In the definitive version the set of buttons will be represented by icons, tags and speech. If the mouse rests more than one second over a button, a tag should appear beside it and its content should be spoken aloud to the user. If the student clicks on a button it will be considered as the choice for the posed question.

7.3.8 If Object

This is an object related to the control of the sequencing of the activity, and it is used to cause a jump to one of two possible continuing steps in the activity depending on the evaluation of a *logical expression*. No input or output is required or produced. The logical expression must be formulated with reserved or user defined variables following specific construction rules. Typically it should be used to evaluate and take actions in regard to conditions such as the performance of the student, the number of times a particular object has been executed within an activity, and so on.

7.3.9 Jump Object

The normal execution of the steps of an activity follows the sequence of declaration of the steps. The Jump Object is used to alter the sequence of execution of the steps of the activity, causing an unconditional jump to another step it indicates. It is a provisional object, for the future versions of the system will treat the sequencing of steps graphically.

7.3.10 Set-Var Object

The Set-Var Object allows the definition of working *variables* and the attribution of variable values using *expressions* involving *pre-defined functions*. The variables in question will contain information that is relevant to the execution of the activity, such as words to be analysed, phonemes contained within words, counters, and so on. The system provides a vast spectrum of functions to generate and deal with variables (see Appendix G for more information).

7.3.11 Make-Set Object

The system contains some basic *sets* needed for the literacy teaching process, such as the *vocabulary*, the *phonemes* used in the language, the *letters*, the *vowels*, the *consonants*, the *numbers*, and so on. The Make-Set Object is used to generate different new sets from the basic ones, for the convenience of the teaching process. For instance, the expert teacher can use this object for generating a set with the phonemes that belong to a particular word. The object can also be used to obtain a fixed or a variable number of words with a particular phonological feature. The many different functions to deal with sets are described in Appendix G.

7.3.12 Match Object

This object is intended to provide a way to interactively match or combine two *lists*. When executed the computer will present the values of the first list in an array of buttons on the screen, to be matched with a hidden second list. As the matches occurs, the values of the second list will be shown. This object is appropriate to be used in word construction exercises. See the examples for its uses in Appendix G.

7.3.13 Making Sense of the Activities Description

The isolated view of the activities as shown in this section might appear complex at first sight. Chapter 8 presents some case studies in the construction of activities

using the objects described in the previous section.

7.4 Teaching Strategy

The expression *teaching strategy* in this thesis means a set of rules to orient the system in the choice of the appropriate activities to propose to the students, based on their background, history and current results. The information about the student is kept in a set of files that comprise the *student model*. The interpretation of the data and the decision making process, however, should be based on the expertise the expert teacher transfers to the system.

The *expert overlay* modelling approach as described in [VanLehn, 1988, p.81-83] has been adopted in this work. This choice is justified because the curricula to be taught in the case of basic literacy from the point of view of phonological awareness are constructed as limited and well defined sequences of skills on phonological features of the language. The *expert model* is a reference for evaluating the students, who are represented in the computer in their student models. Both the expert and the student models are images of the knowledge tree. As there can be different knowledge trees for each literacy programme represented with the system, the expert and the student models are specific to each programme. The canonical representation of the expert model corresponds to the knowledge tree with a hundred percent of *knowledge rate* in every activity of every node of the tree. However, the expert teacher, who is defining the teaching strategy, is provided with means for changing this vision, allowing the expert model to be flexible and not necessarily to be perfect for all the items of the curriculum.

The *student model*, in its turn, starts with no points in the activities of the knowledge tree. As the process goes on the system changes the knowledge rates when appropriate. The agents supposed to promote changes in the student knowledge, and consequently in the student model model, are the *activities*. As described earlier, each activity is related to one node of the tree and is liable to promote some improvement in the student's knowledge. The extent of the real changes depends on the student performance in the activity and on the activities' parameters themselves. Recall that amongst the activities description parameters there is the activity complexity, which will be used to scale the knowledge rate of the student after the activity is completed. Besides that, the type of the activity plays another evaluation role: the *lessons* and *examples* are meant to exercise limited influence on the rating of the student, while *exercises* and *tests* might be more significant in the assessment process.

What differentiates *ITSs* from traditional teaching tools is their ability to adapt to the user's needs and their capacity to try different teaching strategies [Burns and Capps, 1988]. With the authoring functionality proposed in this system, the expert in literacy is able to define a set of strategic rules to guide the delivery of lessons, examples, exercises and tests. A *Teaching Strategy Definition Language (TSDL)* is defined to allow the expert teacher to declare the rules to be followed to deliver the activities in order to facilitate the students to progress in their studies.

The system can evaluate the student's needs based on the student model, expert model and on the teaching strategy just mentioned. Because the expert and student models are based on the curriculum represented by the knowledge tree, the system also has information about the *subject domain* and *skills* addressed by each of the activities. Therefore, the system is able to make an evaluation and indicate the next teaching movements. The activities run by the system deliver evaluation ratings after their execution, so the system can keep the student model up to date. The design of TSDL is based on the style proposed by [Major, 1995], with assertions in English addressing the student model, the knowledge tree, and the attributes of the activities. Its design permits the expert teachers to describe how they interpret the facts about the student development and how they indicate what to do and what not to do next. An example of what it looks like follows.

```
if average (Basic Code) is good and
exercising(Basic Code) is enough and
testing(Basic Code) is enough
then disable lesson(Basic Code), disable test(Basic Code),
suspend exercise(Basic Code) (3).
```

```
if Code knowledge(Bug On Jug Sounds) is poor then
enable lesson(Bug On Jug Sounds),
enable test(Bug On Jug Sounds),
rank lesson Code Knowledge (Bug On Jug Sounds) +100.
```

The components of the language are terms extracted from the knowledge tree and skills, merged with *logic connectives*, *functions* and *scale values*. For instance, in the first line of the code above the word "average" means the average value of tests in the "Basic Code" block of knowledge, compared to the scale value "good". The scale values used in the example are extracted from the qualitative scale for marks (*poor*, *enough*, *good*, *excellent*). Function names are used to indicate actions to be undertaken as in "disable test (Basic Code)".

The examples above show that TSDL deals typically with *Ranking Rules*,

which are used to reestablish the rank of activities after the student model has been modified by the execution of one activity. Their use can virtually affect the rank of all activities, resulting in a complete new ordering scale to choose the next activity for the student. For instance, the instruction *disable lesson (Basic Code)* will delete all the activities of the type *lesson* of the curriculum item *Basic Code* from the ranking of activities. This category of activities can be re-enabled further by a specific *enable* instruction. The instruction *suspend exercise (Basic Code) (3)*, in turn, will keep this category of activities out of the ranking for the next three choices of activities, after which they will be eligible again. The instruction *rank lesson Code Knowledge (Bug On Jug Sounds) + 100* credits a hundred points more to the activities of the category specified, making them *possibly* more likely to be elected as the next activity. The standard difference of points between the *activities* of two neighbour branches in the knowledge tree is 10,000 points. In the same branch *lessons* are quoted 1,000 points higher than *examples*, which are quoted 1,000 points higher than *exercises*, which in turn are quoted 1,000 points higher than *tests* in the ranking. Of course the addition of ranking points does not guarantee, but can accelerate, or slow down, the choice of these activities in the following runs. The complete syntax of TSDL is shown in Appendix F.2.

The *Student Background* file contains information about the previous activities of each student. These clues are important in the choice of the themes and words to be used mainly at the beginning of the teaching process, when little is known about the student. Essentially it will supply the Diagnostic Module with weights for the subjects and vocabulary in the Thematic Vocabularies.

The *Student History* is a log file tracking the work of the student with the system. It contains information on the performance of the student in the exercises (timing and scores). The Student Model file will keep the overlay model, based on the knowledge tree, for each student. As the activities take place the Diagnostic Module records the student ratings in the overlay model.

This approach of giving back to the expert teacher the mission of defining what to do with the student is similar to the one adopted by [Major et al., 1997] in their REDEEM System.

At this point it makes sense to introduce the idea of giving the expert model rates different from a hundred per cent. This is justified as long as the expert teacher feels that the requirements for one particular subject or skill in the tree are too strong and should be relaxed. If this is the case the expert teacher could then add to the teaching strategy rules like:

expertise (Advanced Code) = 80.
expertise (Code Knowledge) = 90.
expertise (Basic Code, Auditory Processing)=85.

This would introduce a reduction of requirements in that particular branch, allowing the student to try more advanced lessons and progress in the sequence of learning. However it must be noted that it can be useful only when the expert teacher perceives that there has been some distortion in the overall evaluation of the activities under that knowledge branch.

The score approach proposed here is simple and direct, and transfers to the teaching strategy scheme the job of considering the whole scenario and selecting the next task. In order to avoid the diagnostic module repeatedly reusing activities, there will be an audit trail file tracking all student movements and scores. These values must be available in TSDL for the expert teacher to deal with.

7.5 Student Model

As mentioned before, the Student Model of the system is based on the overlay model concept, using the Knowledge Tree as the organisational element of the curriculum. Each node of the tree corresponds to a particular segment of the curriculum, and is supposedly able to exercise particular skills. The activities to exercise these skills are the lessons, examples, exercises and tests, as described before. So it is possible to attach activities to each node, to be applied during the teaching. Each activity has its complexity degree declared along with the expert mark and the particular skills it exercises. Later on these initial declarations can be explicitly overridden by the teaching strategy scheme.

Each node of the tree will have one expert model file and as many student model files as there are students. The expert model of one node can be loaded in the memory of the system alone. However, the information relating to the student model in a particular node can only be in memory after the corresponding expert model is loaded.

The expert model contains the following data for each node of the knowledge tree:

- hierarchy number
- name
- type
- skills
- source
- exclude

number of phonetic specifications
 phonetic specification 1
 phonetic specification 2
 ...
 number of lessons
 lesson data (one for each lesson)
 lesson id
 lesson skills
 lesson expert score
 lesson complexity
 number of examples
 example data (one for each example)
 example id
 example skills
 example expert score
 example complexity
 number of exercises
 exercise data (one for each exercise)
 exercise id
 exercise skills
 exercise expert score
 exercise complexity
 number of tests
 test data (one for each test)
 test id
 test skills
 test expert score
 test complexity
 number of children
 child 1
 child 2
 ...

The meanings of most of the above mentioned data is obvious from the names of the fields. The ones that are not obvious are:

- source - this is the name of the vocabulary set to be used as the source of words for the node vocabulary;

- exclude - this is an optional parameter to be used when the node being described has subdivisions and the expert teacher does not want the stimuli (words) present in the subdivisions to be present in this node;
- child i - this is a pointer to a node that is a sub-division of the knowledge represented in this node; there must be one for each sub-divisional node.

The student model uses the basic data structure provided by the expert model and extends it to keep the data about the student. The complementary data structure for each node of the knowledge tree is shown below:

last update time

lesson section

number of visits to lessons

average score in lessons

most complex lesson level visited

lessons are enabled

report of visits (one for each lesson in the node)

date of visit

score

visits already made

average

example section

number of visits to examples

average score in examples

most complex example level visited

examples are enabled

report of visits (one for each example in the node)

date of visit

score

visits already made

average

exercise section

number of visits to exercises

average score in exercises

most complex exercise level visited

exercises are enabled

report of visits (one for each exercise in the node)

- date of visit
- score
- visits already made
- average
- test section
 - number of visits to tests
 - average score in tests
 - most complex test level visited
 - tests are enabled
 - report of visits (one for each test in the node)
 - date of visit
 - score
 - visits already made
 - average

The meaning of most of the fields of the student model are quite obvious. It is worth clarifying that the *score* in each section is the score obtained by the student in the last visit to the activity, and that the *average* is the average score of all the visits already made to that particular activity. Besides that, the attributes that are global to each section (e.g. *number of visits to tests*, *average score in tests* and so on) are a summary of all the visits to all the activities of that kind in the node.

The student and expert models methods are designed to be manipulated by different parts of the system, as described below:

- The Knowledge Tree compiler uses them to create the basic expert model structure, as it compiles the tree and generates the vocabularies of the system.
- The Authoring Tool editor uses them to complete the basic expert model with the information about the activities as long as it appends or alters activities in the tree. It provides the provisional expert's scores at that same time.
- The Student Background file, which is the one that contains information about the students (code, name, background) uses these structures to create the student model when the student is registered to the system, or to alter it later. The Student Background file is described in more detail below.
- The Tutoring System interpreter uses these structures to update the student model while teaching.
- The Strategy Rules interpreter may use them to alter information on the expert model (the scores) and on the student model (the availability of activities)

as it interprets the general strategy rules.

- Again, the Strategy Rules interpreter uses them while evaluating the performance of the student and selecting activities for teaching next.

The Student Background file is the one that contains information about the student. Each student who interacts with the system has a record in the following format:

student code
student name
subjects of interest
personal data
literacy programme

where:

student code: up to five letters or numbers.

student name: up to 45 letters.

subjects of interest: so far the one-word titles of the subjects the student is interested in or familiar with, separated by commas (e. g. " food, cars ").

personal data: for future use.

programme: the two-letter code for the literacy programme to be applied to this student (e. g. "pg", for Phono-Graphix).

7.6 Observations on the Prototype

The prototype for the generation of Intelligent Tutoring Systems for literacy teaching developed in this research intends solely to demonstrate the feasibility of the main proposal of this thesis concerning the combination of the different literacy teaching approaches (Phonological Awareness and Thematic) and of the proposed architecture for organising the knowledge and the expertise required for teaching.

It has been totally coded by the author in the Java Language and it does not have the refinement required for real teaching. Its potential as a model for a generation tool is demonstrated through the case studies developed in Chapter 8. The details for operating and the figures on the tools developed can be seen in Appendix I.

Chapter 8

Case Studies

This chapter aims to demonstrate the use of the tool in the development of an implementation of the methodology proposed by an interactive system. The case study examples described in Chapter 8 will be described as requested in that chapter with the implementation by the author will be discussed. In the next section of the teaching strategies will be shown and characterized.

8.1 Phono-Graphic Application

8.1.1 Pointing out the first sound

One of the first activities of Phono-Graphic is the "pointing out the first sound" activity. It also works with the same methodology as the other activities, consisting of a sound and sign that the student has to identify. The student is asked to point out the first sound of the word.

This activity was developed out with words from a list of 100 words of Man. Code. 1941, according to the author of the methodology. The activity is used to help the student understand the right to sound relationship of the words and to help to understand the left to right relationship of words to characters with a computerized font to represent sounds in words, and to create an activity of "Phono-Graphic speaking strategy". This activity is implemented by the teacher as follows: as an example, Figure 8.1 shows two examples of this reference. In this example, the computer would say: "What is the first sound of the word 'year'?" The student is expected to choose one of the buttons by clicking on it to answer the first sound of the word. After that, the activity will continue with the next example, depending on the student's answer. The computer will say: "The first sound is 'y' or 'e'?" (whether the answer was right or wrong). The student will be able to click on the student would be able to answer "y".

Chapter 8

Case Studies

This chapter aims to demonstrate the use of the word selection scheme and the application of the architecture proposed by implementing teaching activities from the three PA programmes described in Chapter 5, Section 5.3. The activities will be described as suggested in their respective methods manuals and their implementations by the system will be discussed. In the last section an application of the teaching strategies will be shown and commented on.

8.1 Phono-Graphix Applications

8.1.1 Pointing out the first sound

One of the first activities of Phono-Graphix in the teaching of “Basic Code” is the “pointing out the first sound” exercise. In this exercise the teacher shows a picture corresponding to a word and says this word slowly. Then she asks the student to point out the first sound of the word.

This activity can be carried out with words from any of the *branches* of Basic Code, and, according to the authors of the method has as its goals, “to cause the child to understand the sight to sound relationship of text; to cause the child to understand the left to right relationship of text; to cause the child to understand how to represent sounds in words; and to create an automatic Phono-Graphix spelling strategy”. This activity is implemented by the system in a form of an *exercise*. Figure 8.1 shows one example of this interaction. In this example, the computer would say: “What is the first sound of the word *pop*?” The student is expected to choose one of the buttons, by clicking on it twice, to indicate the first sound of the word. After that the activity will continue in the next steps, when, depending on the student’s answer the computer would say an appropriate comment (whether the answer was right or not) and would proceed asking whether the student would like to see another word.

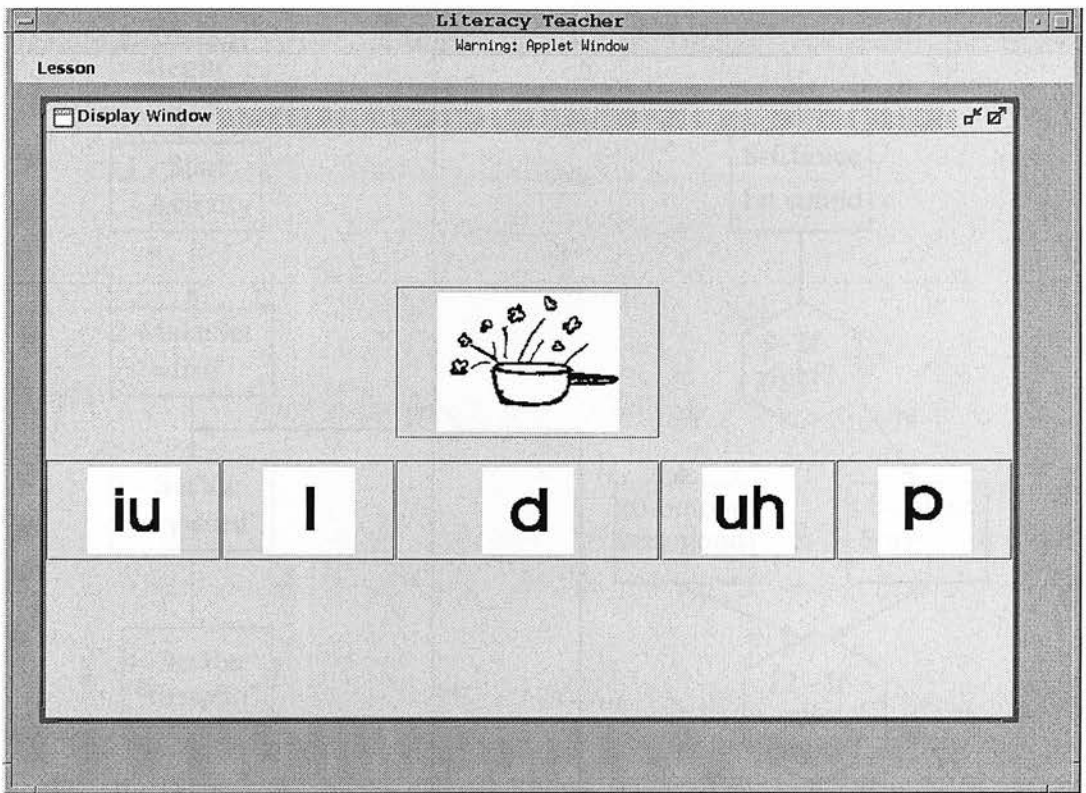


Figure 8.1: Screen for *Pointing the First Sound*

The description of an activity is a computer programming task carried out through the tools available in the system. The expert teacher is expected to get used to the system vocabulary and to the tools in order to define the activity. As discussed on Chapter 7 and shown in Figure 7.2, the future “programming tool” for the activities is expected to be more user friendly than the one available currently, in the form of a flow-chart drawing support tool rather than a language programming tool.

A flow-chart diagram has been drawn, in Figure 8.2, in order to help with the construction of the activity.¹ Such a diagram is similar to the canvas of the screen in possible future versions of the authoring tool.

¹Observe that no activity is built-in to the system; all activities must be described by an expert teacher. The examples presented here are not claimed to be either complete or pedagogically excellent. They are rather to illustrate the use and the potential of the proposed tools.

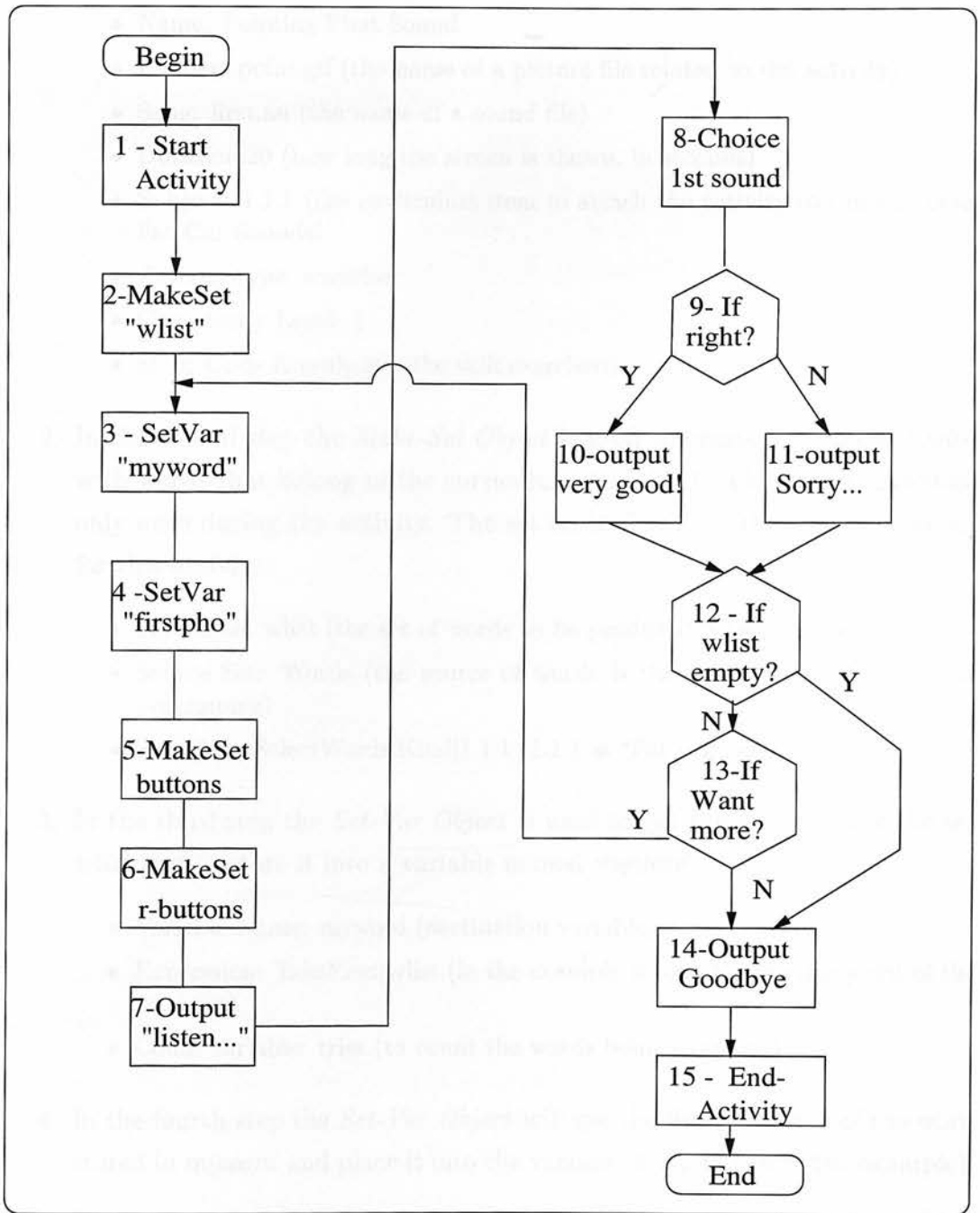


Figure 8.2: Flow chart for the *Pointing First Sound* activity

Each block of the flow chart in Figure 8.2 is discussed below:

1. In the first step, using the *Start-Activity Object*, the expert teacher declares the attributes of the activity:²

²In these descriptions of the activities we generally describe only the most relevant fields of the objects in use. The description of all fields of the objects of the system can be seen in Appendix G.

- Name: Pointing First Sound
 - Banner: point.gif (the name of a picture file related to the activity)
 - Song: first.au (the name of a sound file)
 - Duration:20 (how long the screen is shown, in seconds)
 - Subject: 1.1.1 (the curriculum item to attach the activity to - in this case, Fat Cat Sounds)
 - Activity-type: exercise
 - Complexity Level: 1
 - Skill: Code Knowledge (the skill exercised)
2. In the second step the *Make-Set Object* is used to create a set named *wlist* with words that belong to the curriculum item 1.1.1. This step is executed only once during the activity. The set created will be the source of words for this activity.
- Set Name: *wlist* (the set of words to be produced is named *wlist*)
 - Source Set: Words (the source of words is the whole vocabulary for the programme)
 - Selection: SelectWords:Kind|1.1.1 (1.1.1 is “Fat Cat Sounds”)
3. In the third step the *Set-Var Object* is used to get the first word of the set *wlist* and to store it into a variable named *myword*:
- Variable Name: *myword* (destination variable)
 - Expression: TakeFirst|*wlist* (in the example it took the word *pop* out of the set *wlist*)
 - Count variable: *tries* (to count the words being examined)
4. In the fourth step the *Set-Var Object* will get the first phoneme of the word stored in *myword* and place it into the variable *firstpho* (/p in the example):
- Variable Name: *firstpho*
 - Expression: takephoneme(*myword*:1:1)
5. In the fifth step the *Make-Set Object* will pick-up four randomly chosen English phonemes and place them into a set named *buttons* (in the example it picked up the sounds /uh, /d, /l and /iu, and stored into the set *buttons*):
- Set Name: *buttons*
 - Total of Elements: 4
 - Source Set: Phonemes

- Selection: pickup
6. In the sixth step, with the *Make-Set Object*, the first phoneme of the word stored in *firstpho* will be mixed with the set of phonemes of the previous step. The resultant set will be placed into a *rbuttons* (standing for “randomised buttons”). In the example, the result was the creation of the set *rbuttons* with the elements */iu*, */l*, */d*, */uh* and */p*.
- Set Name: rbuttons
 - Source Set: buttons;firstpho
 - Selection: Random
7. In the seventh step, the *Output Object* is used to make the computer produce an utterance to the student. It will say “listen to this word:”, and immediately after that it will say the word contained in *myword* (*pop*, in the example). To obtain this, the field “Text to speak” in the *Output Object* is codified as “Listen to this word:”+*myword*.
8. The *Choice Object* is used in the eighth step to show a picture of the word (if it is representable and available in the pictures database), with a set of buttons with the written representation of the phonemes in *rbutton*, and to give the student an instruction about the task. If the student clicks once on any button, the phoneme will be said. If the student clicks on the picture, the word will be repeated. When the student clicks twice on any button it is considered to be the answer the student has chosen. The most important fields of the object are:
- Element List: rbuttons (the array of buttons)
 - What to say: “What is the first sound of ”+*myword*
 - Big Button: *myword* (the word to be depicted)
9. In the ninth step the *If Object* makes an evaluation as to whether the student made the right choice. The activity will then proceed to one of the two next steps depending on the evaluation. In the expression to be evaluated (below), the variable *BarElement* is a reserved variable from the system which contains the value of the last button pressed by the student with the system’s pointing device. The important fields of the *If Object* are:
- Expression: *firstpho=BarElement* (compares clicked button to the first phoneme of the word)
 - Yes Branch: 10 (the address of the *Output Object* for the choice is right)

- Yes Counter: ok (incremented when the result of the expression is true)
- No Branch: 11 (address of the *Output Object* to wrong answer)
- No Counter : notok (incremented when the result of the expression is false).

10. In the tenth step the *Output Object* is used to praise the student for the right answer. To obtain this, the field “Text to speak” in the *Output Object* is codified as “Congratulations, you did it right”.

11. In the eleventh step the *Output Object* is used to communicate that the student has failed. The field “Text to speak” in the *Output Object* is codified as “Sorry, the sound ”+BarElement+ “ is not the first sound of the word ”+myword.

12. In the twelfth step an evaluation is made with the *If Object* to verify whether there are still words to be seen in *wlist*. If it is the case, the system proceeds with the next step, otherwise it jumps to the last step of the activity. The expression field of the *If Object* has the value: “SetLength|wlist| < |1”. *SetLength* is a reserved word, destined to give access to the amount of elements of a particular set.

13. In the thirteenth step the *Yes-or-No Object* is used to ask the student whether he or she wants to examine more words. If the answer is *yes* the system jumps to the third step, to get another word from the *wlist*, and to run another cycle of the activity. Otherwise, it goes to the next step. The most relevant fields of the object are:

- Question: “Would you like to try again?”
- Waiting time: 20
- Yes branch: 3 (in this case the activity will continue in the step indicated, for a new cycle)
- No branch: (it is empty, next step assumed)
- Default answer: no (if the student takes more than 20 seconds to answer, “no” is assumed)

14. In the fourteenth step the *Output Object* is used to say the good-bye message as determined by the expert teacher. In this case it is: “In ” +tries+ “ tries you succeeded in ” + ok+ “ and failed in ” +notok. With this construction, the system will use the counters *tries*, *ok* and *notok*, incremented during the activity, to give the student some information about the performance in the task. In the utterance the system will replace the name of the variables by their actual contents.

15. In the last step, with the *End-Activity Object* the system shows the final banner and transfers the control to the diagnostic module.

8.1.2 Word building

Another activity of Phono-Graphix which can be used with various of its curriculum items is the one to order the sounds of the words. In it the teacher shows a picture representing the word in focus, and the pictures of the graphemes (*pictures of the sounds*, in Phono-Graphix terminology) in a randomly ordered row. Next, the teacher points the finger to the first empty space below the sounds and asks “which is this sound”, inducing the student to drag the first sound of the word to the first position. The student then chooses one of the sounds of the sequence and places it into the empty space. The teacher makes a comment, whether right or wrong. If it is right, the teacher points to the next space and prompts the student again. And so it goes up to the last sound left.

This exercise is indicated to “cause the child to understand the sight to sound relationship of text, to understand the left to right relationship of text, to consciously understand how she can represent sounds in words and to create an automatic spelling strategy”, according to the authors of the method [McGuinness and McGuinness, 1998, p.70].

This activity is also implemented in the system in the form of an *exercise*. Figure 8.3 presents an example of this activity. While presenting Figure 8.3 the computer asks the student: “What are the sounds of the word fat?” and then prompts “What is this sound?”, pointing to the first position. After the student clicks one button twice the computer either says “That’s right” or “This sound is not here, try another”, depending on the evaluation it makes by comparing the order of the sound with the right order provided in *button list*. When every position is matched the message “OK, you matched everything” is said and an evaluation rate, in terms of a percentage, is made available in the reserved variable Performance. After that, the activity designer can refer to the value of the variable Performance to decide how to proceed in the activity (using the *If Object*).

The sentences used in the dialogue above correspond approximately to what Phono-Graphix designers recommend as their approach to this activity.³

³From annotations made during the participation of the author in a Phono-Graphix workshop held in London, October, 1998.

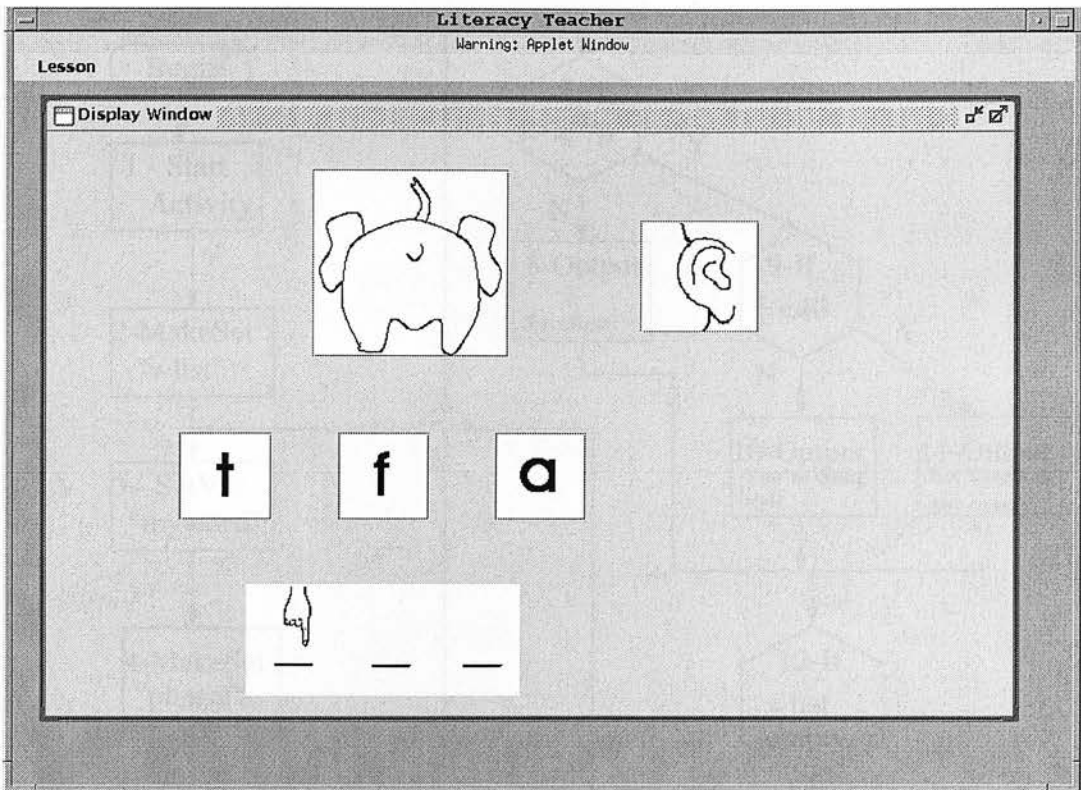


Figure 8.3: Example of the Match Object

Another flow-chart diagram was drawn, in Figure 8.4 to aid the comprehension and construction of this activity.

The Figure 8.4 is discussed below:

1. The first step is to provide general information on the activity with the *Start-Activity Object* (type of activity, complexity, curriculum item, and so on).
2. In the second step, with the *Make-Set Object*, a set named *wlist* with words that belong to the target curriculum item is created.
3. In the third step the *Set-Var Object* is used to get the first word of the set *wlist* into the variable *myword*. In the example, the word was *fat*.
4. In the fourth step the *Make-Set Object* will get the set of phonemes of the word in *myword* into the set *phoset*. In the example, the three elements would be */f*, */a* and */t*.
5. In the fifth step the *Make-Set Object* is used to randomise *phoset* into *phosetout*. In the example the result was */t*, */f*, */a*.

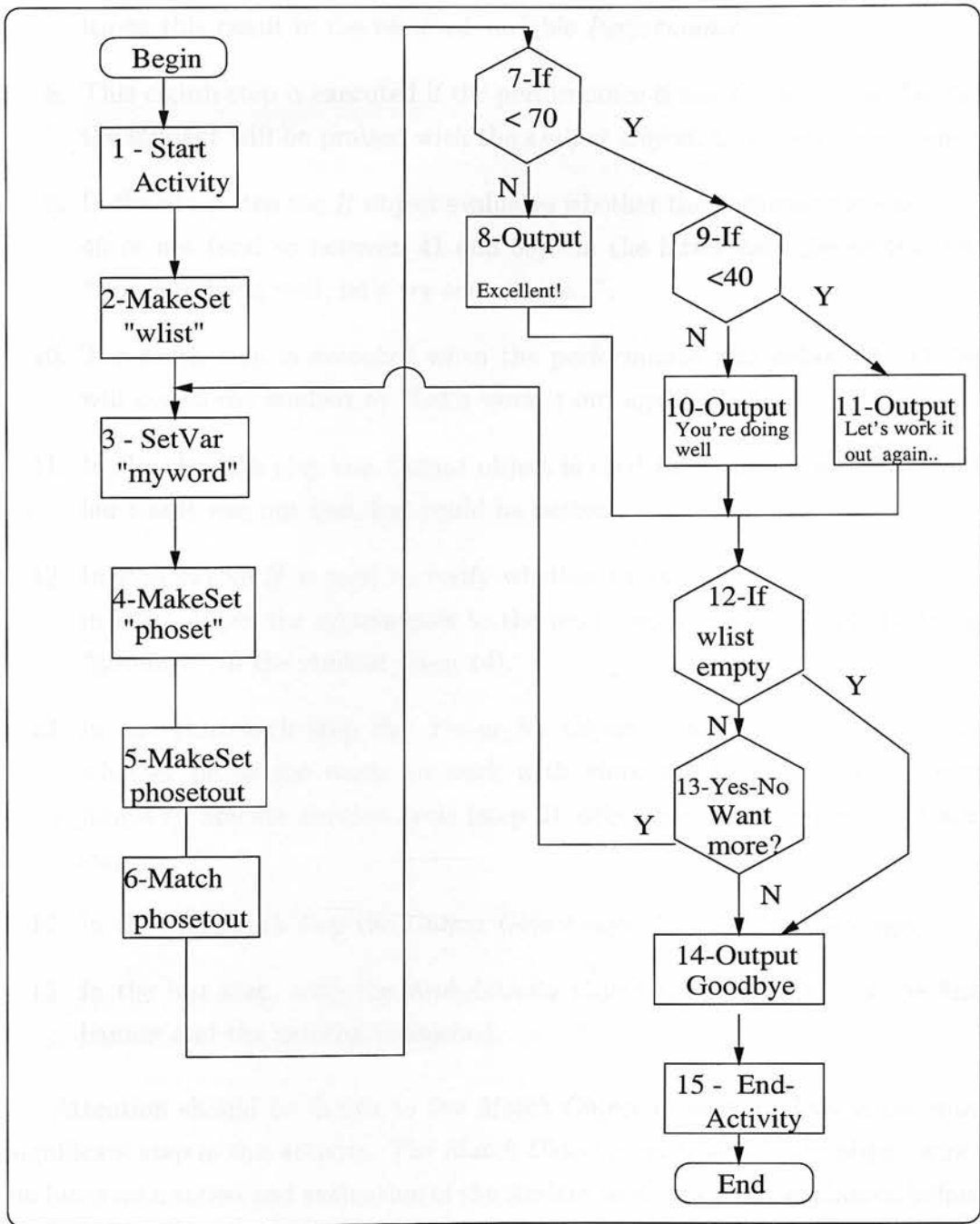


Figure 8.4: Flow chart for the *Word Building* activity

6. In the sixth step the *Match Object* will present the phonemes in *phosetout* and a list with spaces intended for ordering the phonemes, prompting the student to do it.
7. In the seventh step, with the *If Object*, the system verifies whether the student's score was above or below 70%. The execution of the *Match Object*

leaves this result in the reserved variable *Performance*.

8. This eighth step is executed if the performance is not below 70. In this case the student will be praised with the *Output Object*, by saying "Excellent!".
9. In the ninth step the *If* object evaluates whether the performance was below 40 or not (and so between 41 and 69). In the latter case the system says "You are doing well, let's try some more...".
10. The tenth step is executed when the performance was below 40. *Output* will say to the student to "Let's work it out again...".⁴
11. In the eleventh step the *Output* object is used to communicate that his or her result was not bad, but could be better.
12. In step twelve *If* is used to verify whether there are still words to be seen in *wlist*. If yes the system goes to the next step, otherwise it jumps to say "goodbye" to the student (step 14).
13. In the thirteenth step the *Yes-or-No Object* is used to ask the student whether he or she wants to work with more words. If yes, the system jumps to execute another cycle (step 3), otherwise, it proceeds to the next step.
14. In the fourteenth step the *Output Object* says the good-bye message.
15. In the last step, with the *End-Activity* object the system shows the final banner and the activity is finished.

Attention should be drawn to the *Match Object* in step 6. This is the most significant step in this activity. The *Match Object* is quite a complex object which includes interaction and evaluation of the student in its body. As explained before, it prompts the student and controls the progress in relation to the matching processes, assessing the student's performance at the same time. The main fields of the object are:

- Word picture: *myword* (it will show a picture related to the word contained in *myword* - in this case, *fat*).
- Button list: *phoset* (this is the list used as reference to be matched).

⁴Recommendations are made to never be rude or negative in the comments, but rather say phrases like "There is still room for improving" or "This one was nearly right...". However, this is not built-in to the system; it is up to the expert teacher to decide what to say to the student.

- Match list: phosetout (this is the list that is shown to the student).
- Object prompt: “What are the sounds of the word:“+myword (this is the first prompt of the object, said just once).
- Action prompt: “What is this sound?” (this is the prompt for each interaction, while pointing to the next empty space in the word).
- Fail message: “This sound is not here... try another” (this is the message said when the match is wrong).
- Success message: “That’s right” (the message said when it is right).
- Last message: “OK, you matched everything” (the message to be said when the student finishes the match).

8.2 Jolly Phonics Applications

8.2.1 Missing Sound

Jolly Phonics is a more concise method than Phono-Graphix. However the teaching activities of Jolly Phonics are quite similar to those of Phono-Graphix. The activity of looking for the missing sounds of a given word is an activity which has a parallel in Phono-Graphix. It can be applied to any curriculum item. In it the teacher shows a picture representing the word and the graphemes corresponding to the sounds of the word in the right order, but with one of them missing.⁵ Then the teacher points a finger to the empty space and asks “which is the missing sound?”, inducing the student to drag one of the five graphemes that lie below to the empty position. After the student answers, the teacher makes a comment, whether the answer is right or wrong.

This activity is implemented in the system in a form of a *test*. The Figure 8.5 is an example of the activity as implemented in the system. There you can see the picture of a *tap*, and a representation of its sounds, but with the sound /a missing - instead, there is a little hand with the finger pointing to the position, while the computer asks: “What is the missing sound of tap?”. At the bottom of the screen there is a sequence of graphemes from which one should be chosen by the student, including the right one.

⁵The set of symbols for representing the graphemes in Jolly Phonics is slightly different from the set used in Phono-Graphix.

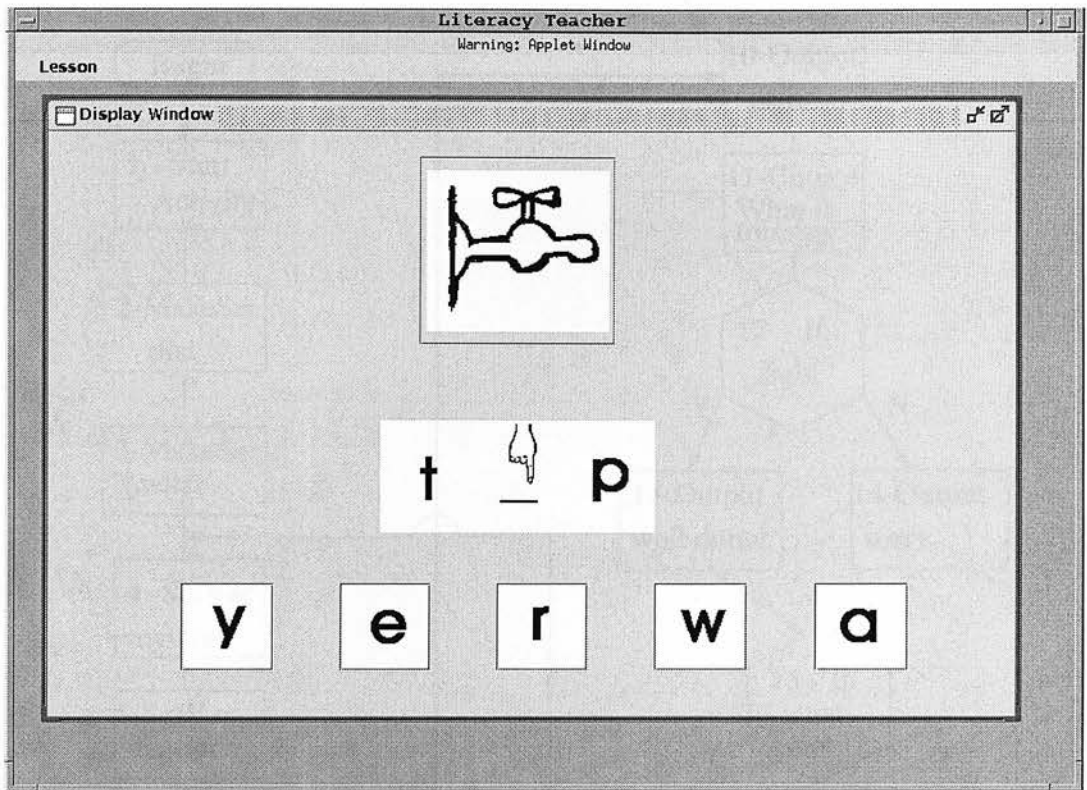


Figure 8.5: Choice for *The Missing Sound*

The flow-chart diagram for the activity is shown in Figure 8.6 and discussed below.

1. The first step is to provide general information on the activity with the *Start-Activity Object*. Its main fields are:

Type of Activity: test

Subject Domain: 1.1.1

Complexity: 2

Skill Explored: Code Knowledge

2. The second step is to obtain a set named *tlist* with 10 words that belong to the curriculum item desired (1.1.1 in this case), using the *Make-Set Object*. The main fields in the object are:

Set Name: tlist

Total Elements: 10

Source: Words

Select Expression: SelectWords:Kind:1.1.1

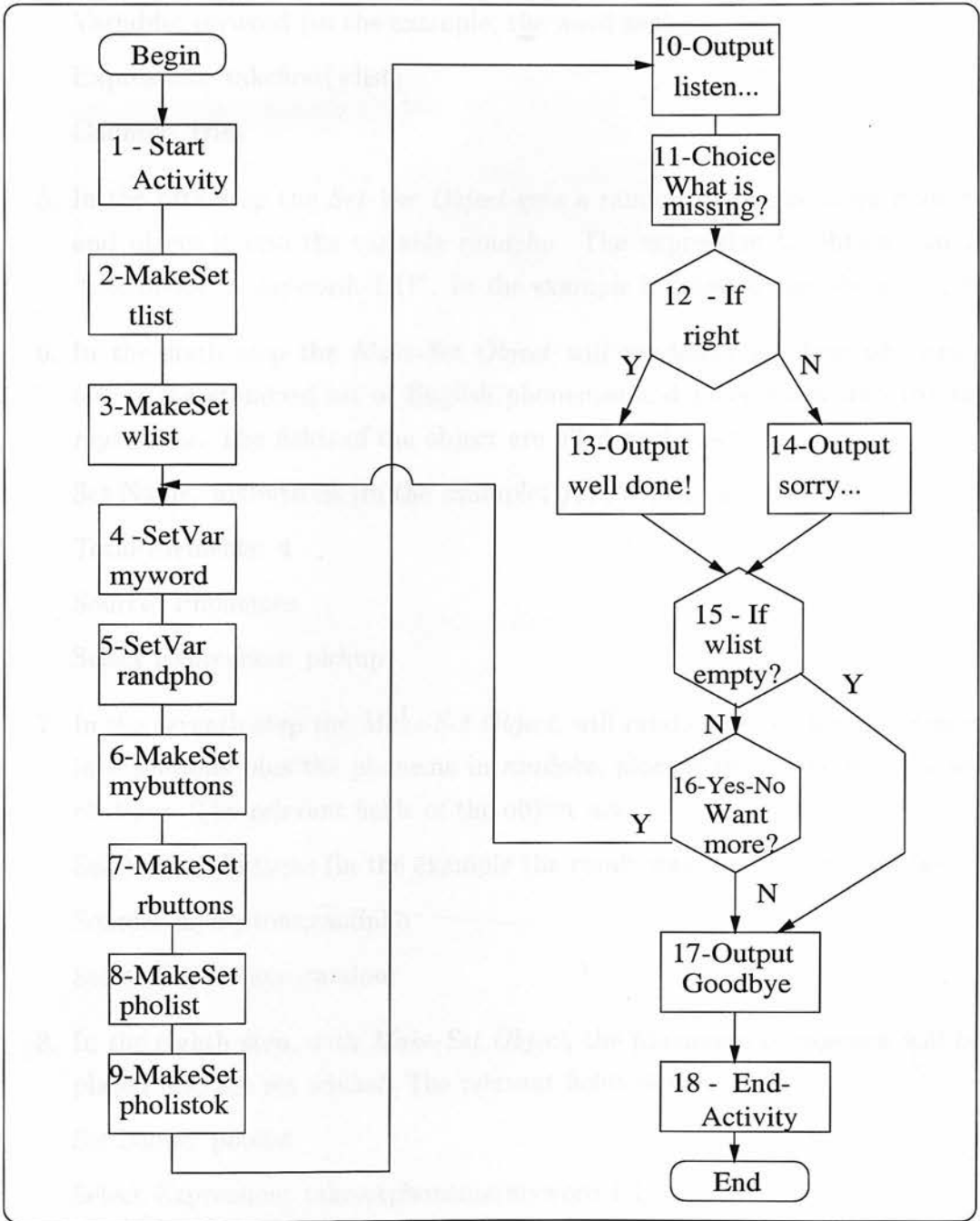


Figure 8.6: Flow chart for the *Missing Sound* activity

3. In the third step, the set obtained in step 2 is randomised producing a set named *wlist*. This step is necessary to allow the words to appear in a different order each time the activity is started.
4. In the fourth step the *Set-Var Object* is used to get the first word of the set into the variable *myword*. The main fields of the object are:

Variable: myword (in the example, the word *tap*)

Expression: takefirst(wlist)

Counter: tries

5. In the fifth step the *Set-Var Object* gets a random phoneme from *myword* and places it into the variable *randpho*. The expression to obtain that is “takephoneme(myword:-1:1)”. In the example it selected the phoneme /a.

6. In the sixth step the *Make-Set Object* will randomly get four phonemes out of the standard set of English phonemes and place them into the set *mybuttons*. The fields of the object are filled as follows:

Set Name: mybuttons (in the example: /e, /r, /w, /y)

Total Elements: 4

Source: Phonemes

Select Expression: pickup

7. In the seventh step the *Make-Set Object*, will randomise the four phonemes in *mybuttons* plus the phoneme in *randpho*, placing the result into the set *rbuttons*. The relevant fields of the object are:

Set Name: rbuttons (in the example the result was: /y, /e, /r, /w, /a)

Source: mybuttons;randpho

Select Expression: random

8. In the eighth step, with *Make-Set Object*, the phonemes of *myword* will be placed into the set *pholist*. The relevant fields are:

Set Name: pholist

Select Expression: takesetphoneme(myword:1:10)

9. In the ninth step the *Make-Set Object* is used for replacing the phoneme in *pholist* which has the same value of *randpho* with a *space*. The resulting set is *pholistok* which have all the phonemes of *myword* except the one in *randpho*. The fields of the object are:

Set Name: pholistok

Source: pholist

Select Expression: Replace:randpho: (replace by space)

10. In the tenth step the *Output Object* is used to command the computer to ask the student to listen to the selected word. The “Text to speak” field of the object is codified as “Listen to this word: ”+myword.
11. In the eleventh step the *Choice Object* asks the student to point out the sound that is missing in the word. The Figure 8.5 is an illustration of this step. There the word “tap” is depicted, and a sequence of graphemes are presented below it. The computer will say “What is the missing sound of tap?” and wait for the student to double click one of the graphemes shown. The relevant fields of the object are:
- Element list: rbuttons
- Text to speak: /”What is the missing sound of ”+myword
- Timeout: 40
12. In the twelfth step the *If Object* evaluates the choice of the student. The relevant fields are:
- Expression: randpho=BarElement (to check if the chosen button corresponds to the phoneme in the *randpho* variable)
- Yes Branch: 13
- Yes Counter: ok
- No Branch: 14
- No Counter: notok
13. In step thirteen the *Output Object* is used to congratulate the student, because the answer was right (“Congratulations, you were correct”).
14. In the fourteenth step however, *Output Object* will tell the student the answer was not right (“Sorry, the missing sound is /”+randpho).
15. In the fifteenth step the *If Object* checks whether there are still words in the word list (in *wlist*). The fields are:
- Expression: SetLength:wlist<1
- Yes Branch: 17
16. In the sixteenth step the *Yes-or-No Object* will ask the student whether he or she should be presented with another word or not. The relevant fields are:

Question: "Would you like to try again?"

Waiting time: 20

Default: no

Yes Branch: 4

17. The seventeenth step is to say goodbye to the student, with the *Output Object*.
18. In the last step, with the *End-Activity Object* the system shows the final banner and the activity is finished.

8.2.2 Adding Sounds

Another interesting way to develop phonological awareness is through the observation that some words are part of other words and that the addition of sounds to some words can result in new words which contain the former ones. This sort of practice is used in various points of the curriculum of Jolly Phonics. This activity is implemented with the system in the form of an *example*. The Figure 8.7 shows one example of the screen generated by the computer while carrying out this activity. In the example the word *at* has been selected and the computer is showing that it can build the word *mat* by placing the phoneme *m* before *at*.

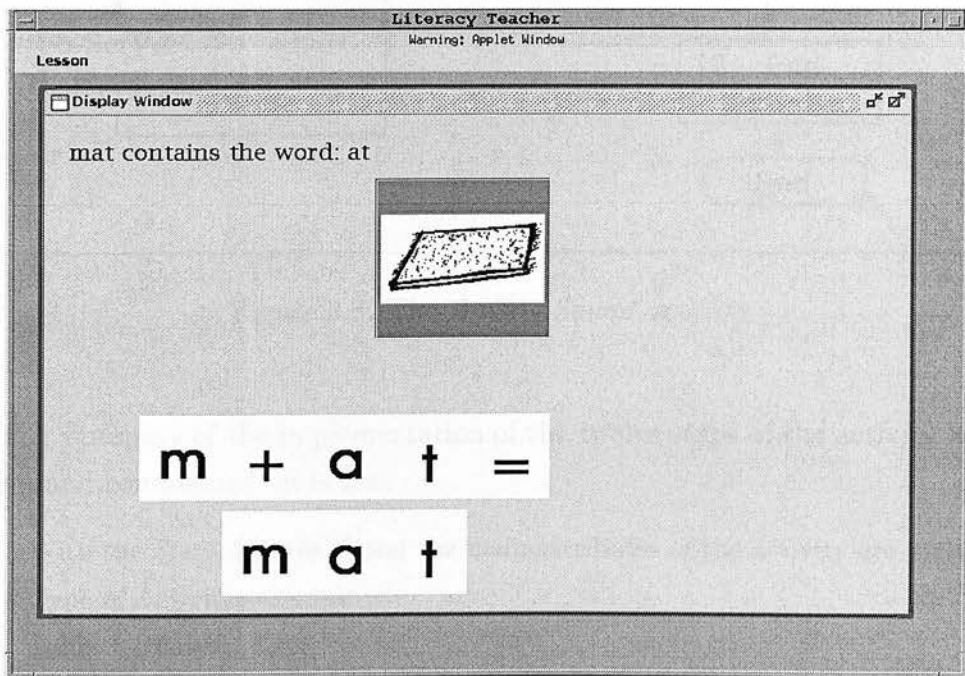


Figure 8.7: The main screen in the *adding sounds* activity

The flow-chart diagram for the activity is shown, in Figure 8.8 and discussed below.

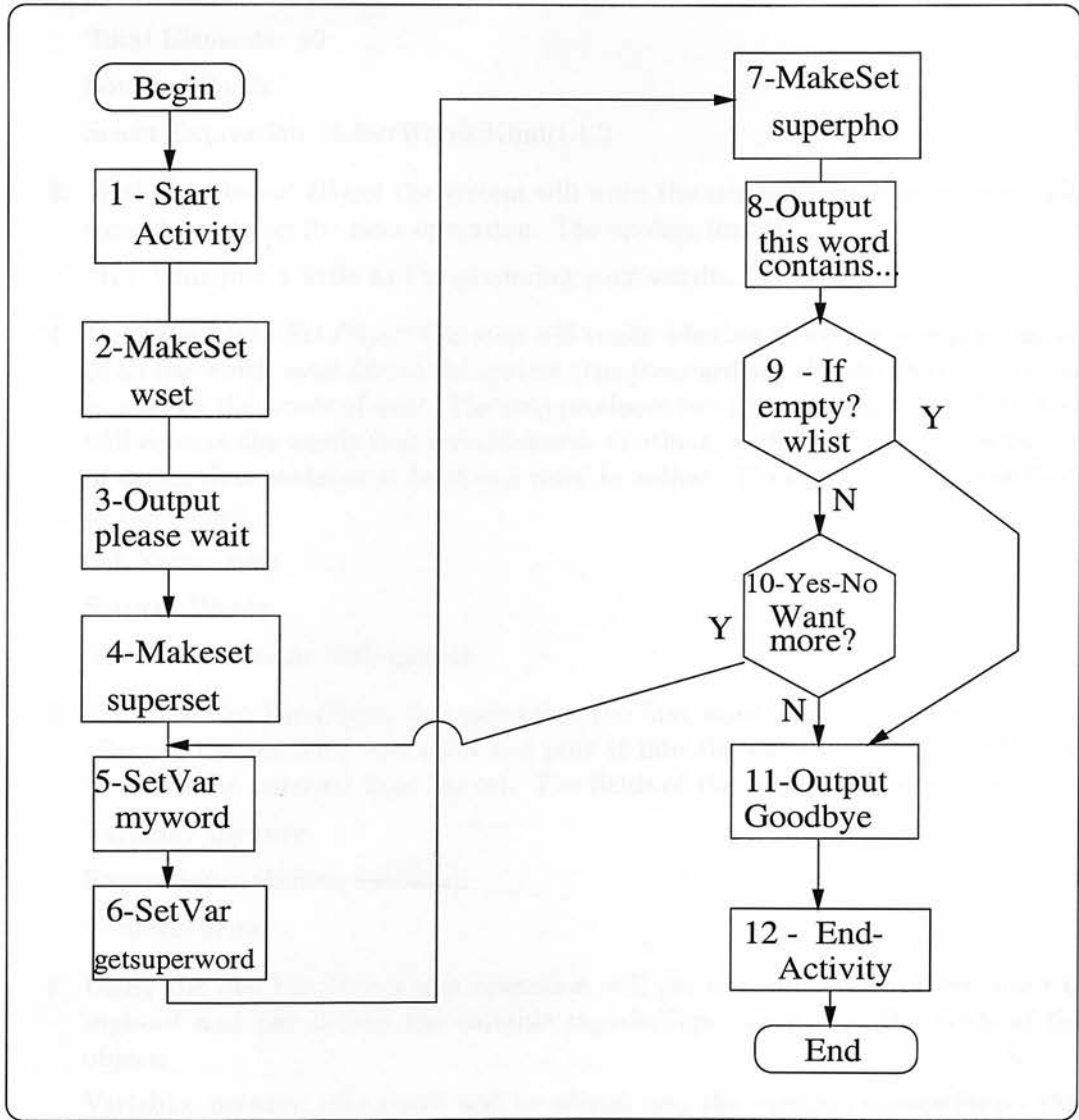


Figure 8.8: The *Adding Sound* activity

The summary of the implementation of the twelve steps of the activity is now shown and commented on below:

1. With the *Start-Activity Object* the main attributes of the activity are declared:
 Type of Activity: example
 Subject Domain: 1.1.2
 Complexity: 2
 Skill Explored: Blending

2. With the *Make-Set Object* 40 of the words of curriculum item 1.1.2 are selected and stored into the set *wset*.
 Set Name: *wset*
 Total Elements: 40
 Source: Words
 Select Expression: `SelectWords:Kind|1.1.2`

3. With the *Output Object* the system will warn the student that it is going to take some time doing the next operation. The spoken text is:
 “Hi! Wait just a little as I’m preparing your words...”

4. With the *Make-Set Object* this step will verify whether there are words in the set of all the words available to the system (the standard set *Words*) that are formed in part by the words of *wset*. The step produces two sets of words: *wsetRes* which will contain the words that are *sub-words* of others; and *wsetSup* which is the set of words that contains at least one word in *setRes*. These are the relevant fields of the object:
 Set Name: *wset*
 Source: Words
 Select Expression: `SetSuperSet`

5. With the *Set-Var Object* this step takes the first word of the resultant set (*wsetRes*) of the previous operation and puts it into the variable *myword*. The first word will be removed from the set. The fields of the object are as follows:
 Variable: *myword*
 Expression: `takefirst(wsetRes)`
 Counter: *tries*

6. Using the *Set-Var Object* this operation will get one *superword* of the word in *myword* and put it into the variable *mywordSup*. These are the fields of the object:
 Variable: *myword* (the result will be placed into the variable *mywordSup* - this extension is automatically generated)
 Expression: `getsuperword(wsetSup)`

7. The *Make-Set Object* is used in this step to generate a set of graphemes that is the product of the comparison of two words, where the second contains the first one. The result of the operation is a set of graphemes with the phonetic sum of the second word in relation to the first one (e.g. for the words *ring* and *spring* the system will generate “s p + r i n g = s p r i n g”). These are the fields of the object to obtain this result:
 Set Name: *superpho*
 Select Expression: `SuperPhonemes:myword:mywordSup`

8. With the *Output Object* the system will show the sequence of graphemes and say a message about the relationship of the words in question. The relevant fields are:

Text to speak: mywordSup+" contains the word: "+myword

Text to write: superpho

9. With the *If Object* the computer will check if there are still words in the set of words. If so it will proceed to the following step, otherwise it will jump to step 11. The fields are:

Expression: SetLength|wsetRes<1

Yes Branch: 11

10. The *Yes-or-No Object* is used in this step to give the student the option to see another example or to leave the activity. The fields are:

Question: "Would you like to know another super-word?"

Waiting time: 100

Default: no

Yes Branch: 5

11. The *Output Object* is used here to say goodbye to the student:

Text to speak: "Bye, I would like to see you again"

12. The *End-Activity Object* finishes the activity.

8.3 THRASS Applications

8.3.1 Identifying Words

THRASS is a method centred on its *Thrasschart* and *Thrasswords* chart, and it goes straight to the point of the spelling of the words. Two activities recommended in the THRASS method have been selected to demonstrate the ability of the model proposed here to represent different PA approaches. The first of these is the Identifying Words test, which is a very basic task that verifies whether the student is able to indicate the written representation of a word that is said to him or her, and shown among other words.

Figure 8.9 presents the screen the computer shows while executing this activity. On this example the student is presented with the picture of a *cap* and a set of buttons with written words to choose from, including the written representation for *cap*.

The flow-chart diagram for the activity is illustrated in Figure 8.10 and discussed below. The activity is implemented in the system in 16 steps, as presented in the following summary including the important fields of each object used.

1. The *Start-Activity Object* is used to declare the main attributes of the activity:

Type of Activity: exercise

Subject Domain: 1.1.1

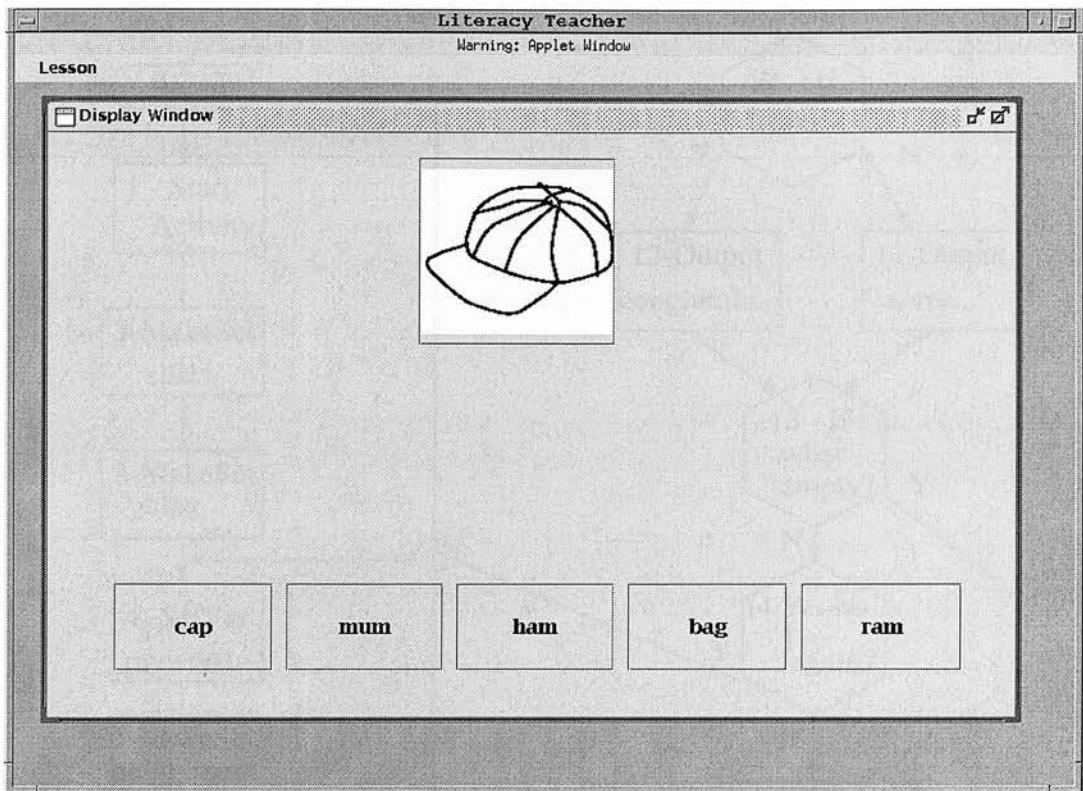


Figure 8.9: Choice for *Identifying Words*

Complexity: 2

Skill Explored: Code Knowledge

2. The *Make-Set Object* is used to select the words of the vocabulary that belong to curriculum item 1.1.1. and place them in the set *slist*. The fields of the object are:

Set Name: *slist*

Source: Words

Select Expression: `SelectWords:Kind:1.1.1`

3. The *Make-Set Object* is used to randomly pick ten elements out of the list *slist* selected in step 2, so the set of words to be used in the activity could be different each new time the activity is entered. The ten randomly chosen words will be placed in the set *wlist*. The fields of the object are:

Set Name: *wlist*

Total Elements: 10

Source: *slist*

Select Expression: `random`

4. The *Set-Var Object* is used to get the first element of the list of words - it will be taken out of the list *wlist* and placed in the variable *myword*. The counter *tries* will be incremented by one unit. The fields of the object follow:

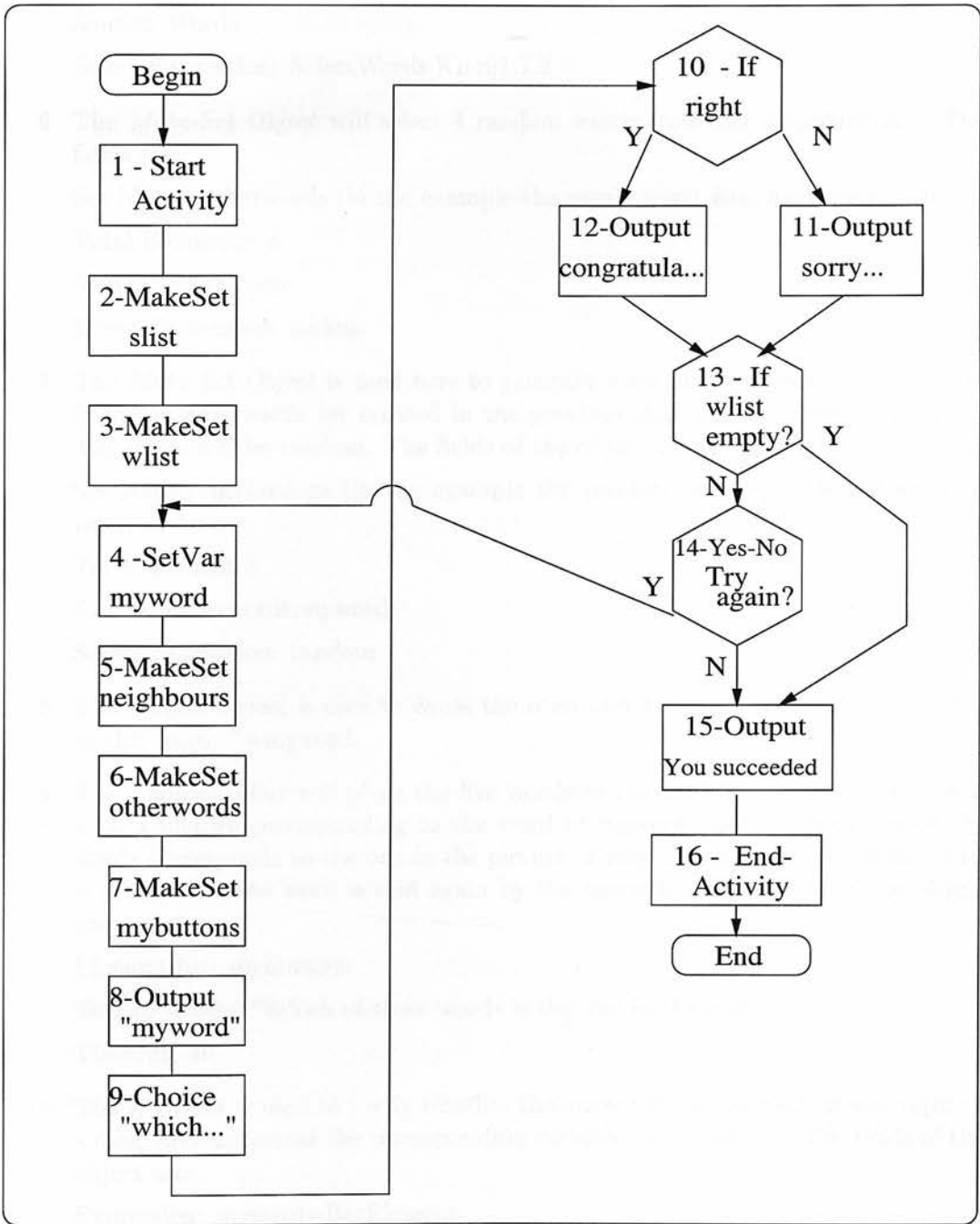


Figure 8.10: Flow chart: the *Identifying Words* activity

Variable: myword (in the example the word was *cap*)

Expression: takefirst(wlist)

Counter: tries

- The *Make-Set Object* will select the words that belong to the “neighbour” curriculum item 1.1.2. The fields are as follows:

Set Name: neighbours

Source: Words

Select Expression: SelectWords:Kind|1.1.2

6. The *Make-Set Object* will select 4 random words from the *neighbour* set. The fields are:
Set Name: otherwords (in the example the words were: *bag, ham, mum, ram*)
Total Elements: 4
Source: neighbours
Select Expression: pickup
7. The *Make-Set Object* is used here to generate a set of five words, with the four from the *otherwords* set created in the previous step and the word in *myword*. The order will be random. The fields of the object are as follows:
Set Name: mybuttons (in the example the product was *cap, mum, ham, bag, ram*)
Tot Elements: 5
Source: otherwords;myword
Select Expression: random
8. The *Output Object* is used to cause the computer to say to the student: "Listen to this word: "+myword.
9. The *Choice Object* will place the five words in the set *mybuttons* on the screen with a picture corresponding to the word in *myword*, asking which one of the words corresponds to the one in the picture. Every time the picture of the word is pointed to, the word is said again by the computer. The fields of the object are:
Element list: mybuttons
Text to speak: "Which of these words is the one in the picture?"
Timeout: 40
10. The *If Object* is used to verify whether the answer from the student was right or wrong, and increment the corresponding variable (*ok* or *notok*). The fields of the object are:
Expression: myword=BarElement
Yes Branch: 12
Yes Counter: ok
No Counter: notok
11. The *Output Object* will convey the message corresponding to a wrong answer: "Sorry, the word "+BarElement+" is not in the picture. It is "+myword
12. The *Output Object* is used here to praise the student for the right answer: "Congratulations, you are right!"

13. The *If Object* will check if the number of tries has reached 10. If yes, the next step will be the good-bye message in step 15. The fields of the object are filled as follows:
 Expression: tries=10
 Yes Branch: 15
14. The *Yes-or-No Object* is used to ask the student if he or she wants to see another set of words. The fields of the object are:
 Question: "Would you like to try it again?"
 Waiting time: 20
 Default: no
 Yes Branch: 4
15. The *Output Object* is used to cause the computer to say: "You succeeded in "+ok+" tries out of "+tries. The variable *tries* has the total of words examined and the variable *ok* has the number of right answers.
16. With the *End-Activity* object the activity is finished.

8.3.2 Alternative spellings

The second activity representing the THRASS method is directly related to spelling. It proposes different possible spellings for a word, and expects the student to point out the right option. This activity might be considered by some teachers as inappropriate because it can introduce misleading and incorrect spellings. However it is one activity from THRASS and it aims to show that it is possible to have different representations for one word, but that there is just one that was elected to represent the word in its written form. The author is not claiming that this or any of the other activities implemented with the system are the best representatives of phonological awareness developing activities. As stated before, the aim of this chapter is to exercise the abilities of the system to deal with different requirements of recognised PA methods.

This activity is implemented in the system in the form of a test. The Figure 8.11 shows an example of what the screen generated by this activity looks like. In the example the picture of a palm of a hand is presented along with five buttons each one with a possible spelling for the word. The student is expected to point out the right one.

Figure 8.12 contains the flowchart correspondent to the activity. Its steps are summarised below.

1. The *Start-Activity Object* used to declare the attributes of the activity:
 Type of Activity: test

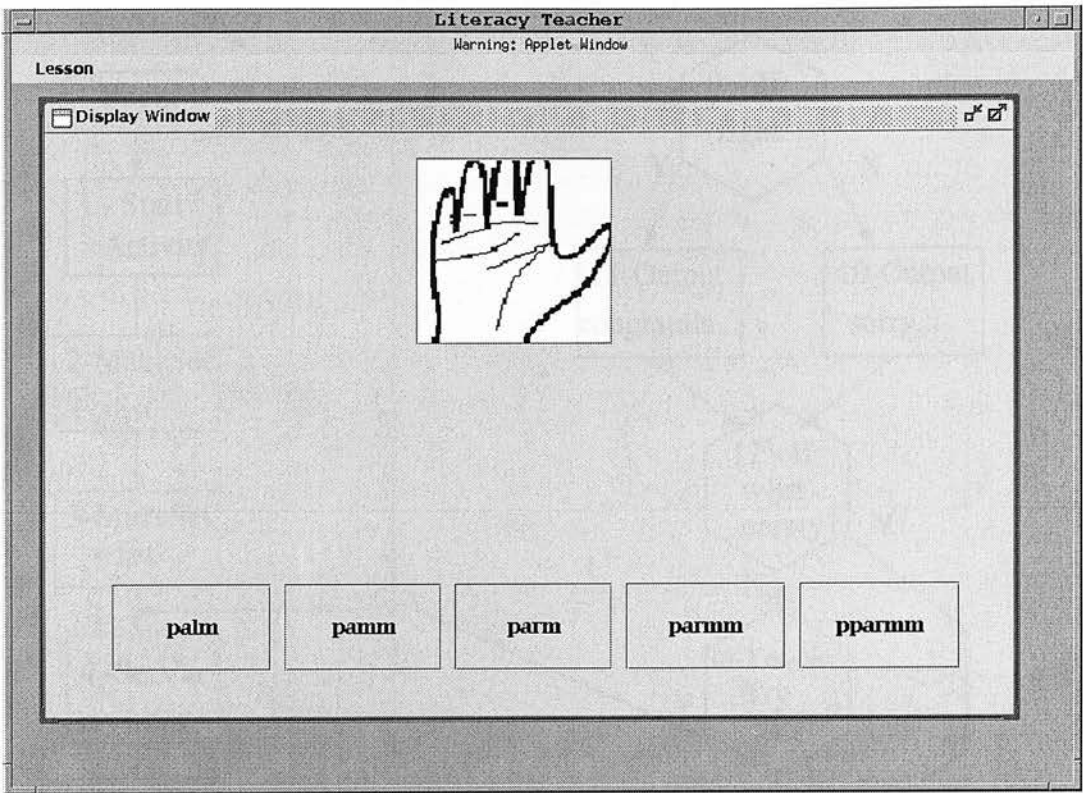


Figure 8.11: The main screen in the *alternative spellings* activity

Subject Domain: 1.1.1

Complexity: 2

Skill Explored: Code Knowledge

2. The *Make-Set Object* is used here to select the words of curriculum item 1.1.1 and to store them into the set *slist*. The fields of the object are:
 - Set Name: *slist*
 - Source: Words
 - Select Expression: `SelectWords:Kind|1.1.1`
3. The *Make-Set Object* is used to pick out 10 randomly chosen words from the *wlist* set, obtained in the previous step. The fields are:
 - Set Name: *wlist*
 - Tot Elements: 10
 - Source: *slist*
 - Select Expression: `random`
4. The *Set-Var Object* is used to take the first word of the resultant set (*wlist*) of the previous operation and put it in the variable *myword*. The first word will be removed from the set. The fields of the object are:
 - Variable: *myword* (in the example it was *palm*)

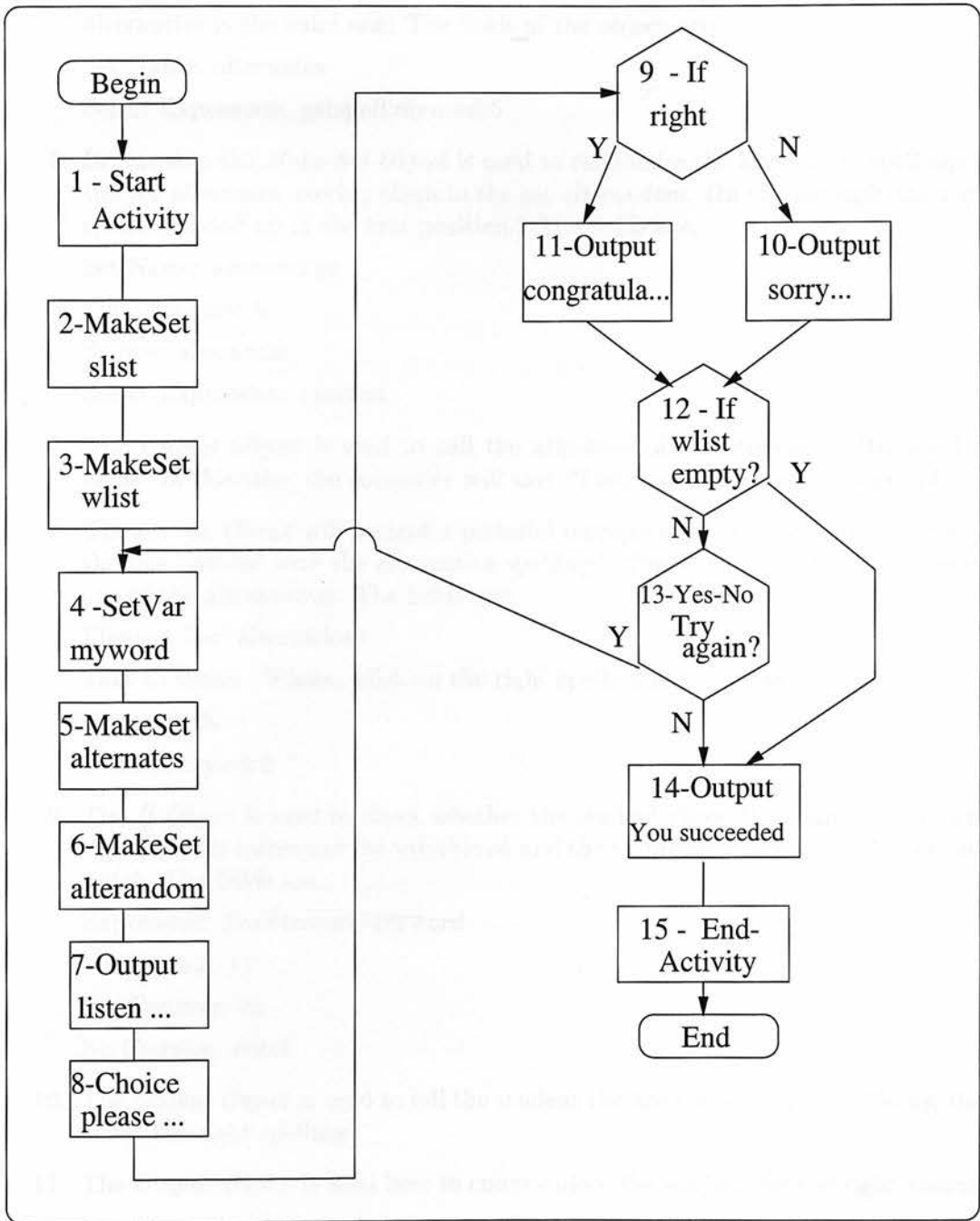


Figure 8.12: Flow chart: the *Alternative Spellings* activity

Expression: `takefirst(wlist)`

Counter: `tries`

- The *Make-Set Object* will get five alternative spellings for the word in *myword* and place them into the set *alternates*. In order to get the alternative spellings the system has a special function that chooses randomly one out of all the possible representations of a phoneme for each one of the phonemes of the word. The first

alternative is the valid one. The fields of the object are:

Set Name: alternates

Select Expression: getspell:myword:5

6. In this step the *Make-Set Object* is used to randomise the alternative spellings in the set *alternates*, storing them in the set *alterandom*. (In the example the right spelling ended up in the first position.) The fields are:

Set Name: alterandom

Tot Elements: 5

Source: alternates

Select Expression: random

7. The *Output Object* is used to call the attention of the student to the word in focus. In this step, the computer will say: "Listen to this word: "+myword
8. The *Choice Object* will present a pictorial representation of the word along with the five buttons with the alternative spellings. The student is invited to choose one of the alternatives. The fields are:

Element list: alterandom

Text to speak: "Please, click on the right spelling for "+myword

Timeout: 30

Picture: myword

9. The *If Object* is used to check whether the student chose the right answer. The right answers increment the variable *ok* and the wrong ones increment the variable *notok*. The fields are:

Expression: BarElement=myword

Yes Branch: 11

Yes Counter: ok

No Counter: notok

10. The *Output Object* is used to tell the student the answer is not right: "Sorry, this is not the right spelling"
11. The *Output Object* is used here to congratulate the student for the right answer.
12. The *If Object* will verify whether there are still words in the set *wlist*. The fields are

Expression: SetLength|wlist| < |1

Yes Branch: 14

13. The *Yes-or-No Object* is used to to ask the student if he or she wants to see another word along with its alternative spellings. The fields are:

Question: "Would you like to try it again?"

Waiting time: 20

Default: no

Yes Branch: 4

14. The *Output Object* will say the goodbye message with an evaluation of the student: "You succeeded in "+ok+" tries and made "+notok+" mistakes"
15. The *End-Activity Object* finishes the activity.

8.4 Summary of the Case Studies

In this chapter six different activities from the three methods used for testing the system's ability in representing phonological awareness tasks were implemented. The existing objects as they have been implemented to date, provided the means to implement the majority of the tasks. However, for some of them some objects had to be extended with new functions (the Set-Var and the Make-Set objects) or with variations in the form of presentation (the Match and the Choice objects). For instance, for the activity shown in Subsection 8.3.2, Alternative spellings, a function named *getspell* had to be developed to offer the alternative spellings required by the task. This fact was predicted while the system was being modelled. A lot of interaction with experts is expected before the system gets to a point where any activity would be straightforwardly implemented with only the available objects and functions. It is rather more likely that there will always be further functions that could be added to support the teaching activities. However, as a test for the prototype and for the modelling approach, the authoring tools and the teaching module have produced good results.

The vocabulary, as seen in the curriculum specifications, is used in the Short Phonix, using the purely phonetic vocabulary that provides a simplified method using the Short Phonix vocabulary. The curriculum specifies that the number of words in the Short Phonix vocabulary should be 100 words, with 50 words distributed across the curriculum.

Chapter 9

Teaching Strategy Application

This chapter describes a simplified literacy teaching method, named Short Phonix, developed to illustrate the teaching strategy scheme proposed.

9.1 General Description of Short Phonix

Short Phonix is partially based on the knowledge tree of the Phono-Graphix method, up to the curriculum item 1.2.3. Its teaching activities are similar to the ones explained earlier in Chapter 8, with some new activities. The curriculum of Short Phonix is shown below.

1 - Reading: Code Knowledge, Segmenting, Blending.

&source=foodonly.

1.1 - Basic Code. &exclude.

VC. CVC.

1.1.1 - Fat Cat Sounds.

[/f,/k,/s,/p,/m,/t] {/a,/o,/aa} {/t,/p,/m,/s}.

&source=foodcommon.

1.1.2 - Bug On Jug Sounds.

[/r,/h,/b,/jh,/d,/m,/t] {/uh,/i,/a} {/g,/d,/m}.

&source=foodcommon.

1.1.3 - Ben Bon Sounds.

[/h,/b,/w,/l,/n,/f,/z,/s]{/e,/o,/i,/uh}{/l,/n,/t,/z,/b,/p}.

&source=foodcommon.

1.2 - Adjacent Consonants: Auditory Processing.

1.2.1 - VCC Words.

VCC.

&source=foodcommon.

1.2.2 - CVCC Words.

CVCC.

1.2.3 - CCVC Words.

CCVC.

{X}{/r}{X}{X}.

{X}{X}{/r}{X}.

9.2 Short Phonix Vocabulary

The vocabulary, as seen in the curriculum specification, is based on the “food” selection, using the purely thematic vocabulary where possible (*foodonly*), and sometimes using the thematic plus common-words vocabulary (*foodcommon*), when the number of available words for the curriculum item is too small. This approach produced 361 words distributed across the curriculum as follows:

- 1.1 - 108 words
- 1.1.1 - 12 words
- 1.1.2 - 8 words
- 1.1.3 - 40 words
- 1.2.1 - 22 words
- 1.2.2 - 115 words
- 1.2.3 - 56 words

9.3 Short Phonix Knowledge Tree and Activities

Figure 9.1 shows the knowledge tree for Short Phonix with the activities attached to each branch. There are 63 teaching activities in the Short Phonix programme. They were generated from 15 different activities, but as they are applied in diverse points of the curriculum they would present different words and produce different results. Tables 9.1 and 9.2 relate all the Short Phonix programme activities to their codes, names, curriculum items and complexities.

Figure 9.1 also provides a good indication of the basic strategy the system would use while selecting the activity to be proposed to the student: the highest priority would be given to the node in the bottom left leaf of the tree (Fat Cat Sounds), the lessons of the lowest complexity coming first, then the examples, then the exercises and finally the tests. Next on the list would be the lessons of second lowest complexity, then the examples, and so on, up to the end of the activities of that node. Then the next node to the right would be explored, and so forth. However, this approach is just a basic guideline, and does not take into consideration the results of the student and the strategies of the expert teacher.

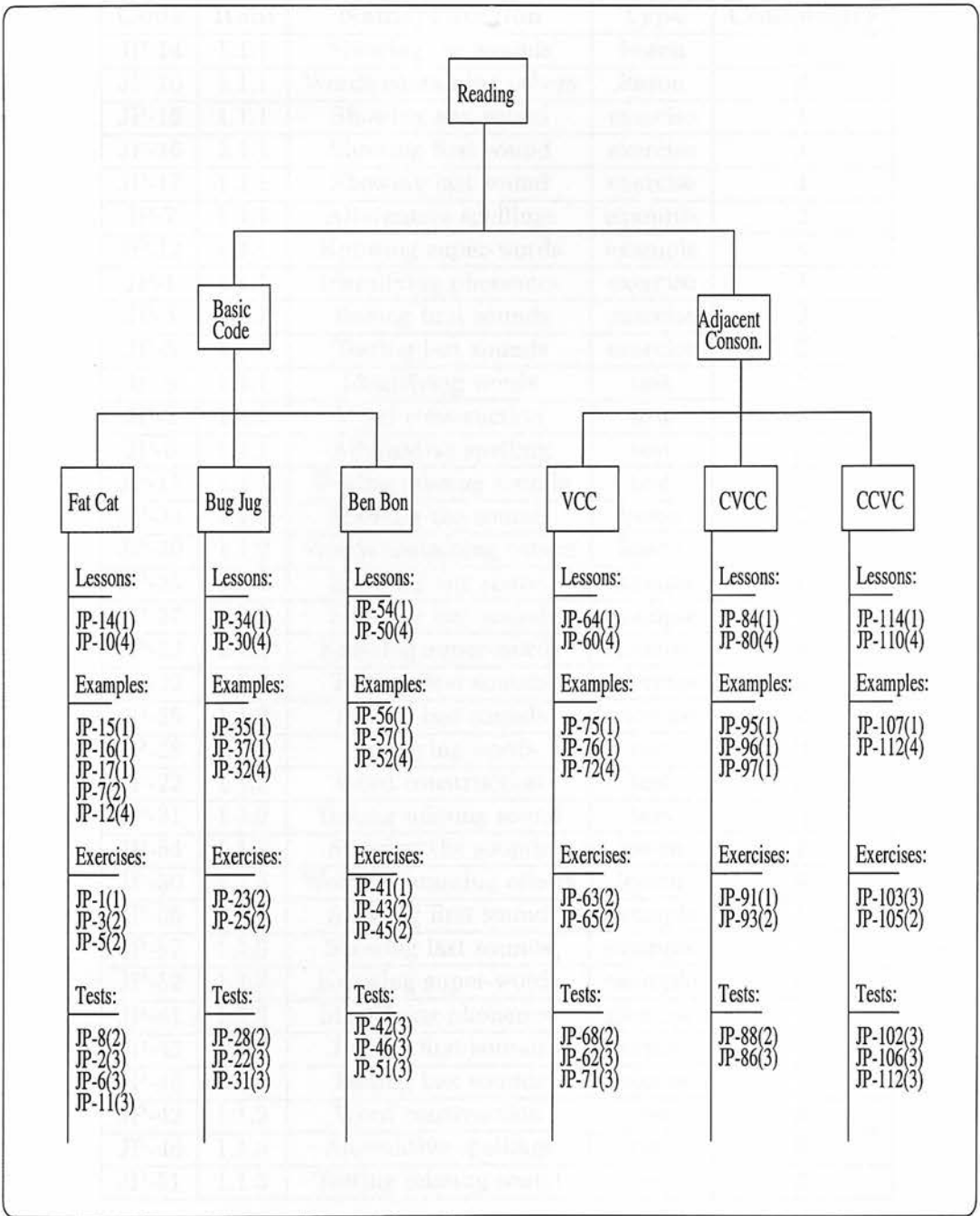


Figure 9.1: The Short Phonix knowledge tree

Code	Item	Name/Function	Type	Complexity
JP-14	1.1.1	Showing the sounds	lesson	1
JP-10	1.1.1	Words containing others	lesson	4
JP-15	1.1.1	Showing any sound	exercise	1
JP-16	1.1.1	Showing first sound	exercise	1
JP-17	1.1.1	Showing last sound	exercise	1
JP-7	1.1.1	Alternative spellings	example	2
JP-12	1.1.1	Knowing super-words	example	4
JP-1	1.1.1	Identifying phonemes	exercise	1
JP-3	1.1.1	Testing first sounds	exercise	2
JP-5	1.1.1	Testing last sounds	exercise	2
JP-8	1.1.1	Identifying words	test	2
JP-2	1.1.1	Word construction	test	3
JP-6	1.1.1	Alternative spelling	test	3
JP-11	1.1.1	Testing missing sounds	test	3
JP-34	1.1.2	Showing the sounds	lesson	1
JP-30	1.1.2	Words containing others	lesson	4
JP-35	1.1.2	Showing any sound	example	1
JP-37	1.1.2	Showing last sound	example	1
JP-32	1.1.2	Knowing super-words	example	4
JP-23	1.1.2	Testing first sounds	exercise	2
JP-25	1.1.2	Testing last sounds	exercise	2
JP-28	1.1.2	Identifying words	test	2
JP-22	1.1.2	Word construction	test	3
JP-31	1.1.2	Testing missing sound	test	3
JP-54	1.1.3	Showing the sounds	lesson	1
JP-50	1.1.3	Words containing others	lesson	4
JP-56	1.1.3	Showing first sound	example	1
JP-57	1.1.3	Showing last sounds	example	1
JP-52	1.1.3	Knowing super-words	example	4
JP-41	1.1.3	Identifying phonemes	exercise	1
JP-43	1.1.3	Testing first sounds	exercise	2
JP-45	1.1.3	Testing last sounds	exercise	2
JP-42	1.1.3	Word construction	test	3
JP-46	1.1.3	Alternative spellings	test	3
JP-51	1.1.3	Testing missing sound	test	3

Table 9.1: Activities for Short Phonix

Code	Item	Name/Function	Type	Complexity
JP-64	1.2.1	Showing the sounds	lesson	1
JP-60	1.2.1	Words containing others	lesson	4
JP-75	1.2.1	Showing any sounds	example	1
JP-76	1.2.1	Showing first sound	example	1
JP-72	1.2.1	Knowing super-words	example	4
JP-63	1.2.1	Testing first sounds	exercise	2
JP-65	1.2.1	Testing last sound	exercise	2
JP-68	1.2.1	Identifying words	test	2
JP-62	1.2.1	Word construction	test	3
JP-71	1.2.1	Testing missing sound	test	3
JP-84	1.2.2	Showing the sounds	lesson	1
JP-80	1.2.2	Words containing others	lesson	4
JP-95	1.2.2	Showing any sound	example	1
JP-96	1.2.2	Showing first sound	example	1
JP-97	1.2.2	Showing last sound	example	1
JP-91	1.2.2	Identifying phonemes	exercise	1
JP-93	1.2.2	Testing first sound	exercise	2
JP-88	1.2.2	Identifying words	test	2
JP-86	1.2.2	Alternative spellings	test	3
JP-114	1.2.3	Showing the sounds	lesson	1
JP-110	1.2.3	Words containing others	lesson	4
JP-107	1.2.3	Showing last sound	example	1
JP-112	1.2.3	Showing super-words	example	4
JP-103	1.2.3	Testing first sounds	exercise	3
JP-105	1.2.3	Testing last sound	exercise	2
JP-102	1.2.3	Word construction	test	3
JP-106	1.2.3	Alternative spellings	test	3
JP-112	1.2.3	Word construction	test	3

Table 9.2: Activities for Short Phonix - continuation

9.4 Short Phonix Teaching Strategy

This is the point where the teacher expertise will help to decide where to start and what to do after an activity is completed. At the beginning it is useful to test the student in order to adjust the level of the activities to his or her current knowledge. Each time the student completes an activity an evaluation is carried out, even when he or she is just visiting a lesson or an example. The result of the evaluation is recorded by the system, so that after an activity is carried out a new configuration of the student model is obtained. The strategy for teaching should be designed so that activities that are appropriate to the student needs would be proposed and so that the student with good results would soon gain access to more advanced tasks. It should, however, allow more simple activities to be rerun and reinforced every time the student fails repeatedly in one area of the curriculum.

In this section an experimental strategy for the Short Phonix programme is presented and discussed. The strategy is defined in terms of the TSDL - Teaching Strategy Definition Language, as described early in Section 7.4 and formally stated in Appendix F.2. Appendix H contains the complete code for the proposed strategy.

The strategy adopted, although long, is simple and repetitive and takes into consideration some general aspect of the programme, but it is not specific in relation to any particular pedagogical issue. It has, however, the means to let the student advance or go back to earlier stages in the curriculum when necessary.

Its main device to let the student progress in the curriculum is the test whether the exercise and test averages are good,¹ as in the following code:

```
if exercising(Fat Cat Sounds) greater than enough
    then disable exercise(Fat Cat Sounds).
if testing(Fat Cat Sounds) greater than enough
    then disable lesson(Fat Cat Sounds),
        disable example (Fat Cat Sounds),
        disable exercise(Fat Cat Sounds).
if testing(Fat Cat Sounds) greater than good
    then disable test(Fat Cat Sounds).
```

¹The words *poor*, *enough*, *good*, and *excellent* are reserved words of the TSDL language and represent *degrees* to evaluate the student's performance. The correspondence for the degrees are: *poor*: up to 49%; *enough*: between 50% and 69%; *good*: between 70% and 89%; *excellent*: above 90%.

The strategy interpretation is based in a ranking for the selection of the activity to be applied next. These tests will disable the activities of the curriculum item in question as soon as the student results get *enough* or *good*. The ranking algorithm then will have no option but to choose activities from the next curriculum item.

However, if the student begins to fail on tests and exercises, lowering the average, there is another device that will re-enable the tests and exercises of the previous levels, so the student can return to practise more elementary subjects. An example of the use of such devices follows:

```
if exercising (V C C Words) less than enough
  then enable example(V C C Words),
       enable exercise(V C C Words),
       enable lesson(V C C Words),
       enable test(Bug On Jug Sounds),
       enable exercise(Bug On Jug Sounds),
       enable test(V C C Words).
```

Another important complementary device of the strategy is the suspension of the *winner* activity from running for some subsequent cycles, so other activities of the same level or slightly superior levels can win the race to be used and guarantee that exercises and tests are executed before a high priority activity is elected to run again. In this case, the suspension for 15 runs worked quite well, but this can be changed when the teacher feels that the subjects should be stressed more before progressing. However, when the student gets to the higher activities of the curriculum the suspension of the eligible activities may result in a lack of alternatives, leaving the algorithm the option of random selection of an activity, which may not be appropriate. For this reason at the end of the strategy rules there is a statement to reduce the suspension factor whenever the student is approaching the end of the programme:

```
if exercising (C C V C Words) greater than enough or
  testing (C C V C Words) greater than enough
  then suspend winner(5).
```

Finally, the strategy just shown illustrates that the knowledge of the expert can be scaled up or down whenever one or more of the curriculum items appears to be more or less relevant to the general progress of the student. The declarations

expertise(Bug On Jug Sounds)=90.
expertise(Auditory Processing)=80.
expertise(Fat Cat Sounds, Blending)=70.

will determine that lower degrees will be considered as *good* or as *enough* for those particular subjects or skills. For example, after the third declaration above (expertise(Fat Cat Sounds,Blending)=70) is executed, a score of 35 points (50% of 70) in *Blending* skills in *Fat Cat Sounds* will be considered *enough* (normally only a performance of 50% would be considered *enough*).

9.5 Strategy Simulation

In order to evaluate the teaching strategy definition scheme and to observe the progression of the students under the strategy rules used as an example for Short Phonix, we developed a tool that replaces the Teaching Module by simulating the results of the activities and communicating them to the Diagnostic Module.

The next three figures are the result of the simulation of the progress of three students, the first with bad performance, always getting a mark between 2 and 4 (Figure 9.2), the second with good performance, achieving marks ranging from 8 to 10 (Figure 9.3), and the third with intermediate performance, with marks from 4 to 9 (Figure 9.4). The graphs contain the codes for the activities in the y-axis ordered by increasing complexity, and the sequence of trials of activities in the x-axis (up to 100 trials). At the intersection point of the two axis there are the performances of the student in each particular trial. For example, in Figure 9.2, the first trial was with the activity *jp-14* and the student scored 3; the second trial was *jp-17* and the score was 4; the third trial was *jp-15* and the score was 3; and so on.

The three graphs show that the Diagnostic Module, while interpreting the teaching strategy rules proposed for Short Phonix, reacts in a way that could be compared to how a human teacher would with students of differing performance. In the first case (Figure 9.2), dealing with a student who performs badly in all activities, the system tries all the simplest tasks in cycles, without progressing to the more complex ones. In the second case (Figure 9.3), as the student supposedly has good results in every task he or she tries, the system allows the student to

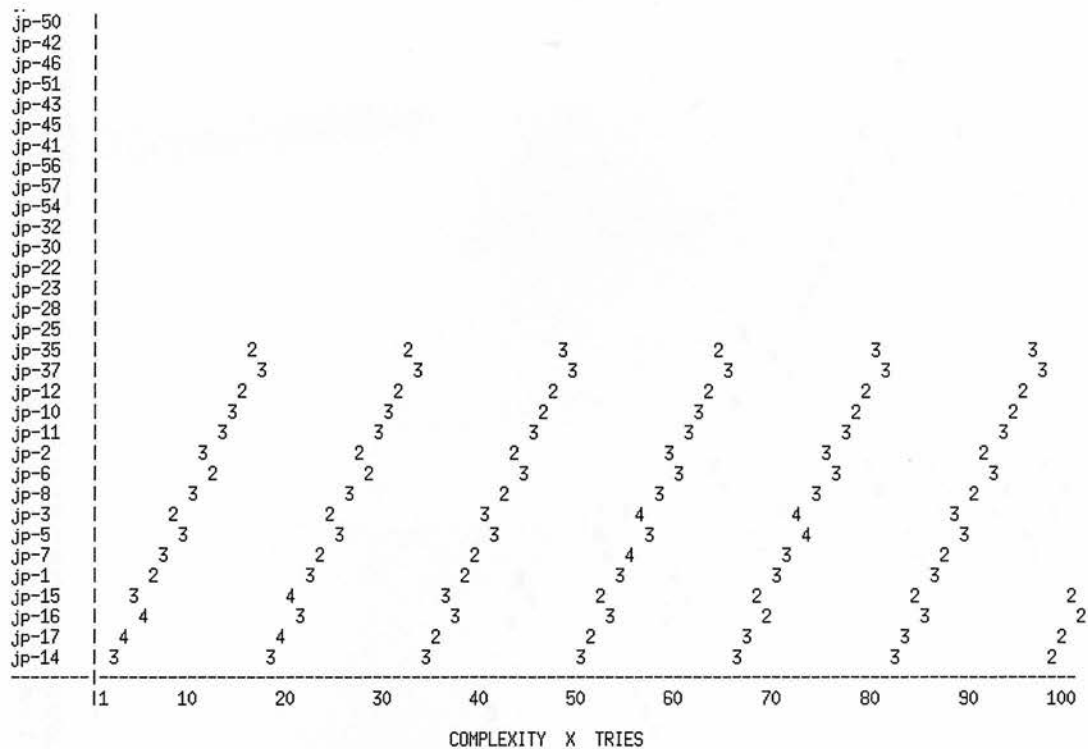


Figure 9.2: The path of the student with difficulties

access the most difficult ones without delay. Finally, the average student (Figure 9.4) progresses slowly but surely, and is able to try new, more complex activities.

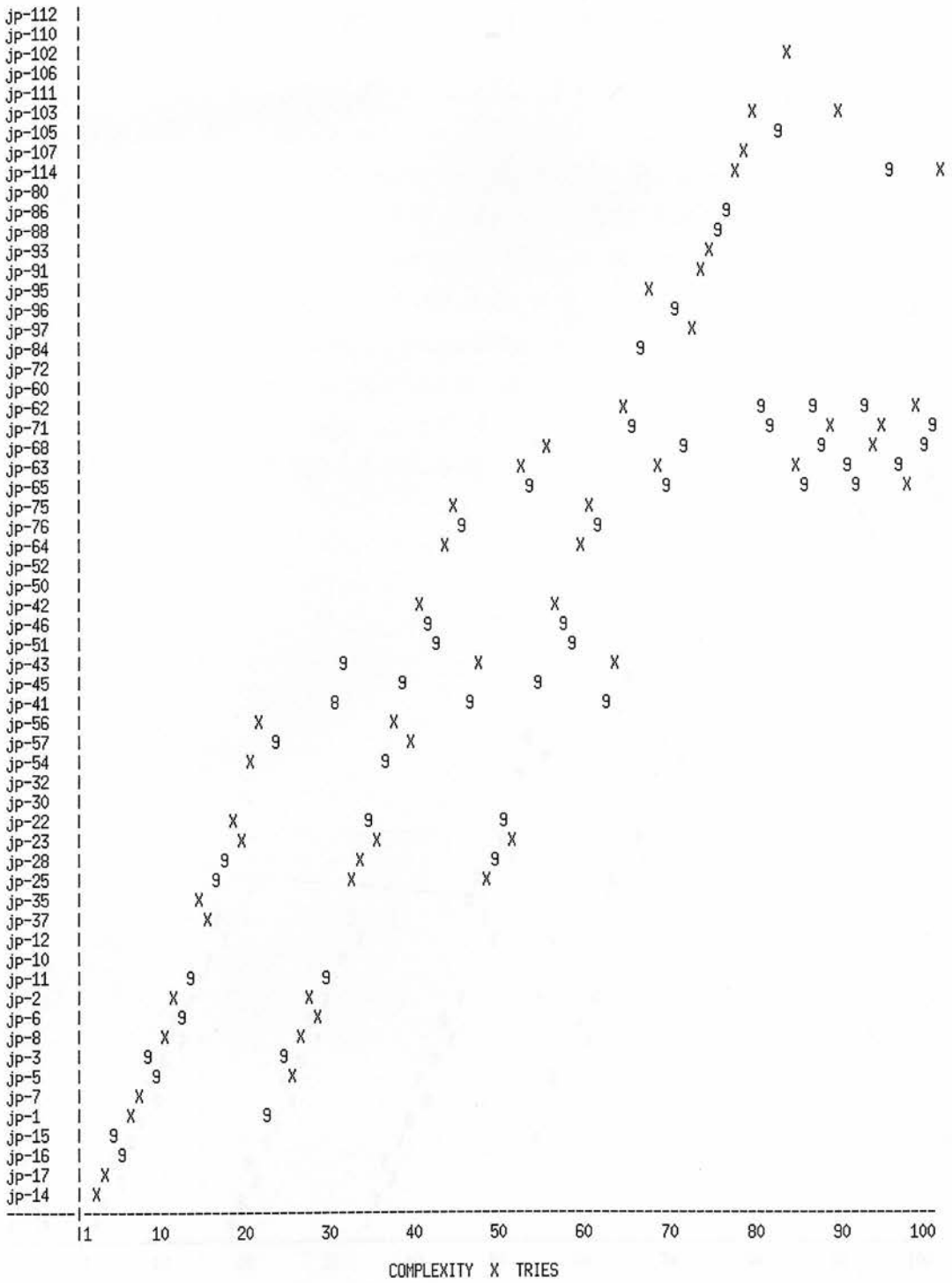


Figure 9.3: The path of the good student

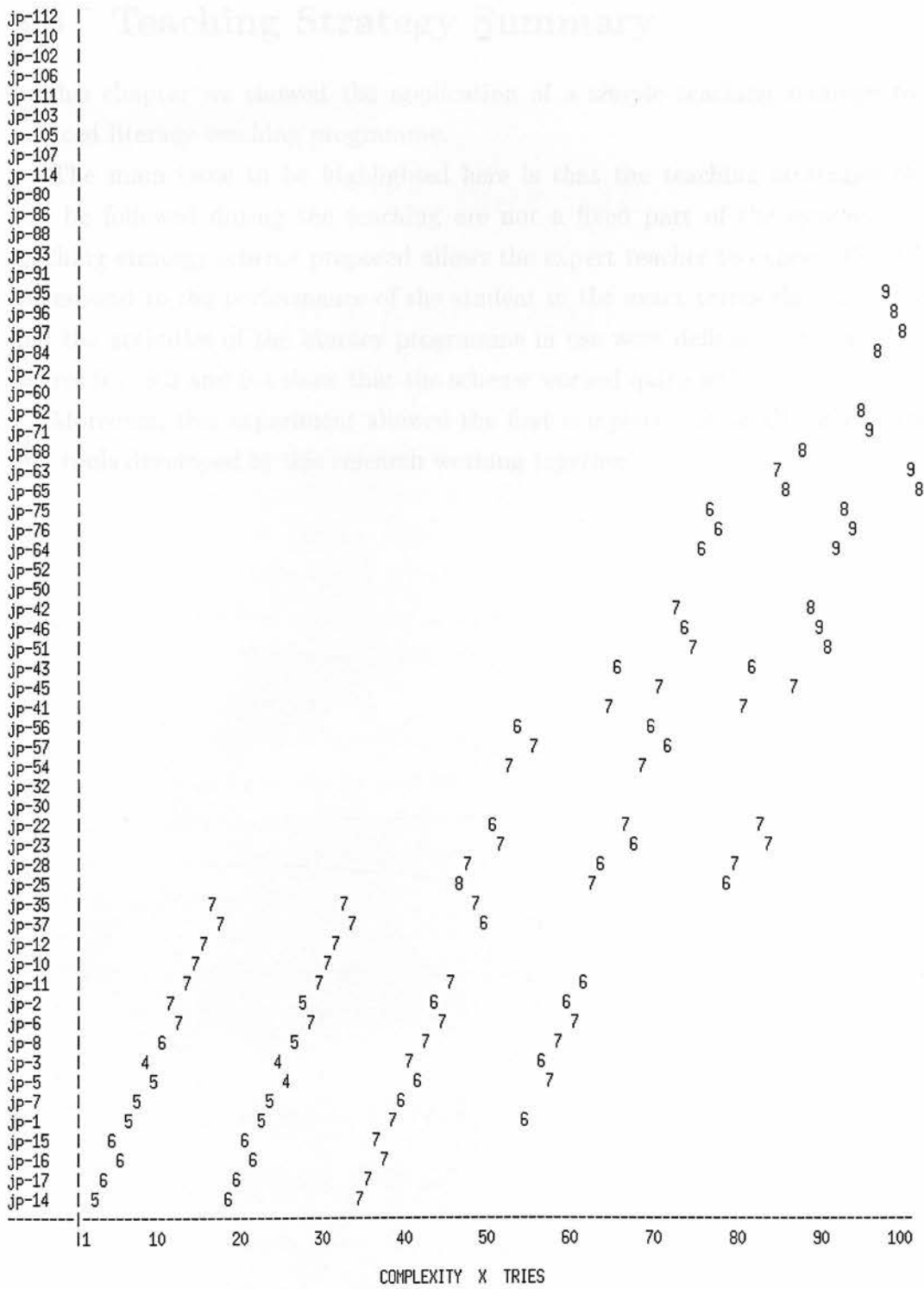


Figure 9.4: The path of the average student

9.6 Teaching Strategy Summary

In this chapter we showed the application of a simple teaching strategy to a reduced literacy teaching programme.

The main issue to be highlighted here is that the teaching strategies that will be followed during the teaching are not a fixed part of the system. The teaching strategy scheme proposed allows the expert teacher to express the rules to respond to the performance of the student in the exact terms the curriculum and the activities of the literacy programme in use were defined. The graphs in figures 9.2, 9.3 and 9.4 show that the scheme worked quite well.

Moreover, this experiment allowed the first complete test of all the concepts and tools developed in this research working together.

Chapter 10

Evaluation of Pedagogical Issues

This research is aimed at the teaching of basic literacy skills and has investigated different approaches to literacy suitable for adults. During the research the virtues of two different literacy approaches, Phonological Awareness and Generative Themes, have been made evident: Phonological Awareness approaches are supported by extensive experimental work and solid theory, whereas Generative Themes approaches are credited with a number of successful practical applications in teaching adults in developing countries around the world. However, apparently they have never been used in a combined way. Our objective became, then, to investigate whether these two approaches could be combined to produce a valuable literacy method to teach adults, deliverable by networked computer systems.

Using methods from Artificial Intelligence we designed tools for capturing the expertise of literacy teachers for:

- selecting vocabularies to be used in literacy teaching as a stimulus for learning, attending to requirements of both phonological awareness and thematic approaches;
- defining the curriculum to be followed in the teaching;
- defining the teaching activities;
- defining the teaching strategy to be applied.

In addition, a teaching environment for carrying out the teaching on the Internet has been developed and implemented in a prototype.

With the methods and tools developed during the research, we generated prototypical material that can be used to check the validity of the combination of the two approaches to form an effective literacy teaching method for adults. The work has been based essentially on theoretical works and on reports and manuals

for practical application of some PA literacy methods. This chapter describes an evaluation task aimed at the assessment of the propositions and outcomes of this research by two literacy experts, in what relates to the pedagogical issues and to the appropriateness of the material generated for use in the literacy context.

The two experts were Barbara Frame, from the Faculty of Education, University of Edinburgh, currently working as a trainer of literacy teachers, and Joyce Watson, from the School of Psychology, University of Saint Andrews, former leading literacy teacher, now doing research in synthetic literacy methods, based in phonological awareness approaches. They were invited to participate in the evaluation task because their range of expertise broadly covered the pedagogical aspects of this research. In the description of the evaluation activities Barbara Frame is referred to as the *Expert-T*, standing for “expert in teaching” whereas Joyce Watson is referred to as the *Expert-R*, standing for “expert researcher”.

The evaluation took place at University of Edinburgh, lasted for three and a half hours, and was somewhat informal in style. It was conducted by the author of this thesis and another researcher, referred to in this chapter as the *Facilitators*.

The evaluation task addressed both specific and high-level questions, and was designed to verify the following issues:

- Are the concepts and tools devised and developed in the research understandable by literacy experts and teachers?
- Are teachers able to make the low level specifications required to describe curricula and curriculum items?
- Do the tools developed in the research produce thematic and phonologic vocabularies similar to vocabularies generated by humans for the same purpose?
- Would the use of these tools for generating vocabularies be of interest to classroom teachers? Would they be of interest to expert teachers when researching and defining literacy methods?
- Are the outcomes of the research, in terms of the modelling of teaching activities and of teaching strategy, compatible with the practice in Phonological Awareness approaches?

The evaluation task was also aimed at obtaining recommendations from the experts on alternative ways of carrying out future work. The task was structured in four parts. The first part was a general presentation of the research. For each

of the subsequent parts a short presentation was made by the first facilitator. Then a questionnaire was distributed. The experts were asked to respond in writing to some specific tasks, and then, still in writing, to a number of broader questions. Finally, at each stage, a discussion on the themes in focus took place. The discussions were recorded on tape and transcribed, and extracts of the transcriptions were added to this thesis. One of the experts took away the copy the questionnaire to add further written comments. The activities are described in more detail below.

10.1 First Part - General Presentation

The main issues of the research were presented to the literacy experts. The presentation structure was:

1. Motivation and scope
2. Phonological Awareness review
3. Review of Freire's work
4. Main question: is it possible to combine these two approaches in a computerised system?
5. The vocabulary selection
6. The curriculum description
7. The design of activities
8. The teaching strategy

The experts were asked if they had any specific questions at this part. No further questions were asked.

10.2 Second Part - Vocabulary generation

The methods used to select words related to a particular theme were further detailed in a specific presentation. Then three questions related to vocabulary selection were proposed to the experts. They were asked to complete a form for the first two questions. The answers to the third one were tape recorded with their agreement. The questions and discussions of their results follow.

Question 2.1

The following samples of vocabularies have being randomly selected by the selection tool for one of these three themes: food, sea or agriculture. Please, determine to which of these themes each one belong (just one choice for each).

1. Vocabulary A - Theme: [sea¹]
abandon alluvium armada astern batter bay boatman border
bow breakwater brine canoe catch chuck commission contam-
inate counter curly cutter displacement distance dust exposed
falls fare fathom fin gourd harpoon horn incoming japan knot
lake lookout mississippi motor nautical oblique pipe plaice plot
pour preserve quartermaster rapid ruff ruffle salvage screw shel-
tered shore shoreline steam steer stem steward striker sway sweep
swirl terrestrial ton tortoiseshell turtle undercurrent voyage warm
whalebone whitefish yangtze
2. Vocabulary B - Theme: [food]
abominable allergic beetroot beside bony breakfast butcher caked
centrepiece chewy collins cream creamy cultivate cup digestible
dip double eating feed fibre fillet fruity gourmet grinder health
hungry indian jelly knead lemon meat mediterranean mixture
nutty oat owe peel perishable piquant poach potpourri puree
regular remove rocket scales scone scraps server shin smoky spe-
ciality sprout squash stew tab trash unwholesome vinegar wait
yolk zest
3. Vocabulary C - Theme: [sea]
abandon airlift anchor astern batter blast bleak blennioid board
brace broadside cabin cape coarse cod commando conch crash
curlew deck desert distance erosion flagship fleet flipper float
globe harbour haul island japan keel liner log marine master
navigable oblique petroleum pipe pirate plenty pull quartermas-
ter raw roar row safe schoolmaster scombroid screw seaside sed-
iment shallows shoreline spinner spot squall steam stretcher tai-
wan tidal tinker traffic uninhabited voyage waterway wave wreck
yacht
4. Vocabulary D - Theme: [food]
abominable array banana bean beside burger burn cinnamon
communion convert crispy crop culinary cup cyprinid dinner dish
drizzle exclusive fat fatty filter fishy grub gut honeyed hostess in-
stead jellied knead liver manna meat miscellaneous nutmeg om-
nivorous owe palate pan pepper pot prepare pulp raise replete
rocket salad sausage savour scone seed sesame soda spread sta-
ple supper taster turnip undressed vinegar wholewheat yolk zest

¹In this question the expected answers are given in the answer spaces.

5. Vocabulary E - Theme: [agriculture]

absorb airlift amoeba barrel base bath blade borderland breed
 broker bud cabbage canal chaff chrysanthemum clove commu-
 nity consumption cow cream cultivator develop dig dip dominion
 economy fallow feather female foam forage gauge geranium gra-
 nary hemlock hoe irrigate jersey knot lichen litchi lotus magnolia
 maple marginal mint november organic peach pear pine planta-
 tion producer pulp purchase rancher reclaim ripe rough seed shop
 sickle soap species spinach stick sting submerged sunflower train
 transport tuft turnip undergrowth vanilla walnut weak weather
 xylem yard zone

Discussion of Question 2.1

The question above was designed to check whether the thematic selection scheme achieved its aim of determining vocabularies that can be globally recognised as related to a particular theme. The experts had no problems in recognising the vocabularies (a) as for *sea*, (b) for *food*, (c) for *sea* and (d) for *food*. The vocabulary (e), for *agriculture*, was recognised by Expert-R, whereas Expert-T considered this vocabulary as belonging to the theme *food*. This is acceptable as although the sample had words strongly related to *agriculture*, such as *breed*, *cow*, *cultivator*, *dig*, *forage*, *hoe*, *irrigate*, *plantation* and many others, a good number of the words related to *agriculture* are also related to *food*, as is the case for *cream*, *maple*, *mint*, *organic*, *peach*, *pear*, *pulp*, *ripe*, *seed*, *spinach*, *sunflower*, *turnip*, and *vanilla*, among others. The final figure for this question is that in nine out of ten judgements carried out, the experts correctly identified the theme of the samples presented.

Question 2.2

To which thematic vocabulary do you think each one of the following words belongs? (Tick one or more themes).²

Word	Sea	Food	Agriculture
bank	[rs]	[]	[]
blow	[rst]	[]	[]
buy	[]	[rt]	[s]
cereal	[]	[rst]	[s]
coast	[rst]	[]	[]
cod	[rst]	[st]	[]
chrysanthemum	[]	[]	[rst]

²The answers are given in the fields in brackets in the following way: r - Expert-R answer; t - Expert-T answer; s - System's answer.

<i>delicacy</i>	[]	[rst]	[]
<i>dig</i>	[]	[]	[rst]
<i>division</i>	[]	[]	[rs]
<i>drown</i>	[rst]	[]	[]
<i>flavoured</i>	[]	[rst]	[]
<i>fertilise</i>	[]	[]	[rst]
<i>habit</i>	[]	[s]	[]
<i>haddock</i>	[st]	[rs]	[]
<i>inundate</i>	[]	[]	[s]
<i>jersey</i>	[]	[rs]	[s]
<i>lunch</i>	[]	[rst]	[]
<i>master</i>	[rst]	[]	[]
<i>nautical</i>	[rst]	[]	[]
<i>november</i>	[]	[]	[rs]
<i>pea</i>	[]	[rst]	[s]
<i>plaice</i>	[rs]	[st]	[]
<i>pollen</i>	[]	[]	[rst]
<i>sailing</i>	[rst]	[]	[]
<i>sailor</i>	[rst]	[]	[]
<i>skin</i>	[]	[rst]	[s]
<i>spoil</i>	[]	[rs]	[rt]
<i>transport</i>	[rst]	[t]	[rst]
<i>violet</i>	[]	[]	[rst]

Discussion of Question 2.2

This question was designed to check whether the experts would select words in a way that is similar to the way the system does. All the words presented in the list came from the three different thematic vocabularies as selected by the computer. By asking the experts to classify them it was possible to compare their judgement with the system's selection of words.

There were 30 words in the list with 38 matches to the themes, according to the system. The reason for this is that some of the words belonged to more than one theme (e.g. *haddock*, *plaice* and *cod* belonged to both *food* and *sea*).

For the theme *sea* the system pointed out 12 words, Expert-T found 10 words and Expert-R pointed out 11 words. All the words indicated by the experts coincided with the system's indications, and at least one expert agreed with each word the system indicated. The experts agreed with each other in 9 words.

For the theme *food* the system pointed out 12 words and each of the experts pointed out 10 words. Expert-T and the system agreed in 7 words, whereas Expert-R and the system coincided in 8 words. The experts agreed in 7 indications among each other.

The system found 14 words as related to the theme *agriculture*, while Expert-T pointed out 7 words and Expert-R indicated 9 words. They all agreed in 6 words. The experts agreed in 7 words, the system and Expert-T coincided in 6 words, whereas the system and Expert-R coincided in 8 words.

The figures in general indicate a high degree of coincidence in the categorisation of the words in relation to the themes. In particular, the numbers for the agreements between the system and an expert and among the experts are quite close.

Table 10.1 shows the relative coincidences among the experts and the system. The figures in this table correspond to the matches of each pair (Expert-T and the system, Expert-R and the system, Expert-R and Expert-T) and finally the coincidences among the three of them (Expert-T, Expert-R and the system). The calculation of each number takes into consideration the agreements either for the words considered related to the subject or not. For example, Expert-R and the system agreed that the word *bank* is related to the theme *sea*, the three of them agreed that *blow* is related to the same theme, and the three of them agree that *buy* is *not* related to *sea*. Thus, the agreement rate is calculated in relation to the thirty words.

Assessors	Sea	Food	Agriculture
Expert-T - system	28/30 (93%)	24/30 (80%)	21/30 (70%)
Expert-R - system	29/30 (96%)	25/30 (83%)	23/30 (76%)
Expert-T - Expert-R	27/30 (90%)	23/30 (76%)	22/30 (73%)
Experts - system	27/30 (90%)	22/30 (73%)	21/30 (70%)

Table 10.1: Comparative Matches Among the Experts and the System

According to Table 10.1 the system performed best for the theme *sea*, and worst for the theme *agriculture*. This result might be credited to a more obvious determination of the objects related to the theme *sea* in comparison to the other two themes. Nonetheless, the results report a substantive match between the experts' and system's evaluations.

Question 2.3

One of the goals of this research is to provide teachers with a means of generating thematic vocabularies. To what extent do you think these methods achieved this?

Discussion of Question 2.3

This question was designed to allow the experts to express their feelings freely about the selection results after they had been in practical contact with the system outcomes through the previous questions. The answers were tape recorded and transcribed. Extract of the transcriptions are given below.

Expert-R: I think I need to know how it is going to provide the teacher with the means. At the moment I know it is going to be delivered by the computer...

Facilitator: The question is, is it the case that teachers have this problem in obtaining thematic vocabularies and would it be interesting for them to have a vocabulary like these ones as a starting point?

Expert-R: I speak for myself, I think it would be very useful as a way of generating vocabulary...

Expert-T: When you are a busy teacher you can not even think about that. There is a computer program that does it for you, and I find it very useful. I can not remember the name, but what it does, is for instance, if you want the children to work with a rhyme family you can ask it...

Facilitator: How do you specify it?

Expert-T: There is a list of all rhyme families. So if a child has particular difficulties with a rhyme the teacher selects it and the child practises with it. It is also a fun thing because the words appear on the screen and are segmented and put back together again...

Facilitator: So do you think a tool for this objective is a good idea?

Expert-R: Yes, it is a good idea...

10.3 Third Part - Curriculum Description and Vocabulary Distribution

In another presentation, the concept of Knowledge Tree, used for the definition of the curriculum to be taught and for the distribution of the words across the curriculum, was given in more detail, along with some examples of the Knowledge Tree Definition Language (KTDL) code used in modelling the curricula of the literacy methods studied in the research.

Tables 6.2, 6.4 and 6.6 were given to the experts and studied as examples of the distribution of the vocabularies across the programmes' curricula. After the presentation, the following evaluation questions were given to the experts:

Question 3.1

Let $\{/b,/k,/d,/t\}$ $\{/u,/a,/i,/ei,/e\}$ $\{/d,/t\}$ be the phonetic specification for a particular curriculum item of programme X. Please indicate which of the following words would be in the group of words for that item.³

[rst] bad [] byte [rst] bed [] tilde
[rst] bit [rs] could [rst] dad [] that
[rst] date [rst] dead [rst] did [rst] kate

Discussion of Question 3.1

This question was designed to let the experts familiarise themselves with the proposed notation for specifying phonetic features of words. It also checks whether the phonetic coding, obtained from the Festival System, is easy to understand and compatible with the coding used in literacy teaching. The experts had no difficulty indicating the words that comply with the specification. The only disagreement occurred in relation to the word “could”, which Expert-T did not recognise as a three-phoneme word. For her, the correct coding would be $/c /u /l /d$. Expert-R and the system accepted the word as a three phoneme word, coded $/c /u /d$.

Question 3.2

Observe the following set of words: {baits, bathes, belt, bilge, boards, boats, bones, capes, coast, decks}.

Which of the following phonetic groups do you think the words belong to? (V stands for Vowel, C stands for Consonant and X stands for any phoneme).

[rst] CVCC

[] CCVC

[] $[/b/,/k,/h,/jh,/d,]$ $\{/e,/o,/i,/a\}$ $\{/g,/k,/s,/t\}$ {X}

³The representation used here for phonemes corresponds approximately to the same range of phonemes as IPA, but uses some alternative symbols. For clarity, some of them are illustrated here:

$/u$ as in “put”

$/i$ as in “sit”

$/e$ as in “bed”

$/oo$ as in “saw”

$/dh$ as in “this”

$/ii$ as in “see”

$/a$ as in “cat”

$/ei$ as in “day”

$/o$ as in “of”

$/ou$ as in “no”

$/jh$ as in “june”

[s] [/b/,/k/,/d,] {/e/,/ou/,/oo/,/i/,/ei} {/n/,/s/,/t/,/l/,/k/,/p/,/d/,/dh} {/s/,/t/,/jh/,/z/,/l}
 [] [/b/,/k/,/d,] {/e/,/oo/,/i,} {/n/,/s/,/t/,/l/,/k/,/p/,/d} {/s/,/t/,/jh/,/l}
 [rst] XXXX

Discussion of Question 3.2

This question tests the system notation and allows the evaluation of the distribution of words throughout the curriculum, as the experts classification can be compared to the system's classification. According to the system, the set of words would match the first, the fourth and the sixth expressions. Both the experts agreed with the first and the sixth expressions, but not with the fourth. This result might be due to the apparent complexity of the fourth expression.

Question 3.3

To which phonological categories do you think each one of the following words belongs? (Tick one or more categories)

Word	CVCC	{C}{V}{C}{/n/,/s/,/ou/,/ii}	CVCV	Other
bound	[rst]	[]	[]	[]
calls	[rst]	[rst]	[]	[]
fence	[rst]	[rst]	[]	[]
fix	[s]	[rst]	[]	[]
furrow	[]	[s]	[rst]	[]
garden	[s]	[s]	[]	[rs]
gets	[rst]	[rst]	[]	[]
given	[st]	[st]	[]	[r]
group	[]	[t]	[]	[rs]
heaven	[st]	[rst]	[]	[r]
heavy	[]	[st]	[rs]	[]
hello	[]	[rs]	[rst]	[]
holy	[]	[s]	[rst]	[]
Jersey	[t]	[s]	[rs]	[]
knots	[s]	[rst]	[]	[]
lakes	[st]	[rst]	[]	[]
leafy	[]	[s]	[rst]	[]
lex	[s]	[rst]	[]	[]
lights	[rst]	[rst]	[]	[]
lots	[rst]	[rst]	[]	[]
makes	[rst]	[rst]	[]	[]
mallow	[]	[rs]	[st]	[]
many	[]	[s]	[rst]	[]
maybe	[]	[rs]	[st]	[]
still	[]	[t]	[]	[rs]

<i>stone</i>	[]	[rt]	[]	[s]
<i>think</i>	[rst]	[]	[]	[]
<i>worry</i>	[]	[s]	[rst]	[]
<i>worst</i>	[rs]	[]	[]	[t]

Discussion of Question 3.3

This question, similarly to the previous ones, verifies: (1) whether the proposed scheme for defining the phonological features of the words is understandable by teachers; (2) whether it is comprehensive enough to express different needs; (3) whether it selects words effectively and in a similar way to expert teachers.

The analyses of the answers indicates a positive response for the three questions above. The system found 49 matches of the words of the sample with the three specified categories, and indicated that three words did not belong to any of the categories. Expert-T found 37 matches, 32 of which were in agreement with the system. Expert-R also found 37 matches, 33 of which matched the system's indication.

Table 10.2 establishes the same kind of relationship as Table 10.1 in the case of Question 2.2. Here there are 29 words and the rate of matching is calculated among the pairs of assessors (and the three assessors in the last line), counting the coincidences for each of the different categories indicated.

Assessors	CVCC	{C}{V}{C} {/n,/s,/ou,/ii}	CVCV	Other
Expert-T - system	23/29 (79%)	16/29 (55%)	25/29 (86%)	22/29 (76%)
Expert-R - system	21/29 (72%)	17/29 (58%)	27/29 (93%)	25/29 (86%)
Expert-T - Expert-R	20/29 (68%)	15/29 (51%)	25/29 (86%)	22/29 (76%)
Experts - system	20/29 (68%)	14/29 (48%)	25/29 (86%)	22/29 (76%)

Table 10.2: Comparative Matches Among the Evaluations

The rates of agreement in Table 10.2 although generally not very high, show similar levels between the various combinations in pairs, which indicates that the specifications of the phonological features of the words were understood by the teachers, were able to express different but very close phonological features, and selected the words in a way that is similar to that of the teachers. It is important to say that this task (and the previous ones) require precision and considerable short term memory effort. The worse rates in the case of the second expression ({C}{V}{C} {/n,/s,/ou,/ii}) indicate that more complex expressions might be more difficult for the teachers to process, which does not necessary signify they cannot be created by them after some practice. The experts appeared to do it

in a relaxed way, and did not seem to feel they had been put in the position of being tested themselves.

Question 3.4

The table shown below contains samples of words for some of the curriculum items of Phono-Graphix (VCC, CVCC and CCVC words). Some of these words were extracted from the set of original words for Phono-Graphix and some from the system selections for the themes food and agriculture (or both). Are there words in Table 10.3 that you would be unhappy to use as representatives of the curriculum items in consideration? Please circle them and explain why they are inappropriate (whether they are phonetically misleading, phonetically complex, contextually improper, non existent or whether they are inappropriate for any other reason).

Curriculum Item	Words
VCC	able, act, alp, amp, and, apple, armed, arms, ask, asp, axe, eats, eagle, east, eaten, eats, elf, elk, elm, else, end, even, iced, ink, inp, its, oats, often, oils, oink, old, oven
CVCC	baits, baked, bakes, bards, barks, basin, beans, beats, beaten, beefs, bent, bend, bind, bites, bitten, bison, boards, boast, bound, box, bowls, bulb, bulk, bulls, bump, buns, burns, camp, castle, cattle, cent, certain, child, chimp, coast, count, damp
CCVC	black, blip, block, bloom, blop, blouse, brain, bran, bread, break, breed, breeze, brick, broth, bruise, broth, bruise, brush, clam, clean, clone, cream, crop, cube, freeze, fresh, fries, shrub

Discussion of Question 3.4

This question aims to investigate the following questions: (1) Are there problematic words in the sample? Which ones? Where did they come from? (2) Will teachers be willing and able to amend word lists generated for their use? (3) How do generated words compare to original ones? For this comparison we traced the words that were originally in the Phono-Graphix scheme. They are: (a) VCC words: act*, alp, amp, and*, ant*, apple*, asp, eagle, east*, elf, elk, elm, end, inp, ink, oink; (b) CVCC words: bend, boast, box*, bump, camp, castle, cattle, cent, certain, child, chimp, coast, count, damp; (c) CCVC words: black, blip, block*, blop, blouse, brain, bread, break, breeze, brick, clam, clean. The words marked with an asterisk were also present in the sets generated by the system.

The experts considered that some of the words in the sample could represent a problem for inclusion in literacy programmes, at least for the initial stages, which are the stages relevant to this research.

For Expert-T these were the problematic words:

- Words with the sounds /*ar* and /*ur*, such as *arms*, *armed*, *bards*, *barks*, *boards* and *burns*, which in Scotland would be pronounced with the “r”, differently from in England.
- The word *box*, considered as a CVC word, because of the acceptance of the sound “ks” as a unique phoneme.
- The word *certain*, considered as a CVCVC word, with the vowel /*ai* sounding before the /*n*; likewise, other words ending with the phoneme /*n*, such as *eaten*, *oven*, and *bison* are also problematic because the assumption that the *Schwa* vowel before the /*n* do not sound is not accepted by some teacher in other parts of Britain.
- Words such as *black*, *blip* and *block* were considered doubtful for in some places the sound “bl” would be considered as a unique phoneme - in this case, these words would be considered as CVC words; the same occurs with other blending such as *br*, *cl*, *cr*, *fr* and so on.
- Words such as *inp* and *oink* were placed in the category of non-words, though were accepted as necessary for developing and testing the encoding and decoding skills.

Expert-R pointed out basically the same problems, adding the word *often*, which has two accepted pronunciations; she felt that it should be avoided in early stages of teaching.

In all cases the problems pointed out arise in words extracted from both the system’s and the Phono-Graphix’s vocabularies, and were related either to differences of pronunciation in different regions, or to different interpretation of the way phonemes are coded in the English language. As mentioned in Chapter 2, it has been generally accepted that the English language has *around* 44 phonemes, and this does not include *blended phonemes*, which would cause a significant rise in the number of phonemes to be dealt with. This is a real problem in literacy teaching regardless of the methods or the media being used to help teaching. The possible ways to deal with these problems are discussed more comprehensively in Chapter 11. However, it was made clear that there are some possible solutions to

avoiding such words whenever they are considered as a problem. No words were considered particularly inappropriate for any reason other than their phonological complexity.

Question 3.5

Do you think these methods could be useful for classroom teachers while looking for stimuli to be used in the teaching of literacy?

Discussion of Question 3.5

The answers to Questions 3.5 were recorded on tape and transcribed abstracts are presented below. This and the following questions are intended to allow the expert to express their impressions while using the scheme for defining phonetic features of words, and to collect their suggestions for improvement.

Expert-T: I think the first advantage is that the vocabulary the children will be using would be in a context, so they are using vocabulary for reading, spelling and writing which leads to the theme, so that helps. The next step is to consider what you would imagine that the teachers would do with the words in the categories according to the knowledge tree.

Expert-R: I've been finding great difficulty in making up sentences and little stories and trying to put everything into a context and I have to sit down with a dictionary and look up words which are compliant with the ones the children have already learned, so could see this is being very helpful indeed, because I could just have the theme ready and look it up and fit them into the different categories that I would be needing as a teacher. That's a lot of hard work. ... to put everything into little sentences ... and the computer could do it for me. It has to be within a certain group of letters, with certain constraints, and it could generate the group of words to fit the criteria. And now using all programmes, I presume you could use Jolly Phonics, or Phono-Graphix or the school's own programme, whatever it is, and just fit it in. It will take a little understanding, I think, to know what to do, but, given that, I would not see a problem.

Facilitator: Do you think it would be something that only specialised teachers like yourself could do, or do you think that it would be something that classroom teachers could do?

Expert-R: There are some very good classroom teachers. It would not be for all of them, to be fair, but it is a sort of thing that could possibly go into in colleges of education with students of research, and the new ones would come out able to do it and maybe help the old ones...

Expert-T: There is certainly research to indicate that effective teachers of literacy do not necessarily have explicit subject knowledge, and one of the things they find difficult to talk about is phonemes and graphemes, so something like this could be almost a staff development exercise and actually using this the teacher would build up a knowledge of how to analyse words by graphemes or by phonemes... and getting children involved in that kind of discussion helps as well, because it mixes awareness of the language at word level, and at phoneme and grapheme. So I do not think it would do any dis-service. I think people would just agree... because, I think there is disagreement about where a phoneme begins and where a phoneme ends, and people have different points of view according to their backgrounds and pronunciation.

Expert-R: You say that not all actual teachers would be able to do it, but, for instance, in England the students of universities and secondary schools Reading Diploma were actually used as literacy consultants and even around schools, if they still have literacy consultants, these people could probably be taken as the staff to have it, and they could go into the schools and do it in service, or the colleges of education could do it in service. I think you would need a staff to try it, it could not go straight to the classroom...

Question 3.6

How do you think these methods could be improved?

Discussion of Questions 3.6

The answers to Question 3.6 were recorded on tape. Abstracts of the transcriptions are presented below.

Facilitator: (after reading the question aloud) For instance, there is this question of the symbols used to represent the phonemes...

Expert-R: Well this is not a problem, it is just a case of different representation, the slash is just a means to say that is not the name of the letter, because not everybody does this way...

Facilitator: Do you tend to use IPA generally or do you have variations on that that you use?

Expert-T: I use variations of it.

Expert-R: In a school if a child has difficulty with language, this is how it would be coded, but we do not expect them to do this. This is a particular code system for specialists.

Facilitator: and if you are trying to do things like describing rhymes families what coding do you use in sounds?

Expert-T: I use simple long and short vowels...

10.4 Fourth Part - General Evaluation

The principal characteristics of the Authoring Tool for describing the teaching activities and the principles for defining the Teaching Strategy were described in another short presentation. The prototype of the system was demonstrated. Then another set of questions were addressed to the experts. The answers were recorded on tape and the transcriptions follow. Expert-R asked for a copy of the questions and later sent some written comments to each question, which are also presented after the tape transcription.

Question 4.1

Do you think the outcomes of this research can be useful for teaching illiterate adults to read?

Discussion of Question 4.1

The answers to the question 4.1 were recorded on tape. Abstracts of the transcription are presented below.

Expert-T: I think the role of a computer is well known as a motivator, and if you're talking about adults, then it enables you to please yourself... and learning in your own way would be quite helpful for adults. We know that phonological awareness is a key issue in literacy. I know that this is a very first draft and that it would have to be much more exciting in terms of the presentation, graphics and so on. The difficulties with the speech synthesiser would need to be sorted out.

Expert-R: I would say the same, that it would be very useful, and that people would be very motivated to do it, I worry a little bit about the font in one or two of them, because this is not the way that children use to do it, maybe it would not matter for adults at all. And I know that there are problems with graphics, for the copyright, but you can find some way with students, some are real artists...

Facilitator: Oh yes, but the main idea in the thesis is to test the ideas, particularly the idea of "can you combine thematic and PA approaches in ways that provide useful tools"...

Expert-R: Oh yes, I would think so, yes, very much so.

Expert-R written comments:

Advantages: working at own pace, using familiar vocabulary, and using semantic categories.

Consider: the quality and the type of graphics to use for adults.

Question 4.2

Do you think they can be adapted to teach children?

Discussion of Question 4.2

The answers to the question 4.2 were recorded on tape. Abstracts of the transcription are presented below.

Expert-T: I have seen children in the nursery sitting and using the computer to learn the letters and sounds and they seem to be very interested, and we are talking about three and four years old.

Expert-R written comments:

The graphics seem to be suitable for teaching children. The flexibility of the programme would be very useful for the teacher.

Question 4.3

Would it be of value if used by students without the direct supervision of a teacher?

Discussion of Question 4.3

The answers to the question 4.3 were recorded on tape. Abstracts of the transcription are presented below.

Expert-T: It will depend on how good the program was, because in the topic of taking the teacher away somehow you have to make the computer stand in for the teacher, and it will be important to know where each of the adult learners would be in the curriculum, to start with what they know. I wonder if the program can do that.

Expert-R: I don't see that as a big problem because I think that if you had someone in charge of it, who does some sort of pre-test as to know who knows nothing, who knows a bit better, who knows something more, and initially they are all given, and part of the program would seek you and it should be able to work out your level, there should always be someone at hand, but does not really take part until somebody has a problem, and then they will go and maybe help them, and at the end of the lesson, each two hours things would be printed out so she would be able to access how you got on and in the next there would be necessary a pre-test to give you the lesson for the next two hours... I think it won't be worth it to put somebody beside you, but I think you could need somebody on hand who you could call if you needed... and if you do make a mistake the machine is there to tell you, you don't need someone to tell you, you are an adult and you have got discernment.

Expert-T: It also depends on the context. If you are talking about a remote mountain village and there is a computer there but there is no teacher, nobody there knows about it, then you have got the problem of getting to grips with switching the computer on and I remember how difficult that was when they arrived at the schools, so you have to go over that, and then, the software itself, which has to be installed. Obviously the best solution is to have teachers there who would introduce the basics, providing some kind of goal and purpose for the people... The way it is presented is in a very linear sequential method of learning, isn't it, and we know that there are people who like the big picture first before you go through all those small steps... Maybe it would be interesting to have an introduction as part of the software as well...

Expert-R written comments: Adult students would be able to work without supervision during sessions. Each session would need: introduction (with the teacher), independent work (without the teacher), and conclusion (with the teacher). The teacher would have to be present to assist with individual problems.

Question 4.4

Do you have any opinion about transferring the scheme presented to a different language? (e.g. Portuguese)

Discussion of Question 4.4

The answers to the question 4.4 were recorded on tape. Abstracts of the transcription are presented below.

Expert-T: It seems to be agreed that English is the most difficult language in this regard. In Portuguese the one-to-one correspondence is much easier.

Expert-R: Well, I think that to South Africa it might be a bit difficult. There should be about ten different dialects. This is no different from England and Scotland. We are finding all these problems, we are just after finding the problems and sorting them out, and I think somebody needs to help you with the accents and dialects...

Expert-R written comments: English is difficult. Might even be easier in Portuguese. The activities in this evaluation served to illustrate the accent problem in English, all of which have to be taken into account when delivering a programme.

Question 4.5

Please, feel free to make any other comments on this research.

Discussion of Question 4.5

The answers to the question 4.5 were recorded on tape. Abstracts of the transcription are presented below.

Expert-T: In terms of general comments, it is interesting research, because it is bringing together a number of things that are very much in vogue, the Phonological Awareness, the fact that it is presenting the work in different ways, the visual, the auditory, the written text as well, to help illiterate adults to learn to read.

Expert-R written comments: A tremendous amount of work has gone into this research. It has possibly been made more complex by trying to consider both illiterate adults and beginning child readers. The life experience of adults would have a bearing upon the selection of thematic vocabularies. Some of the words in question 4, third part, would be outwith the experience of young children. The teacher, however, could select for appropriateness and be able to generate other words.

10.5 Comments on the results of the evaluation

The evaluation task, as here described, confirmed with practical and objective questions many of the claims that had been made during the development of the research. The expert teachers who assessed it were able to understand and work with the concepts and tools developed in the research, and explicitly stated the necessity of these kind of tools for teachers while preparing literacy programmes or even regular classes.

It must be noted that the experts had only very limited contact with the research before the evaluation task, and their formal introduction to the concepts and notations used in the evaluation lasted for less than one hour of presentation. The evaluation task itself took three hours, where fourteen questions, six of them with heavy cognitive load, were answered. Due to the time constrictions the cognitive questions were perhaps given less time than the ideal to be elaborated. However, the results obtained largely confirm the appropriateness of the concepts and tools assessed.

Chapter 11

Discussion and Conclusion

This research was born out of a proposal for the EBAC Project - Basic Education Assisted by Computers - as submitted to CAPES/Brazil and The British Council in 1996 by a group of researchers from the University of Brasilia. The two organisations had invited submission of project proposals on *basic education*. The theme *literacy* had just emerged as a possibility at that time. However, after arriving in Scotland, the sad news about the death of Paulo Freire at 76 years of age, led to a decision to carry out research related to his legacy. Paulo Freire's method for teaching literacy, with the Generative Themes concept then emerged as one of the possible basis of the research.

However, the study of recent work in teaching reading in English inclined the research towards phonological awareness approaches. Many questions arose at that time, related to the possibility of exploring these different points of view and applying them in a complementary basis in some way. Eventually the main question to be answered in this research emerged: *Is it possible to find enough stimuli to cover all the aspects of a phonological awareness programme in a thematic vocabulary suitable for teaching reading to adults?*

This chapter summarises the outcomes of the research, discusses the main questions and their possible answers and comments on the open areas for future work on this research.

11.1 Outcomes of the research

A whole environment had to be set up for conveying and supporting the attempt to answer the main and other related questions. This led to the study and proposition of an intelligent tutoring system generator to test and make use of the results of the main research. Almost forty thousand lines of Java code has been written since the research began. Many concepts, models and tools have been

developed and implemented as part of this thesis. The main elements developed are summarised below:

- The *LTS Dictionary* is a dictionary specific to the research with the features needed for phonological processing, selection of words according to complexity and derivation of related words (inflections and synonymy).
- The *Seed Set* concept was created in relation to the thematic selection of words from the LTS dictionary based on the occurrence of specific words named *seeds* in the description of the words in the dictionary entries. This concept is implemented through a tool which allows the selection with various selection parameters.
- The *Knowledge Tree* concept comprehends a hierarchical structure designed to organise the curriculum to be taught and to permit the description of the phonological features required for the vocabulary of each curriculum item. To implement the concept the Knowledge Tree Definition Language was developed. The concept was further used as the basis for the definition of the overlay models of the expert and of the student.
- The *Activities* concept formalises and classifies the teaching activities as *lessons*, *examples*, *exercises* and *tests* in order to support the definition of the assessment scheme and of the teaching strategy to be used while teaching.
- The concepts of *Objects* and *Steps* are used to describe the teaching activities step by step.
- The *Authoring Tool* was designed to allow the definition of the activities and their integration into the knowledge tree of the literacy method in use.
- The *Teaching Strategy* model allows the expert teacher to define the teaching strategy, according to the knowledge tree and the performance of the student while being submitted to the literacy programme in use. A language has been defined to allow the definition of the strategy (Teaching Strategy Definition Language) and an interpreter of the language has been written to implement it.
- The *Short Phonix Literacy Programme* is a reduced literacy programme based on the three literacy programmes most closely studied in this project (Phono-Graphix, Jolly Phonics and THRASS), developed to demonstrate

the features of the proposed model and system, including a teaching strategy model and a thematic vocabulary for the theme *food* with a database of pictures to be used in conjunction with the words.

- The *Integrated Prototype* - encompasses all the concepts, models and tools above mentioned, which were integrated in a prototype, implemented as a Java Applet which can be run from any point of the Internet through a common browser. The student model is updated in the original site at the end of each activity.

11.2 Questions and Answers

As said before, many questions arose and were answered during this research. The main question on *the possibility to find enough stimuli to cover all the aspects of a phonological awareness programme in a thematic vocabulary suitable for teaching reading to adults* has a positive answer in Chapter 6 (Tables 6.1, 6.3 and 6.5) after the discussion of the material presented in Chapters 4 and 5. The four thematic vocabularies obtained automatically by the tools developed in this research proved to contain enough words to cover the phonological requirements of three different PA methods, in a comparable way to the vocabularies used originally in those methods.

The next important question occurring throughout the research was: *Is it possible to combine these two different approaches to teaching reading (PA and Generative Themes) to obtain a method that would be suitable for teaching adults using computer systems and the Internet?* This is a question that remains an open one, for no objective experiment on teaching real students has been carried out so far. However, the prototype developed during the research shows that the model and the tools proposed and implemented are consistent and allow literacy teaching activities related to phonological awareness and generative themes to be carried out through the Internet, while updating the student model and adapting to the student's performance in learning.

The affirmation that *literacy teaching activities* related to the approaches in focus in this research are delivered by the prototype is supported by the evaluation task described in Chapter 10, where many of the outcomes of the research were evaluated by literacy experts. In connection with the evaluation task some other important questions related to the pedagogical issues of the research were asked and answered. The most important of such questions are:

1. *Is the concept of Seed Sets understandable and possibly useful for teachers*

either to be used in conjunction with the system proposed in this thesis or as a stand-alone tool to help in dealing with thematic vocabularies?

2. *Is the concept of Knowledge Tree, including the constructions related to the phonological features of the words, appropriate for describing PA curricula?*
3. *Would expert teachers be able to understand and work with the concepts and tools proposed in this thesis? Would classroom teachers be able to benefit from these concepts and tools?*
4. *Would illiterate adults possibly benefit from a system such as the one proposed in this thesis?*
5. *Would a system like this be also interesting for helping in teaching children?*
6. *To what extent could a system like this be autonomous in the teaching of literacy?*

The answers to these questions were positive or encouraging in the case of the last one, as described in Chapter 10. Some foreseeable derivative questions arose on various aspects such as the human-machine interfaces and interaction devices, the quality of speech, the phonological coding and the operational support. Most of the answers are addressed in the indications for future work, in Section 11.4.

11.3 Sensitive Points

Some of the options and development in some aspects of the research, although constituting valuable contributions, should be redesigned before further work is carried out. This section lists and comments on these issues.

- The graphical interface with the student: a general layout of the screens must be defined, with standard sizes for pictures, written material, buttons and other elements. The use of proper fonts for teaching reading and the definition of a clear way for representing letters, graphemes and phonemes are strongly required.
- The interface with the expert teacher: the development of more user-friendly interfaces is essential for the expert teacher to deal with the vocabulary, the definition of the knowledge tree and the description of activities. This will allow the involvement of expert teachers in testing the concepts and tools of the system, and in experimenting with other literacy methods.

- **Vocabulary:** the improvement of the dictionary used as sources of words for the system, including data on the frequency of use and age of acquisition of words is fundamental for supporting further work. It has been stressed in this thesis that the number of words available for the curriculum items in PA methods could be significantly improved with the use of more complete and concise source of words.
- **Speech Generation:** the improvement of the speech generation facility is fundamental for this research. The use of speech synthesis liberates the designers and producers from the time-consuming task of recording dialogues. The drawbacks so far have already been pointed out: generated speech is not yet natural enough to be acceptable in a task as sensitive to listening as literacy teaching is. Besides, the generation of isolated phonemes is not audible and clear enough to be appropriate for literacy teaching. Moreover, for the fulfilment of the aims of this research a device is needed that is able to produce speech in Portuguese as it is spoken in Brazil. Although it is to be expect that in a few years time this problem will be solved, it seems that for the time being the only solution for practical application will be the prototypical generation of speech while developing the ITSs, and its recording for the final versions of the systems. Some research on speech libraries would be useful at this moment.
- **Phonological Coding:** The ability to deal with different phonetic coding is another important aspect to be improved. During the Evaluation Task described in Chapter 10, some words were considered problematic for being used in the first stages of the teaching process due to the different phonetic coding in different regions or in different approaches to teaching reading. The first solution, already available, is the exclusion of these words either manually, by pointing out each one of them, or automatically, by specifying the phonetic or letter-group pattern of the undesirable words. The second and possibly a more definitive solution, is to create schemes to adjust the phonetic representation of the words to the particular phonology of the region or to the specificities of the method in use. The solution must keep compatibility with the speech device in use.

11.4 Future Work

As expected, this research has confirmed the theme as an open area for future developments. Some of these are related to questions not completely answered yet,

while others simply represent improvements and openness to new development. The relation of the desirable future work on the material presented in this thesis follows.

- The development of animation and cartoon generation support, suitable for literacy teaching.
- The development of the issues related to the personal data from the student, which would produce information to help on the decisions about strategy and selection of teaching stimuli.
- The development of thesaurus-like relationships between the words of the system and the functions to explore these relationships while selecting stimuli for teaching (broader term, narrower term, equivalence, and so on). This is an interesting research area, in two main aspects: how to get the relationships and how to store and use them.
- The development of support to provide pictures to be used by the system. In the same way that some words have different meanings described with different words, they might have different pictorial representation in diverse contexts. Consequently, it is not trivial to have a pictures database, as the context must be taken into consideration.



Figure 11.1: The two “tops”

Figure 11.1 illustrates two different pictorial representations for the word *top*, one for the broad *agriculture* context (the top of a mountain) and the other for the *food* context (the representation of an ice-cream with an arrow pointing to the topping on top of it). A research on automatic acquisition and generation of pictures associated with meanings is an interesting development of this work.

- The incorporation of speech recognition facilities - This represents another fundamental development for the success of this project, as communication with illiterate people is primarily verbal. Although research on the consequences of literacy in speech recognition and production is not conclusive in regard to pronunciation [McGuinness, 1997, p.137], [Morais and Kolinsky, 1995], it seems reasonable to think that the articulation of words by illiterate people from different regions of Brazil might vary significantly more than the articulation of literate computer users. This could imply extra training for the speech recognition algorithms to learn the different accents and speech versions of words.
- The development of handwriting recognition facilities - New devices are coming on the market with writing boards and software for learning and recognising individual calligraphy. As happened with other kinds of devices such as scanners and digital cameras, there is a tendency for popularisation of such devices in the near future. However, for the purposes in focus here, it must be considered that illiterate people do not have calligraphy skills developed at all and work from scratch must be carried out in their cases. This is another point for serious investigation.
- The incorporation of musical support - The filling of gaps with thematic songs is another requirement of tutoring systems which has not been tackled so far in this work, and will probably represent a great deal of interest.
- Literacy teaching in Portuguese - Concerning pedagogical issues, there is another open area. Traditionally Portuguese literacy teaching has been carried out based on syllabic stressing, for the language is regular enough to be taught this way. The new question is *what impact would a phonemic approach have on the literacy learning process in Portuguese?*

11.5 Final Evaluation

As a final evaluation, this work has contributed to the teaching of reading by formalising and providing a framework for generating vocabulary to support phonological awareness and generative themes approaches. It has also opened a valuable horizon of new possibilities and opportunities to be explored in future work.

Bibliography

- Adams, M. J. (1990). *Beginning to Read - Thinking and Learning about Print*. Massachusetts Institute of Technology, Palatino - USA, 1st edition.
- Anderson, J. R. (1988). The expert module. In Polson, M. C. and Richardson, J. J., editors, *Foundations of Intelligent Tutoring Systems*, chapter 2, pages 21–53. Lawrence Erlbaum Associates Publishers, London.
- Balmuth, M. (1982). *The Roots of Phonics - A Historical Introduction*. McGraw-Hill Book Company, New York.
- Bell, L. C. and Perfetti, C. A. (1994). Reading skill: Some adult comparisons. *Journal of Educational Psychology*, 80(2):244–255.
- Black, A. W., Taylor, P., and Caley, R. (1998). The Festival Speech Synthesis System. System documentation, Centre for Speech Technology Research - University of Edinburgh. <http://www.cstr.ed.ac.uk/projects/festival/manual-1.3.0/>.
- Bloom, B. S., editor (1956). *Taxonomy of Educational Objectives*, volume I. David McKay Company, Inc, USA, 1st edition.
- Bradley, L. and Bryant, P. (1991). Phonological skills before and after learning to read. In Brady, S. A. and Shankweiler, D. P., editors, *Phonological Process in Literacy - A Tribute to Isabelle Y. Liberman*, chapter 3, pages 37–45. Lawrence Erlbaum Associates, Publishers, Hillsdale, New Jersey.
- Bradley, L. and Bryant, P. E. (1983). Categorizing sounds and learning to read - a causal connection. *Nature*, 301:419–421.
- Bruneau, J., Chambreuil, A., Chambreuil, M., Chanier, T., Dulin, M., Lotin, P., and Nehemie, P. (1991). Cognitive science, artificial intelligence, new technologies: How to cooperate for a computer-assisted learning to read system. In Birnbaum, L., editor, *The International Conference on the Learning Sciences. Proceedings of the 1991 Conference*, pages 74–82, Illinois, USA.

- Burns, H. L. and Capps, C. G. (1988). Foundations of intelligent tutoring systems: An introduction. In Polson, M. C. and Richardson, J. J., editors, *Foundations of Intelligent Tutoring Systems*, chapter 1, pages 1 – 19. Lawrence Erlbaum Associates Publishers, Hillsdale, New Jersey.
- Carvalho, M., Pain, H., and Cox, R. (1999). Automatic vocabulary generation for a literacy teaching ITS. In Cumming, G., Okamoto, T., and Gomez, L., editors, *ICCE99 - Proceedings of the 7th International Congress on Computers in Education*, volume 55 of *Frontiers in Artificial Intelligence and Applications*, pages 784–791, Amsterdam. IOS Press.
- Chandler, D., editor (1983). *Exploring English with Microcomputers*. Council for Educational Technology, Leicester, Great Britain.
- Colton, D., Cole, R., Novick, D. G., and Sutton, S. (1996). A laboratory course for designing and testing spoken dialogue systems. In *International Conference on Acustics, Speech and Signal Processing*, pages 1129–1132, Atlanta.
- Coupland, J. (1983). Software - and historical overview. In Chandler, D., editor, *Exploring English with Microcomputers*, pages 99–107. Council for Educational Technology, Leicester, Great Britain.
- Davies, A. and Ritchie, D. (1996). *THRASS - Teaching Handwriting Reading and Spelling Skills - Teacher's Manual*. Collins Educational, London.
- Diack, H. (1960). *Reading and the Psychology of Perception*. Peter Skinner Publishing Limited, Nottingham.
- Diringer, D. (1947). *The Alphabet - A Key to the History of Mankind*. Hutchinson's Scientific and Technical Publications, London.
- Downing, J. (1979). *Reading and Reasoning*. W and R Chambers, Great Britain.
- Ejerhed, E. and Church, K. (1995). Written language corpora. In Cole, R. A., editor, *Survey of the State of the Art in Human Language Technology*, chapter 12.2, pages 445–454. Center for Spoken Language Understanding, Oregon.
- Eurich, N. P. (1990). *The Learning Industry - Education for Adult Workers*. The Carnegie Foundation For the Advancement of Teaching, Princeton, New Jersey.
- Flesch, R. (1955). *Why Johnny Can't Read - and what you can do about it*. Harper and Row, Publishers, New York.

- Freire, A. M. A. (1997). *A voz da esposa: A trajetoria de Paulo Freire. Instituto Paulo Freire: Biografias*, 1. <http://www.ipf.org.br>.
- Freire, P. (1970). *Pedagogy of the Opressed*. Penguin Books, England.
- Freire, P. and Macedo, D. (1987). *Literacy: reading the Word and the World*. Routledge and Kegan Paul, London, 1st edition.
- Fries, C. C. (1962). *Linguistics and Reading*. Holt, Rinehart and Winston, Inc., New York.
- Gelb, I. J. (1952). *A Study of Writing*. The University of Chicago, Chicago, London, 1st edition.
- Goswami, U. and Bryant, P. (1990). *Phonological Skills and Learning to Read*. Psychology Press, East Sussex, Great Britain.
- Hatcher, P. J., Hulme, C., and Ellis, A. W. (1994). Ameliorating early reading failure by integrating the teaching of reading and phonological skills: The phonological linkage hypothesis. *Child Development*, 65:41–57.
- Hawkrige, D. (1990). Software for schools: British reviews in the late 1980's. In Boyd-Barret, O. and Scanlon, E., editors, *Computers and Learning*, chapter 1.7, pages 88–108. Addison-Wesley Publishing Company, Wokingham, England.
- Howe, J. A. M. (1981). Teaching handicapped children to read: a computer based approach. In Lewis, R. and Tagg, D., editors, *Computers In Education*, pages 433–440. North-Holland Publishing Company, Amsterdam.
- Hutchins, J. (1998). *Literacy Software*. British Dyslexia Association, England, 2nd edition.
- Ide, N. M. and Sperberg-McQueen, C. M. (1995). The TEI: History, goals and future. *Computer and the Humanities*, 29:5–15.
- Johnston, R. S. and Watson, J. (1997). Developing reading, spelling and phonemic awareness skills in primary school children. *Reading*, 31:37–40.
- Kohler, W. (1969). *The Task of Gestalt Psychology*. Princeton University Press, Princeton, New Jersey.
- Lees, P. and Chapman, B. L. M. (1986). New technology and new techniques in the teaching of reading to the hearing-impaired. *Journal of British Association of the Teachers of the Deaf*, 1(10):10–16.

- Lewkowicz, N. K. (1980). Phonemic awareness training: What to teach and how to teach it. *Journal of Education Psychology*, 72(5):686–700.
- Liberman, I. and Shankweiler, D. P. (1991). Phonology and beginning reading: A tutorial. In Rieben, L. and Perfetti, C. A., editors, *Learning to Read: Basic Research and Its Implications*, chapter 1, pages 3–17. Lawrence Erlbaum Associates, Publishers.
- Lilley, I. M. (1967). *Friedrich Froebel, a selection of his writings*. University Printing House, Cambridge.
- Limage, L. (1990). Adult literacy and basic education in Europe and North America - from recognition to provision. *Comparative Education*, 26(1):75–88.
- Lloyd, S. (1998). *The Phonics Handbook*. Jolly Learning Ltd, England, 3rd. edition.
- Maclean, M., Bryant, P., and Bradley, L. (1987). Rhymes, nursery rhymes, and reading in early childhood. *Merril-Palmer Quarterly*, 33(3):225–281.
- Macleod, I. and Lally, M. (1981). The effectiveness of computer controlled feedback in handwriting instruction. In Lewis, R. and Tagg, D., editors, *Computers In Education*, pages 291–296, Amsterdam. North-Holland Publishing Company.
- Major, N. (1995). Modelling teaching strategies. *Journal of Artificial Intelligence in Education*, 6(2/3):117–152.
- Major, N., Ainsworth, S., and Wood, D. (1997). Redeem: Exploiting symbiosis between psychology and authoring environments. *International Journal of Artificial Intelligence in Education*, 8:317–340.
- Mann, V. A. (1991). Phonological abilities: Effective predictors of future reading ability. In Rieben, L. and Perfetti, C. A., editors, *Learning to Read: Basic Research and Its Implications*, chapter 9. Lawrence Erlbaum Associates, Publishers, Hillsdale, New Jersey.
- Manning, C. D. and Schutze, H. (1999). *Foundations of Statistical Natural Language Processing*. The MIT Press, Cambridge, Ma.
- McGuinness, C. and McGuinness, G. (1998). *Reading Reflex*. Penguin Books, Harmondsworth, Middlesex, England, 1 edition.

- McGuinness, D. (1997). *Why Our Children Can't Read*. The Free Press, New York, 1st edition.
- Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., and Miller, K. (1993). Introduction to Wordnet: An On-line Lexical Database. <http://www.wordnet.com>.
- Mizoguchi, R., Sinita, K., and Ikeda, M. (1996a). Knowledge engineering of educational systems for authoring system design. In *Euro AIED-96*, pages 593–600, Lisbon.
- Mizoguchi, R., Sinita, K., and Ikeda, M. (1996b). Task ontology design for intelligent educational/training system. In *Workshop on Architectures and methods for Designing Cost-effective and Reusable ITSs, ITS96*, Montreal. <http://advlearn.lrdc.pitt.edu/its-arch/papers/mizoguchi.html>.
- Montessori, M. (1912). *The Montessori Method*. Burr Printing House, London, 1st edition.
- Moraes, J., Bertelson, P., Cary, L., and Alegria, J. (1986). Literacy training and speech segmentation. *Cognition*, 24:45–64.
- Morais, J., Cary, L., Alegria, J., and Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, 7:323–331.
- Morais, J. and Kolinsky, R. (1995). The consequences of phonemic awareness. In de Gelder, B. and Morais, J., editors, *Speech and Reading - A comparative Approach*, chapter 15, pages 317–337. Erlbaum (UK) Taylor & Francis, Publishers, East Sussex.
- Moseley, D. (1991). Breaking through the sound barrier - learning to read with a talking computer. In Satow, F. and Gatherer, B., editors, *Literacy Without Frontiers - Proceedings of the 7th European and 28th United Kingdom Reading Association Annual Conference*, pages 104–113, Heriot-Watt University, Edinburgh.
- Pain, H. (1981). A computer aid for spelling error classification in remedial teaching. In Lewis, R. and Tagg, D., editors, *Computers In Education*, pages 297–302, Amsterdam. North-Holland Publishing Company.
- Pain, H. (1985). *A computer tool for use by children with learning difficulties in spelling*. PhD thesis, University of Edinburgh, Department of Artificial Intelligence, Edinburgh, Scotland.

- Paris, S. and Parecki, A. (1993). Metacognitive aspects of adult literacy. Technical Report TR93-09, National Center of Adult Literacy, University of Michigan.
- Pennington, B. F., Orden, G. C. V., Smith, S. D., Green, P. A., and Haith, M. M. (1990). Phonological processing skills and deficits in adult dyslexics. *Child Development*, 61:1553–1778.
- Perfetti, C., Beck, I., Bell, L. C., and Hughes, C. (1987). Phonemic knowledge and learning to read are reciprocal: A longitudinal study of first grade children. *Merril-Palmer Quarterly*, 33(3):283–319.
- Perfetti, C., Beck, I., Bell, L. C., and Hughes, C. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly*, 23(3):263–284.
- Perfetti, C. A. and Marron, M. A. (1995). Learning to read - literacy acquisition by children and adults. Technical Report TR 95-07, National Center on Adult Literacy, University of Pittsburgh.
- Pratt, A. C. and Brady, S. (1988). Relation of phonological awareness to reading disability in children and adults. *Journal of Educational Psychology*, 80(3):319–323.
- Roget, P. M. (1994). *The Greenwich Roget's Thesaurus*. Greenwich Editions.
- Rozin, P. and Gleitman, L. R. (1977). The structure and acquisition of reading. II: The reading process and the acquisition of the alphabetic principle. In Reber, A. S. and Scarborough, D. L., editors, *Toward a Psychology of Reading - The Proceedings of the CUNY Conferences*, pages 55–141, Hillsdale, NJ. Lawrence Erlbaum Associates, Publishers.
- Shankweiler, D. P. (1991). The contribution of Isabelle Y. Liberman. In Brady, S. A. and Shankweiler, D. P., editors, *Phonological Processes in Literacy: a Tribute to Isabelle Y. Liberman*, chapter 0, pages xiii–xvii. Lawrence Erlbaum Associates, Publishers.
- Sharples, M. (1983). A construction kit for language. In Chandler, D., editor, *Exploring English with Microcomputers*, pages 51–58. Council for Educational Technology, Leicester, Great Britain.
- Smith, T. L. (1912). *The Montessori System in Theory and Practice*. Haper and Brother Publishers, London.

- Snowling, M. J. (1996). Annotation: Contemporary approaches to the teaching of reading. *Journal of Child Psychology and Psychiatry*, 37(2):139–148.
- Southgate, V. and Roberts, G. R. (1970). *Reading - which approach?* University of London Press Ltd, London, England.
- Sparck Jones, K. (1991). Two tutorial papers: Information retrieval and thesaurus. Technical report no. 234, University of Cambridge Computer Laboratory.
- Stanovich, K. E., Cunningham, A. E., and Cramer, B. B. (1984). Assessing phonological awareness in kindergarten children: Issues of task comparability. *Journal of Experimental Child Psychology*, 38:175–190.
- Townley, H. M. and Gee, R. D. (1980). *Thesaurus-Making*. Andre Deutsch Limited, London.
- Treiman, R. (1991). The role of intrasyllabic units in learning to read. In Rieben, L. and Perfetti, C. A., editors, *Learning to Read: Basic Research and Its Implications*, chapter 11, pages 149–160. Lawrence Erlbaum Associates, Publishers.
- Treiman, R. (1996). Onsets and rimes as units of spoken syllables: Evidence from children. *Journal of Experimental Child Psychology*, 39:161–181.
- Tunmer, W. E., Herriman, M. L., and Nesdale, A. R. (1988). Metalinguistic abilities and beginning reading. *Reading Research Quarterly*, 23(2):134–158.
- Tunmer, W. E. and Rohl, M. (1991). Phonological awareness and reading acquisition. In Sawyer, D. J. and Fox, B. J., editors, *Phonological Awareness in Reading*, volume 1 of *Language and Communication*, chapter 1, pages 2–30. Springer-Verlag, New York.
- Turner, T. C. (1993). Literacy and machines: An overview of the use of technology in adult literacy programs. Technical Report TR 93-3, National Center of Adult Literacy, University of Pennsylvania.
- VanLehn, K. (1988). *Student Modelling*, chapter 3, pages 55–78. Lawrence Erlbaum Associates, Publishers, New Jersey.
- Vellutino, F. R. and Scanlon, D. M. (1987). Phonological coding, phonological awareness, and reading ability: Evidence from a longitudinal and experimental study. *Merrill-Palmer Quarterly*, 33(3):321–363.

Wenger, E. (1987). *Artificial Intelligence and Tutoring Systems*. Morgan Kaufmann Publishers, Inc, Los Altos, California.

Yopp, H. K. (1988). The validity and reliability of phonemic awareness tests. *Reading Research Quarterly*, 23:159-177.

Appendix A

Literacy Teaching Dictionary Format and Example

Each entry of the dictionary developed in this project is in the following format: a number of tagged records, a brief description of the tagged records, and a number of public records, if any, of the records.

A set format of the tagged records follow:

- * [W1] - The word, the writer, a personal or public record.
- * [L1] - The frequency of occurrence of the letter in the dictionary itself and a number of public records, if any.
- * [P1] - The phonetic representation of the letter in the Literacy Teaching System.
- * [F1] - Inflected form: plural of singular.
- * [F2] - The frequency of the plural form.
- * [P2] - The phonetic representation of the plural form.
- * [I1] - Inflected form: third person singular of verb.
- * [F3] - The frequency of the third person singular.
- * [P3] - The phonetic representation of the third person singular.
- * [I2] - Inflected form: plural of noun and third person singular of verb.
- * [F4] - The frequency of the form itself.
- * [P4] - The phonetic representation of the form itself.
- * [I3] - Inflected form: the personal of the verb.
- * [F5] - The frequency of the personal.

* [W2] - The word, the writer, a personal or public record.

Appendix A

Literacy Teaching Dictionary Format and Example

Each entry of the dictionary developed in this research corresponds to a “headword” and consists of a number of tagged records. The tags come in brackets ahead of the records.

A description of the tagged records follows:¹

- [HWD] - Headword: the written representation of the main word of the entry.
- [FWD] - The frequency of occurrence of the headword (obtained in a counting in the dictionary itself and a number of public corpora used for this end).
- [PWD] - The phonetic representation of the headword (according to the Festival System).
- [IF1] - Inflected form: plural of noun only.
- [FF1] - The frequency of the plural form.
- [PF1] - The phonetic representation of the plural form.
- [IF2] - Inflected form: third person singular of verb only.
- [FF2] - The frequency of the third person singular.
- [PF2] - The phonetic representation of the third person singular.
- [IF3] - Inflected form: plural of noun and third person of singular of verb.
- [FF3] - The frequency of the form above.
- [PF3] - The phonetic representation of the form above.
- [IF4] - Inflected form: the gerund of the verbs.
- [FF4] - The frequency of the gerund.

¹This structure is based on the Cobuild II dictionary, with some extensions. We have listed and described only the records that are currently relevant for this project.

- [PF4] - The phonetic representation of the gerund.
- [IF5] - Inflected form: past tense and past participle of verb.
- [FF5] - The frequency of the form above.
- [PF5] - The phonetic representation of the form above.
- [IF6] - Inflected form: past tense of verb only.
- [FF6] - The frequency of the form above.
- [PF6] - The phonetic representation of the form above.
- [IF7] - Inflected form: past participle of verb only.
- [FF7] - The frequency of the form above.
- [PF7] - The phonetic representation of the form above.
- [IF8] - Inflected form: comparative of adjective.
- [FF8] - The frequency of the form above.
- [PF8] - The phonetic representation of the form above.
- [IF9] - Inflected form: superlative of adjective.
- [FF9] - The frequency of the form above.
- [PF9] - The phonetic representation of the form above.
- [SYN] - A synonym of the headword.
- [FYN] - The frequency of the synonym.
- [PYN] - The phonetic representation of the synonym.
- [DEF] - The definition of the word. This definition might come from two different dictionaries. If the sentence is preceded by the plus sign, it comes from the Collins English Dictionary. If it is preceded by an asterisk it comes from the Collins Cobuild II.
- [EXA] - The sentences in this field are examples of use of the headwords or of any of the derived forms of the word.
- [NSY] - The number of syllables of the word.

An example of an entry of the dictionary follows. As some fields of the entry are optional not all of them are present in the example.

- [HWD]enclose
- [FWD]000074
- [PWD]/i/n/k/1/ou/z
- [IF2]encloses

- [FF2]000021
- [PF2]/i/n/k/l/ou/z/i/z
- [IF4]enclosing
- [FF4]000049
- [PF4]/i/n/k/l/ou/z/i/ng
- [IF5]enclosed
- [FF5]000195
- [PF5]/i/n/k/l/ou/z/d
- [SYN]include
- [FYN]000137
- [PYN]/i/n/k/l/uu/d
- [DEF]+to close; hem in; surround+to surround (land) with or as if with a fence+to put in an envelope or wrapper, esp— together with a letter+to contain or hold*If a place or object is enclosed by something, the place or object is inside that thing or completely surrounded by it.*If you enclose something with a letter, you put it in the same envelope as the letter.
- [EXA]*The rules state that samples must be enclosed in two watertight containers...*Enclose the pot in a clear polythene bag...*The surrounding land was enclosed by an eight foot wire fence.*...the enclosed waters of the Baltic.*I have enclosed a cheque for 10 pounds...*He tore open the creamy envelope that had been enclosed in the letter...*The enclosed leaflet shows how Service Care can ease all your worries.
- [NSY]2

Appendix B

Survey for Evaluating Relevance of the Target Vocabulary

B.1 Survey description

The aim of this survey is to evaluate the proportion of thematically relevant words obtained in the various thematic selections carried out in this research.

The size of the selected vocabularies range from around eight hundred to six thousand words. A normalised random sample of a hundred words has been used in order to evaluate their relevance to the proposed themes.

In the first version of the survey, the samples were printed on paper and a small group of students was invited to determine, for each word in each sample, whether they were (a) - relevant to the proposed theme, (b) - irrelevant to the proposed theme or (c) - ignored by the respondent (either the word or its relevance to the theme). The printed samples were introduced by a short text explaining the context the words were supposed to belong to, as follows:

“Dear friend, this is going to take you around 3 minutes... The following words are supposedly related to “food” in some way.

- Please put an “x” in the space before those you think would not come up in a discussion related to “food”.

- If you don't know some word, please mark it with a hyphen.

Thanks for your help, please send it back to ...”

The survey was processed manually. The respondents were asked to identify themselves, so any particular further classification (by age, or native language) could be obtained and used later on in the processing.

In order to do this more efficiently, a second version of the survey on the Internet was implemented. A JAVA Applet was developed to automatically offer the respondents a selection of different vocabularies to be evaluated. In this

survey, the respondents are classified by age group (younger than 18, from 18 to 35 and older than 35) and by native language (English native speaker or not). The Applet can be started in any computer connected to the Internet provided with a common Internet browser. The identification window is shown below.

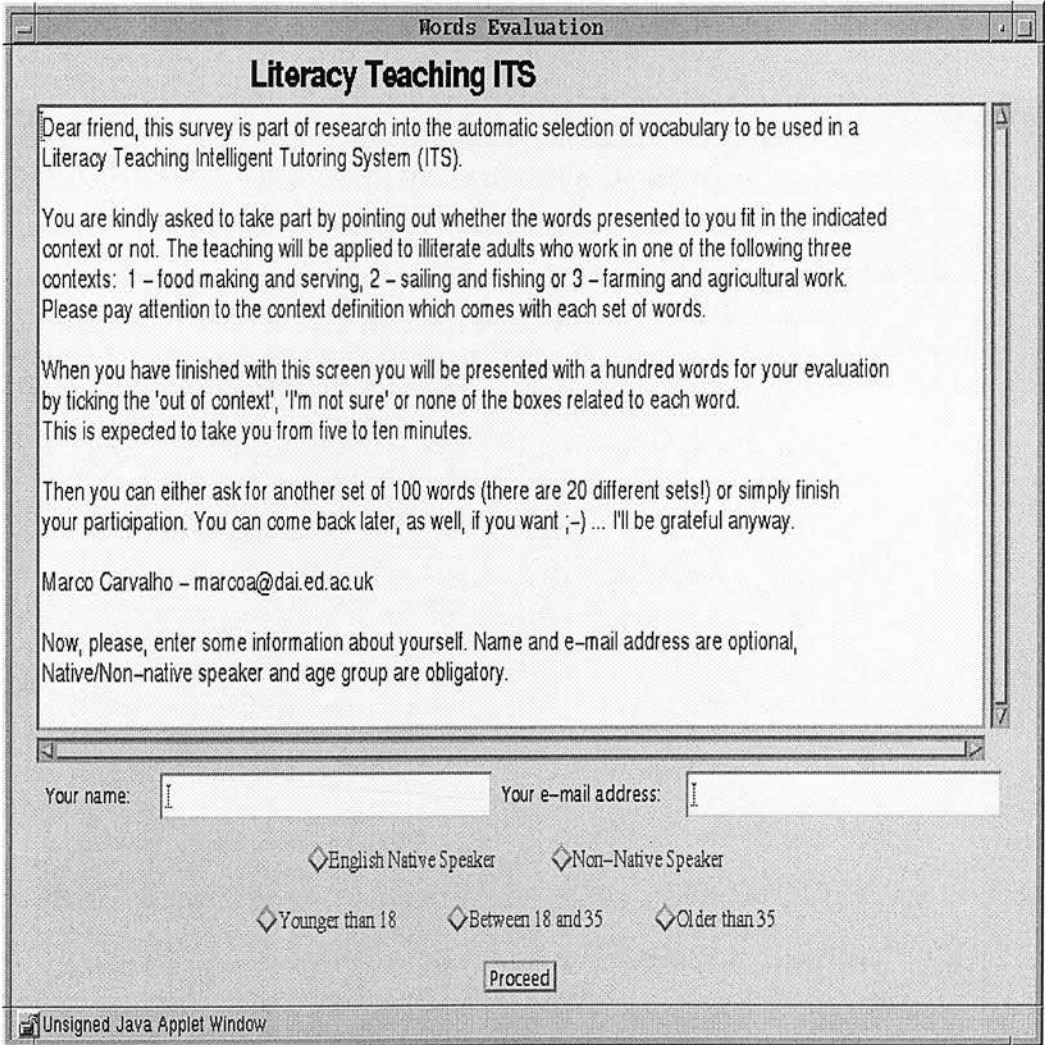


Figure B.1: Survey identification window

After the identification the respondent is provided with a set of words to be evaluated according to the proposed context. These words come on the screen with buttons to determine whether they are out of context or they are ignored. The data screen is shown in Figure B.2.

This version of the survey can be turned into a powerful research tool since it enables the evaluation to be carried out in a much larger scale, classified by native language and age groups. A program to automatically select and interpret the answers program is also available. However, due to the number of respondents,

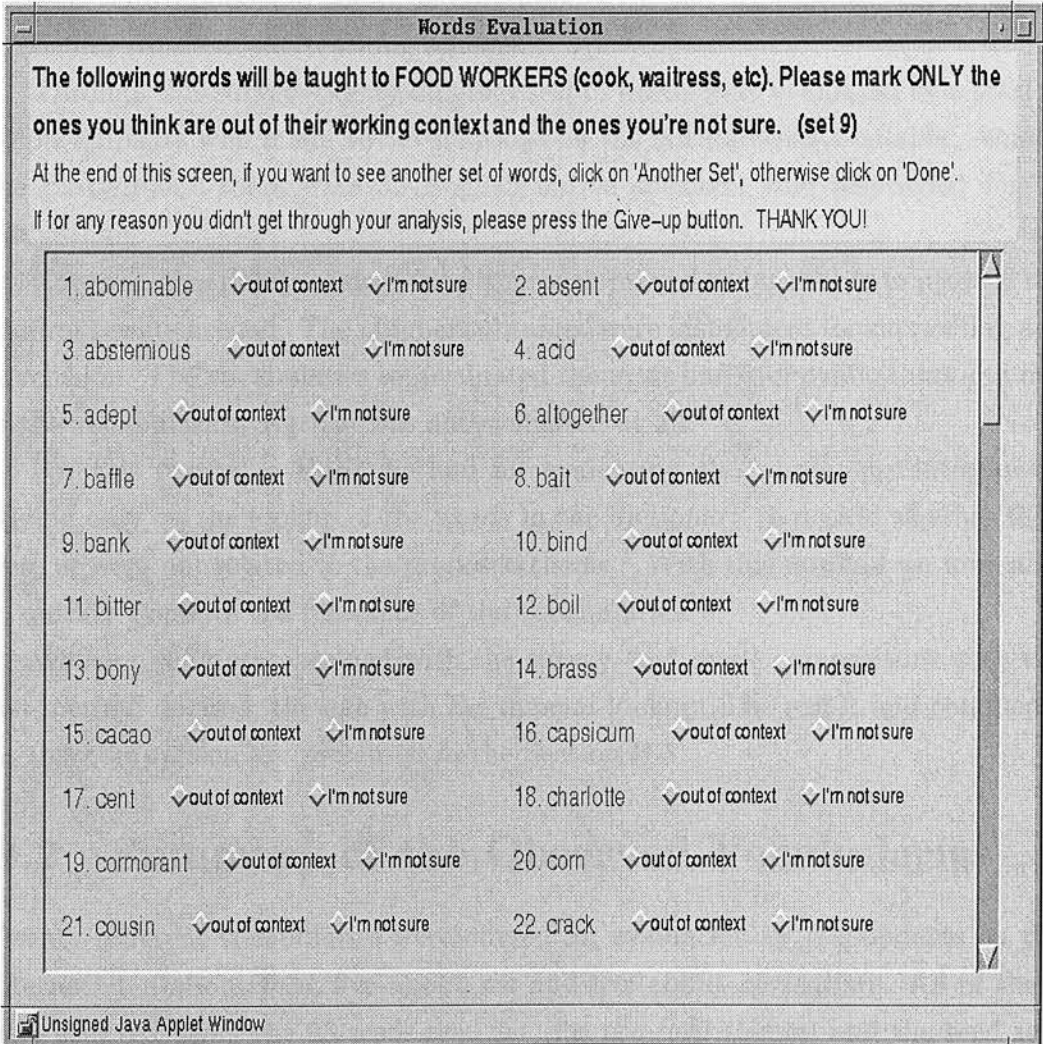


Figure B.2: Survey data window

we were unable to obtain enough reliable data to effectively use in the research. In order to be statistically significant it would be necessary to address a public much larger than the one we were able to. And besides, when addressing a more general public, some analysis on the adequacy of the data would have to be developed to filter incomplete and non-serious answers that could be obtained. In the long run, with quantitatively satisfactory data, the age and the language class of the respondents could be correlated with the appropriateness of use of particular words in the teaching scheme.

Through this survey, twenty different sets of words were evaluated by a total of 53 participants who made 107 evaluations on the vocabularies available. Among the participants there were 32 English native speakers who responded for 76 evaluations.

However, this public consultation approach proved to be unable to provide the prompt results needed. The numbers obtained were insufficient for supporting any conclusion. The vocabularies set evaluated the most had four evaluations and one of the vocabularies (Agri2s) was not evaluated at all.

For this reason, a third method for evaluating the words was introduced, relying only on the lookup of the words in the dictionary to decide whether they were or were not related to the proposed theme.¹ With this method we were able to quickly evaluate the relevance of the vocabularies.

We kept the data obtained with the survey, and made comparisons with the last method devised, the one with the manual lookup. The result and comments on this comparison are presented in the Section B.3.

B.2 Summary of the Observed Vocabularies

Twenty different vocabularies were offered for evaluation by respondents on the Internet: ten about *food*, five about *sea* and five about *agriculture*. All of them were selected using the 93 seeds seed set (for the *food* theme) and the seed sets for *sea* and *agriculture* presented in Section 4.2 - The Thematic Selection.

A summary of the characteristics of each vocabulary and the relevance indexes follows. In the summary, the field *selection method* gives the following information: the number of seeds required for a word to be selected, whether the sentences with examples of use of the words have been scanned for seeds or not, whether not-words (words that should not be present in the description

¹As the decision about the appropriateness of the words was taken based on the frequency of appearance of the words in public corpora, in principle, every word in the dictionary could be considered appropriate.

of the entries) have been considered, and whether multiple occurrences of seeds were counted (relax) or not (strict). For some of the selections, the information *lower case* means that only *head-words* starting with lower case were considered in that particular selection. The field *irrelevant words* contains the percentage of words considered irrelevant to the theme in focus by more than 30 percent of the respondents. The field *unknown words* contains the percentage of words declared as unknown by more than 30 percent of the respondents. The field *unquestioned words* contains percentage of words that no respondent considered neither irrelevant nor unknown.

Vocabulary: Food1; total of words: 3355

Selection method: 1 seed, examples, not words, strict.

Total respondents: 4

Irrelevant words: 31%

Unknown words: 4%

Unquestioned words: 28%

Vocabulary: Sea1; total of words: 3107

Selection method: 1 seed, examples, not words, strict.

Total respondents: 2

Irrelevant words: 32%

Unknown words: 39%

Unquestioned words: 46%

Vocabulary: Agri1; total of words: 6708

Selection method: 1 seed, examples, not words, strict.

Total respondents: 4

Irrelevant words: 33%

Unknown words: 2%

Unquestioned words: 39%

Vocabulary: Food2; total of words: 1703

Selection method: 2 seeds, examples, not words, relax.

Total respondents:

Irrelevant words: 17%

Unknown words: 1%

Unquestioned words: 58%

Vocabulary: Sea2; total of words: 1453

Selection method: 2 seeds, examples, not words, relax.

Total respondents: 3

Irrelevant words: 13%

Unknown words: 6%

Unquestioned words: 56%

Vocabulary: Agri2; total of words: 3323

Selection method: 2 seeds, examples, not words, relax.

Total respondents: 4

Irrelevant words: 35%

Unknown words: 11%

Unquestioned words: 24%

Vocabulary: Food2l; total of words: 1589

Selection method: 2 seeds, examples, not words, relax, lower case.

Total respondents: 4

Irrelevant words: 34%

Unknown words: 2%

Unquestioned words: 46%

Vocabulary: Sea2l; total of words: 1138

Selection method: 2 seeds, relax, examples, not words, lower case.

Total respondents: 3

Irrelevant words: 48%

Unknown words: 6%

Unquestioned words: 27%

Vocabulary: Agri2l; total of words: 3011

Selection method: 2 seeds, examples, not words, relax, lower case.

Total respondents: 2

Irrelevant words: 69%

Unknown words: 11%

Unquestioned words: 22%

Vocabulary: Food2s; total of words: 1467

Selection method: 2 seeds, examples, not words, strict.

Total respondents: 2

Irrelevant words: 38%

Unknown words: 9%

Unquestioned words: 55%

Vocabulary: Sea2s; total of words: 1122

Selection method: 2 seeds, examples, not words, strict.

Total respondents: 3

Irrelevant words: 28%

Unknown words: 9%

Unquestioned words: 37%

Vocabulary: Agri2s; total of words: 2662

Selection method: 2 seeds, examples, not words, strict.
Total respondents: 0
Irrelevant words: -
Unknown words: -
Unquestioned words: -

Vocabulary: Food2sl; total of words: 1363
Selection method: 2 seeds, strict, examples, not words, lower case.
Total respondents: 2
Irrelevant words: 23%
Unknown words: 13%
Unquestioned words: 68%

Vocabulary: Sea2sl; total of words: 856
Selection method: 2 seeds, examples, not words, strict, lower case.
Total respondents: 4
Irrelevant words: 9%
Unknown words: 6%
Unquestioned words: 61%

Vocabulary: Agri2sl; total of words: 2419
Selection method: 2 seeds, examples, not words, strict, lower case.
Total respondents: 1
Irrelevant words: 35%
Unknown words: 8%
Unquestioned words: 57%

Vocabulary: Food2w; total of words 1234
Selection method: 2 seeds, without examples, not words, relax.
Total respondents: 3
Irrelevant words: 16%
Unknown words: 1%
Unquestioned words: 72%

Vocabulary: Food2sw; total of words: 1002
Selection method: 2 seeds, without examples, not words, strict.
Total respondents: 3
Irrelevant words: 11%
Unknown words: 0%
Unquestioned words: 73%

Vocabulary: Food3; total of words: 1086
Selection method: 3 seeds, examples, not words, relax.
Total respondents: 3
Irrelevant words: 4%

Unknown words: 7%
Unquestioned words: 78%

Vocabulary: Food3s;
Subject: Food; total of words: 846
Selection method: 3 seeds, examples, not words, strict.
Total respondents: 4
Irrelevant words: 9%
Unknown words: 6%
Unquestioned words: 70%

Vocabulary: Food3l; total of words: 1012
Selection method: 3 seeds, examples, not words, relax, lower case.
Total respondents: 4
Irrelevant words: 11%
Unknown words: 6%
Unquestioned words: 60%

B.3 Comparison of Evaluations

The vocabularies evaluated through the survey on the Internet were also evaluated through the manual lookup process. Table B.1 compares the results.

As stated before, the data obtained in the Internet Survey are not conclusive because of the relatively small number of answers obtained. In Table B.1, the selections that come first are less restrictive than the ones that come later, so a decrease in the rates of irrelevant words towards the bottom of the table should be expected. Actually, this tendency is generally verified in the column related to the manual lookup. The data for the column related to the survey, although correlating broadly with the tendency of reducing the rates of irrelevant words as the selections requires tougher conditions, are much less stable.

This is so because in some cases, only one or two persons evaluated a particular vocabulary. Our hypothesis, is that different people could be more prepared to identify words belonging to one context than to another, so they could misjudge or not be motivated to evaluate a particular context, influencing negatively the assessment. In a more extensive poll, questions such as “would you be happy in evaluating a vocabulary related to the theme *sea*?” could be asked, giving the respondent the option to answer on subjects he or she is more interested in or more familiar with. We could not, however, go that far in the preparation of the survey. As said above, we decided to show the data and keep all the information on the survey in this thesis for what it may represent in future research.

Appendix C

Festival System Phonemes

Vocabulary Name	Percent of Irrelevant Words (Lookup)	Percent of Irrelevant Words (Survey)
Food1	46	31
Sea1	49	32
Agri1	48	33
Food2	22	17
Sea2	25	13
Agri2	28	35
Food2l	16	34
Sea2l	20	48
Agri2l	16	69
Food2s	16	38
Sea2s	20	28
Agri2s	-	-
Food2sl	16	23
Sea2sl	15	9
Agri2sl	20	35
Food2w	10	16
Food2sw	10	11
Food3	7	4
Food3s	11	9
Food3l	10	11

Table B.1: Comparison of Evaluations

Appendix C

Festival System Phonemes

Number	Example Word	Phono-Graphix	Festival System
1	bad, boy, bat	b	b
2	day, dog, dish	d	d
3	fat, fan, fox	f	f
4	hit, hot, hot	h	h
5	keep, key, kiss	k	k
6	loud, log, life	l	l
7	mad, man, me	m	m
8	no, not, not	n	n
9	pit, pan, pear	p	p
10	run, red, root	r	r
11	sit, sit, sorry	s	s
12	trap, top, tip	t	t
13	very, vet, view	v	v
14	wash, win, win	w	w
15	zoo, zip, zero	z	z
16	go, get, get	g	g
17	chair, chin, chip	ch	ch
18	joke, jog, june	j	jh
19	sing, ring, ring	ng	ng
20	theatre, thin, thin	th	th
21	that, them, this	th/	dh
22	sugar, ship, she	sh	sh
23	garage, vision, vision	zh	zh
24	yellow, yes	-	y

(ctd...)

Appendix D

Number	Example Word	Phono-Graphix	Festival System
25	bat, hat, cat	a	a
26	end, bet, bed	e	e
27	about, ago	-	@
28	ill, sit, sit	i	i
29	pot, hot, of	o	o
30	up, cut, run	u	uh
31	foot, book, put	oo	u
32	car, car, arm	ar	aa
33	beat, seem, see	ee	ii
34	port, law, saw	aw	oo
35	urn, her, her	er	@@
36	food, soon, too	/oo	uu
37	late, cake, day	ae	ei
38	ice, time, my	ie	ai
39	oil, oil, boy	oi	oi
40	boat, home, no	oe	ou
41	out, out, how	ou	au
42	here, near	-	i@
43	air, hair	-	e@
44	poor	-	u@

Appendix D

Knowledge Tree Codes

D.1 The Phono-Graphix Knowledge Tree

1 - Reading: Code Knowledge, Segmenting, Blending.

&source=foodsmall.

1.1 - Basic Code.

VC. CVC. &exclude.

1.1.1 - Fat Cat Sounds.

[/f,/k,/s,/p,/m,/t] {/a,/o,/aa} {/t,/p,/m,/s}.

&source=foodbig.

1.1.2 - Bug On Jug Sounds.

[/r,/h,/b,/jh,/d,/m,/t] {/uh,/i,/a} {/g,/d,/m}.

&source=foodbig.

1.1.3 - Ben Bon Sounds.

[/h,/b,/w,/l,/n,/f,/z,/s]{/e,/o,/i,/uh}{/l,/n,/t,/z,/b,/p}.

1.2 - Adjacent Consonants: Auditory Processing.

1.2.1 - VCC Words.

VCC. &source=foodbig.

1.2.2 - CVCC Words.

CVCC.

1.2.3 - CCVC Words.

CCVC.

1.3 - Advanced Code: Reading Coded Text, Reading Uncoded Text.

1.3.1 - More Than One Letter Sounds.

1.3.1.1 - Ch And Sh Sounds.

{s,c}{h}{\${}}. {s,c}{h}{\${}}{\${}}. {\${}}{s,c}{h}{\${}}.

{\${}}{s,c}{h}{\${}}{\${}}. {\${}}{\${}}{s,c}{h}{\${}}.

{\${}}{\${}}{s,c}{h}. {\${}}{\${}}{\${}}{s,c}{h}. {\${}}{\${}}{\${}}{\${}}{s,c}{h}.

1.3.1.2 - Th Sounds.

&source=foodbig.

th\$. th\$\$\$. \$th\$. \$\$th. \$\$\$th.

1.3.1.3 - Ck Sounds.

\$\$ck. \$\$ck\$. \$\$\$ck. &source=foodbig.

1.3.1.4 - Double L Sounds.

\$ll. \$ll\$. \$\$ll. \$\$\$ll. \$\$\$ll\$.

1.3.2 - Multiple Representation For A Sound: Homonymous Processing.

1.3.2.1 - Ou Sounds. % for boat, note, cone, snow, etc.

{/ou}{*}. {X}{/ou}{*}. {X}{X}{/ou}. {X}{X}{/ou}{*}. {*}{/ou}.

1.3.2.2 - Au Sounds. % for how owl, cow, town, house, ground, etc.

{/au}{*}. {X}{/au}{*}. {X}{X}{/au}{*}. {*}{/au}.

1.3.2.3 - Er Sounds. % for her, hair, girl, shirt, burn, bird,...

{/e@}{*}. {X}{/e@}{*}. {X}{X}{/e@}{*}. {*}{/e@}.

1.3.2.4 - R Sounds. % for rat, wreck, ride, wrap, trap, risk...

{/r}{X}. {/r}{X}{X}. {/r}{X}{X}{X}. {X}{/r}{X}.

{X}{/r}{X}{X}. {X}{X}{/r}{X}.

1.3.2.5 - Ee Sounds. % for see, east, chief, knee, leave...

{/ii}{X}. {/ii}{X}{X}. {/ii}{X}{X}{X}. {X}{/ii}{X}.

{X}{/ii}{X}{X}. {X}{X}{/ii}{X}. {X}{/ii}.

{X}{X}{/ii}. {X}{X}{X}{/ii}.

1.3.2.6 - Ei Sounds.

\$a\$e. \$a\$e\$. \$\$a\$e. \$\$a\$e\$.

{/ei}{*}. {X}{/ei}{X}. {X}{/ei}{X}{X}. {X}{X}{/ei}{X}.

{X}{X}{/ei}{X}{X}. {*}{/ei}.

1.3.2.7 - Do Sounds.

{/u}{*}. {X}{/u}{X}. {X}{/u}{X}{X}. {X}{X}{/u}{*}. {*}{/u}.

1.3.2.8 - U Sounds.

{/uh}{*}. {X}{/uh}{X}. {X}{X}{/uh}{X}. {*}{/uh}.

1.3.2.9 - E Sounds.

{/e}{X}. {/e}{X}{X}. {/e}{X}{X}{X}. {X}{/e}{X}.

{X}{/e}{X}{X}. {X}{X}{/e}{X}. {*}{/e}.

1.3.2.10 - Ai Sounds.

{/ai}{X}. {/ai}{X}{X}. {X}{/ai}. {X}{/ai}{X}.

{X}{X}{/ai}. {X}{X}{/ai}{X}. {*}{/ai}.

1.3.2.11 - S Sounds.

{/s}{X}. {/s}{X}{X}. {/s}{X}{X}{X}. {X}{X}{/s}{X}. {X}{X}{/s}.

1.3.2.12 - Z Sounds.

{/z}{*}. {X}{/z}{*}. {X}{X}{/z}{*}. {*}{/z}.

1.3.2.13 - L Sounds.

{/l}{X}. {/l}{X}{X}. {/l}{X}{X}{X}. {X}{/l}{X}. {X}{/l}{X}{X}.
{X}{X}{/l}. {X}{X}{X}{/l}.

1.3.3 - Multiple Sounds For One Representation: Nonsense Word Spelling.

1.3.3.1 - Words With Ow.

{o}{w}{*}. {\$}{o}{w}{*}. {\$}{\$}{o}{w}{*}. {\$}{\$}{\$}{o}{w}{*}.

1.3.3.2 - Words With A.

{/a}{X}. {/a}{X}{X}. {/ei}{X}. {/ei}{X}{X}. {X}{/a}{X}.
{X}{/ei}{X}. {X}{/a}{X}{X}. {X}{/ei}{X}{X}.

1.3.3.3 - Words With Ea.

{e}{a}{*}. {\$}{e}{a}{*}. {\$}{\$}{e}{a}{*}.

1.4 - Multi-Syllable: Reading Coded Text, Word Splitting.

1.4.1 - Chunking.

XXX/i/ng. XXX/@/r. XXX/o/n. XXX/ii. XXXX/l/ii.
XXX/p/l. XX/ei/n. XXX/ei/n.

1.4.2 - Special Endings: High Frequency Words Processing.

1.4.2.1 - Ate Verbs.

*/ei/t. % as generate, elucidate...

1.4.2.2 - Adjectives.

*/@/n. % as in American
*/@/n/z. % as in Brazilians
*/ii/@/s. % as in rosaceous
*/@/b/l. % as in eatable
*/@/n/t. % as in vibrant

1.4.2.3 - Past Tense.

&source=foodbig. */t/i/d. */r/i/d. */z/d. */ch/t.
\$\$\$\$ed. \$\$\$\$\$ed. \$\$\$\$\$\$ed. \$\$\$\$\$\$\$ed.

1.4.2.4 - Action.

*/sh/@/n. &source=foodbig.

1.4.2.5 - Gerund/Present Continuous.

*/i/ng.

1.4.2.6 - Common Nouns And Adjectives.

&source=foodbig. */i/z/@/m. */@/n/t.

1.4.2.7 - Ly Adverbs.

&source=foodbig. */l/ii.

D.2 The Jolly-Phonics Knowledge Tree

% Jolly Phonics Knowledge Tree

% and

% Vocabulary Selection Source File

1 - Reading.

&source=foodsmall.

1.1 - Satipn Sounds: Letter Recognition, Blending, Identifying Sounds.

&G={/s,/a,/t,/i,/p,/n}.

1.1.1 - First Satipn Words.

GG. GGG. % two and three phonemes only.

1.1.2 - More Satipn Words. % with 4, 5 and 6 phonemes.

GGGG.

GGGGG.

GGGGGG.

1.2 - Chrmd Sounds: Letter Recognition, Blending, Identifying Sounds.

1.2.1 - First Chrmd Words. % with the vowel /e

&G={/c,/e,/h,/r,/m,/d}.

GG. GGG.

1.2.2 - More Chrmd Words. % with the already known vowels

&G={/c,/e,/h,/r,/m,/d,/a,/i}.

GG. GGG. GGGG. GGGGG. GGGGGG. % from 2 to 6 phonemes.

1.2.3 - Satipn Plus Chrmd Sounds. % all the already known sounds.

&G={/c,/e,/h,/r,/m,/d,/s,/a,/t,/i,/p,/n }.

GG. GGG. GGGG. GGGGG. GGGGGG.

1.2.4 - Tricky Satipn Plus Chrmd Sounds: Tricky Words.

&G={/c,/e,/h,/r,/m,/d,/s,/a,/t,/i,/p,/n }.

GG. GGG. GGGG. GGGGG. GGGGGG.

&source=tricky.

1.3 - Goulfb Sounds: Letter Recognition, Blending, Identifying Sounds.

1.3.1 - First Goulfb Words.

&G={/g,/o,/uh,/l,/f,/b}. % Jolly Phonics: g o u l f b

GG. GGG. GGGG. % with two, three and four phonemes only.

1.3.2 - More Goulfb Words. % with all the known vowels

&G={/g,/o,/uh,/l,/f,/b,/e,/a,/i}.

GG. GGG. GGGG. GGGGG. GGGGGG.

1.3.3 - CVC Goulfb Plus Chrmd Plus Satipn Words.

&G={/g,/o,/uh,/l,/f,/b,/s,/t,/p,/n,/c,/h,/r,/m,/d}.

&H={/a,/i,/e,/o,/uh}.

GHG.

1.4 - Aioa Sounds: Letter Recognition, Blending, Identifying Sounds.

&H={/ei,/j,/ou,/ai,/ii,/oo}. % Jolly Phonics: ai j oa ie ee or

&G={/s,/t,/p,/n,/c,/h,/r,/m,/d,/j}.

1.4.1 - VC And CVC Aioa Plus Known Consonants.

HG. GHG.

1.4.2 - CCVC And CVCC Aioa Plus Known Consonants.

GGHG. GHGG.

1.5 - Zwng Sounds.

&G={/z,/w,/ng,/v}. % Jolly Phonics: z w ng v lit oo long oo

&H={/u,/uu,/a,/i,/e,/o,/uh,/ei,/ou,/ai,/ii,/oo}.

&I={/u,/uu,/a,/i,/e,/o,/uh,/ei,/ou,/ai,/ii,/oo,/z,/w,/ng,/v}.

1.5.1 - VC And CVC Zwng Sounds.

HG. GHG.

1.5.2 - Other Zwng Sounds.

III. IIII. IIIIII. IIIIIII. IIIIIIII.

1.5.3 - Words Ending With Zwng Sounds.

{*}{H}{H}{G}.

{*}{G}{H}{G}.

1.6 - Yxchshth Sounds.

&G={/y,/x,/ch,/sh,/dh,/th}. % Jolly Phonics: y x ch sh voi th unv th

&H={/u,/uu,/a,/i,/e,/o,/uh,/ei,/ou,/ai,/ii,/oo}.

&I={/u,/uu,/a,/i,/e,/o,/uh,/ei,/ou,/ai,/ii,/oo,/y,/x,/ch,/sh,/dh,/th}.

1.6.1 - VC And CVC And CV Yxchshth Sounds.

HG. GHG. GH.

1.6.2 - Words Ending With Yxchshth Sounds.

{*}{I}{G}.

1.6.3 - Words Starting With Yxchshth Sounds.

{G}{I}{*}.

1.7 - Quou Sounds.

&G={/k,/w,/au,/oi,/y,/uu,/@@,/aa}. %JP: qu ou oi ue er ar

&H={/u,/uu,/a,/i,/e,/o,/uh,/ei,/ou,/ai,/ii,/oo,/y}.

&I={/u,/uu,/a,/i,/e,/o,/uh,/ei,/ou,/ai,/ii,/oo,/y,/@@,/aa,/oi,/au}.

1.7.1 - VC And CVC And CV Quou Sounds.

HG. GHG. GH.

1.7.2 - Words Ending With Quou Sounds.

{*}{I}{G}.

1.7.3 - Words Starting With Quou Sounds.

{G}{I}{X}{X}. {G}{I}{X}.

1.8 - Alternative Spellings.

1.8.1 - Long A.

{C}{/ei}. {C}{/ei}{C}.

1.8.2 - Long E.

{C}{/ii}. {C}{/ii}{C}.

1.8.3 - Long I.

{C}{/ai}. {C}{/ai}{C}.

1.8.4 - Long O.

{C}{/ou}. {C}{/ou}{C}. {C}{C}{/ou}. {C}{V}{C}{/ou}.

1.8.5 - Long U.

{C}{/y}{/uu}{*}. {/y}{/uu}{*}.

1.8.6 - Little OO.

{C}{/u}{C}.

1.8.7 - Long OO.

{C}{/uu}{C}.

1.8.8 - ER Sound.

{*}{/@@}. {C}{/@@}{C}.

1.8.9 - OR Sound.

{C}{/oo}. {C}{/oo}{C}.

1.8.10 - OI Sound.

{C}{/oi}. {C}{/oi}{C}.

1.8.11 - OU Sound.

{C}{/ou}. {C}{/ou}{C}.

1.9 - Blending.

1.9.1 - Initial Consonant Blending.

1.9.1.1 - Xl Blending.

{/b,/c,/f,/p,/s}{/l}{*}.

1.9.1.2 - Xr Blending.

{/b,/c,/d,/f,/g,/p,/t,/sh,/th}{/r}{*}.

1.9.1.3 - Sx Blending.

{/s}{/t,/c,/m,/n}{*}.

1.9.1.4 - TW Blending.

{/t}{/w}{*}.

1.9.1.5 - Sxr Blending.

{/s}{/c,/p,/t}{/r}{*}.

1.9.2 - Final Consonant Blending.

1.9.2.1 - Lx Blending.

{*}{/l}{/b,/d,/f,/m,/n,/p}.

1.9.2.2 - Xt Blending.

{*}{/c,/f,/n,/p,/x}{/t}.

1.9.2.3 - MPandNd Blendings.

{*}{/m}{/p}. {*}{/n}{/d}.

1.10 - Rhymes.

1.10.1 - Ink Rhymes.

{*}{/i}{/ng}{/k}.

1.10.2 - Ice Rhymes.

{*}{/ai}{/s}.

1.10.3 - Us Rhymes.

{*}{/uh}{/s}.

1.10.4 - Ock Rhymes.

{*}{/o}{/k}.

1.10.5 - In Rhymes.

{*}{/i}{/n}.

1.10.6 - Op Rhymes.

{*}{/o}{/p}.

1.11 - Irregular Words: Tricky Words.

1.11.1 - All Tricky.

&source=tricky.

{*}.

D.3 The THRASS Knowledge Tree

% THRASS Knowledge Tree

% and

% Vocabulary Selection Source File

1 - Literacy Teaching: Handwriting, Reading,
Spelling, Assessing.

&source=foodsmall.

1.1 - Consonant Phonemes.

1.1.1 - Phoneme B.

{/b}{V}{C}. % bird

b\$\$.

\$bb{*}.

\$\$bb{*}. % rabbit

1.1.2 - Phoneme C.

{/k}{V}{C}. % cat

{/k}{V}{C}{C}. % kitten

{C}{V}{/k}. % duck

{C}{/k}{V}{C}. % school

{/k}{V}{V}{C}. % queen

1.1.3 - Phoneme Ch.

{/ch}{V}{C}. % chair

{X}{X}{/ch}. % watch

1.1.4 - Phoneme D.

{/d}{V}{C}. % dog

{C}{V}{/d}{X}. % ladder

1.1.5 - Phoneme F.

{/f}{V}{C}. % fish

{C}{V}{/f}{C}. % coffee

{C}{V}{C}{/f}{*}.

ph\$\$\$.

\$\$ph.

\$\$\$ph.

\$\$ph\$.

- \$\$\$ph\$\$\$. % dolphin
 ph*.
 *ph.
- 1.1.6 - Phoneme G.
 {/g}{V}{C}. % gate
 {V}{/g}. % egg
 {C}{C}{V}{/g}.
 {X}{X}{/g}{X}.
 {X}{X}{/g}{X}{X}.
- 1.1.7 - Phoneme H.
 {/h}{V}{C}{C}. % hand.
 {/h}{V}{C}.
- 1.1.8 - Phoneme J.
 {/jh}{X}{X}. % jam
 {/jh}{X}{X}{X}.
 {/jh}{X}{X}{X}{X}.% giant
 {C}{V}{/jh}. % cage
 {C}{C}{V}{/jh}. % bridge
- 1.1.9 - Phoneme L.
 {/l}{V}{C}. % leg
 {C}{V}{/l}. % bell
- 1.1.10 - Phoneme M.
 {/m}{V}{C}.
 {C}{V}{/m}. % lamb
 {C}{V}{/m}{V}. % hammer
 \$\$\$mm*.
- 1.1.11 - Phoneme N.
 {/n}{V}{C}. % net
 {C}{V}{/n}{V}. % dinner
 {/n}{V}. % knee
- 1.1.12 - Phoneme Ng.
 {X}{/ng}. % ink
 {X}{X}{/ng}. % king
 {X}{X}{X}{/ng}.
- 1.1.13 - Phoneme P.
 {/p}{V}{C}.
 {/p}{V}{C}{C}{V}. % panda

- \$\$\$pp\$. \$pp\$. \$pp\$\$\$. \$\$\$pp\$\$\$. \$\$\$pp\$. % hippo
 1.1.14 - Phoneme R.
 {/r}{V}{C}. % rain
 {C}{V}{/r}{V}. % cherry
 {/r}{V}{C}{C}. % wrist
 1.1.15 - Phoneme S.
 {/s}{V}{C}. % sun
 {C}{C}{/s}.
 {V}{/s}. % ice
 \$\$\$ss. \$\$ss. \$\$\$ss\$. % dress, horse
 1.1.16 - Phoneme Zh.
 {*}{/zh}.
 {*}{/zh}{X}.
 {*}{/zh}{X}{X}. % treasure
 {*}{/zh}{X}{X}{X}.
 {*}{/zh}{X}{X}{X}.
 {*}{/zh}{X}{X}{X}{X}.
 {*}{/zh}{X}{X}{X}{X}{X}.
 {/zh}{*}.
 1.1.17 - Phoneme Sh.
 {/sh}{V}{C}. % chef
 {/sh}{V}{C}{C}. % shark
 {*}{/sh}{/θ}{/n}. % station
 1.1.18 - Phoneme T.
 {/t}{V}{C}. % tap
 {C}{V}{/t}{V}. % letter
 1.1.19 - Phoneme Th.
 {/th}{V}{C}. %
 {/th}{V}{C}{C}. % thumb
 {*}{/th}.
 1.1.20 - Phoneme Dh.
 {/dh}{X}{X}. %
 {/dh}{X}{X}{X}. %
 {/dh}{X}{X}{X}{X}.
 {*}{/dh}{X}. % feather
 {*}{/dh}{X}{X}.
 1.1.21 - Phoneme V.

	{/v}{*}.	% voice
	{*}{/v}.	% sleeve
1.1.22	- Phoneme W.	
	{/w}{X}{X}.	% wheel
	{/w}{X}{X}{X}.	% water
1.1.23	- Phoneme Y.	
	{/y}{X}{X}.	%
	{/y}{X}{X}{X}.	% yawn
	{/y}{X}{X}{X}{X}.	
1.1.24	- Phoneme Z.	
	{C}{V}{/z}.	% cheese
	{C}{C}{V}{/z}.	% sneeze
	{*}{/z}{V}.	% laser
1.2	- Vowel Phonemes.	
1.2.1	- Phoneme A.	
	{/a}{C}{X}.	% ant
	{C}{/a}{C}.	
1.2.2	- Phoneme Ei.	
	{C}{/ei}{C}.	% tape
	{C}{/ei}{C}{V}.	% baby
	{C}{C}{/ei}{C}.	% snail
	{C}{C}{/ei}.	% tray
1.2.3	- Phoneme Air.	
	{X}{/e@}{*}.	
	{*}{/e@}.	% hair
	{*}{/e@}{/r}.	% square
1.2.4	- Phoneme Aa.	
	{C}{/aa}{C}.	% car
	{C}{V}{C}{/aa}{*}.	% banana
1.2.5	- Phoneme E.	
	{C}{/e}{C}.	% bed
	{C}{C}{/e}{*}.	% bread
1.2.6	- Phoneme Ii.	
	{C}{/ii}{C}.	% beach
	{C}{C}{/ii}{*}.	% tree
	{C}{/ii}.	% key

{C}{V}{C}{/ii}. % pony
 1.2.7 - Phoneme Ea.
 {/i@}{*}. % ear
 {C}{/i@}{*}. % deer
 {X}{X}{/i@}{*}.
 1.2.8 - Phoneme Er.
 {*}{/@}. % teacher, collar, doctor, zebra
 {C}{V}{/@}{C}. % lion, circus
 {C}{V}{C}{/@}{C}. % garden, fossil, circus
 1.2.9 - Phoneme Ir.
 {C}{/@@}. % fur
 {C}{/@@}{C}. % fern, shirt, worm
 1.2.10 - Phoneme I.
 {C}{/i}{C}. % tin
 {*}{/i}{/t}. % rocket
 1.2.11 - Phoneme Ai.
 {C}{/ai}{C}{V}. % tiger
 {C}{/ai}{C}. % kite, light
 {C}{C}{/ai}. % fly
 1.2.12 - Phoneme O.
 {C}{C}{/o}{C}. % frog
 {C}{/w}{/o}{*}. % swan
 {/w}{/o}{*}.
 1.2.13 - Phoneme Ou.
 {C}{/ou}{C}. % nose, boat, note
 {C}{C}{/ou}{*}. % snow
 1.2.14 - Phoneme Oi.
 {C}{/oi}{*}. % coin
 {*}{/oi}. % toy
 1.2.15 - Phoneme Oo.
 {C}{/u}{*}. % book, bull
 1.2.16 - Phoneme Uu.
 {C}{/uu}{*}. % moon
 {C}{C}{/uu}. % screw, glue
 1.2.17 - Phoneme Oor.
 {X}{/u@}{*}. % moor
 {X}{X}{/u@}{*}.

{X}{X}{X}{/u@}{*}.

1.2.18 - Phoneme Aw.

{C}{/oo}. % saw

{C}{/oo}{C}. % fork, ball, sauce door

1.2.19 - Phoneme Ow.

{X}{/au}. % cow

{X}{/au}{X}. % house

{X}{/au}{X}{X}.

1.2.20 - Phoneme Uh.

{C}{C}{/uh}{C}. % glove

{C}{/uh}{C}. % bus

Appendix E

Vocabularies

E.1 Food Vocabulary

The thematic vocabulary for *food* as selected by the system, consisting of 1179 words, follows.

- abominable, detestable, abstemious, temperate, adept, alimentary, allergic, allergy, allergies, almond, almonds, ambrosia, aniseed, anorexia, aphrodisiac, appetite, appetites, appetizing, apple, apples, aromatic, fragrant, array, arrays, arrayed, asparagus, assam, astringent, astringents, aubergine, aubergines, authentic, avocado, avocados
- bag, bags, bait, baits, bake, bakes, baking, baked, bamboo, bamboos, banana, banquet, banquets, barbecue, barbecues, bard, bards, base, bases, bottom, top, basic, fundamental, bean, beans, beef, beefs, beefing, moan, beefsteak, beet, beetroot, beside, bib, bind, binds, binding, bound, biscuit, biscuits, bitter, bitterest, blanch, blanches, blanching, blanched, bland, blander, bloody, blossom, blossoms, board, boards, boil, boils, boiling, boiled, bonus, bonuses, gratuity, bony, booth, booths, cubicle, bountiful, plentiful, bowl, bowls, bran, breakfast, brit, brits, broccoli, broth, buffet, bulb, bulbs, bun, buns, burger, burgers, burn, burns, burning, burnt, butcher, butchers, butter, buttery
- cabbage, cabbages, cafeteria, cafeterias, canteen, cake, cakes, caked, encrusted, caller, callers, calorie, calories, candyfloss, capsicum, capsicums, carafe, caramel, carcass, carcasses, carob, carobs, carp, carps, carrion, casein, cater, caters, catering, catered, caviar, celery, celestial, centrepiece, cereal, cereals, champ, champs, char, chars, charring, charred, charlotte, cheese, cheeses, cheesy, cheesier, cheesiest, chestnut, chestnuts, chew, chews, chewing, chewed, chewy, chewier, chewiest, chicory, chilli, chillies, chocolate, chocolates, cinnamon, citrus, clam, clams, clean, cleaner, cleanest, dirty, coat, coats, cocoa, coconut, coconuts, cod, collins, communion, condiment, condiments, cone, cones, confection, confections, connoisseur, connoisseurs, consume, consumes, consuming, consumed, convert, converts, converting, converted, cook, cooks, cooking, cooked, cookery, cookie, cookies, coriander, cow, cows, cracker, crackers, cream, creamy, creamier, creamiest, creole, creoles, patois, crisp, crisper, soggy, crispy, crispier, crispiest, crop, crops, crunchy, cube, cubes, cucumber, cucumbers, cuisine, cuisines, culinary, cultivate,

cultivates, cultivating, cultivated, cup, cups, cupboard, cupboards, curd, curds, curry, curries, custard, cyprinid

- dainty, daintiest, dairy, defrost, defrosts, defrosting, defrosted, freeze, delectable, delightful, delicacy, delicious, tasty, dessert, desserts, dessertspoon, devour, devours, devouring, devoured, dice, diet, diets, dietary, digest, digests, digesting, digested, digestible, digestive, dine, dines, dining, dined, dinner, dinners, dionysus, dip, dips, dipping, dipped, dunk, dish, dishes, dishwasher, dishwashers, distinct, distribute, distributes, distributing, distributed, dive, dives, diving, dived, double, drain, drains, draining, drained, dress, dresses, dressing, dressings, dried, dehydrated, drinkable, dripping, drizzle
- eater, eaters, eating, eats, edible, inedible, eliminate, eliminates, eliminating, eliminated, enrich, enriches, enriching, enriched, excess, excesses, exclusive, execrable, deplorable, extra, additional
- faint, fainter, faintest, fare, fares, fashionable, fast, faster, fastest, quick, slow, fat, fatter, fattest, overweight, thin, fatten, fattens, fattening, fattened, fatty, fatter, feast, feasts, feed, feeds, feeding, fed, fennel, fibre, fibres, strand, fig, figs, fillet, fillets, filling, fillings, filter, filters, filtering, filtered, fishy, fitting, fittings, flavour, flavours, flavoured, flavouring, flavourings, fleshy, float, floats, floating, floated, florentine, flour, flours, floured, fluffy, fluffier, fluffiest, soft, foil, fond, fonder, forage, forages, foraging, fork, forks, fresh, new, frosted, frozen, frugal, thrifty, fruity, fruitier, fruitiest, frumenty, fry, fries, frying, fried
- garnish, garnishes, ginger, glaze, glazes, goat, goats, gooey, gooier, gooiest, gorge, gorges, ravine, gourmet, gastronome, grab, grabs, grabbing, grabbed, grain, grains, gravy, gravies, grease, oil, greens, gridiron, grill, grills, grinder, grinders, grub, grubs, gruel, gum, gums, gut, guts, gutting, gutted
- habit, habits, hack, hacks, hacking, hacked, haddock, hamburger, hamburgers, hazel, hazels, health, healthful, healthy, hearty, heartier, heartiest, jovial, helping, helpings, hen, hens, herb, herbs, herbage, hole, holes, honeyed, hospitality, host, hosts, hostess, housekeeping, hunger, hungry, full, hunt, hunts, hunting, hunted, search, ice, iced, icing, frosting
- indian, indigestible, ingest, ingests, ingested, ingredient, ingredients, inside, insides, outside, insipid, instead, intense, extreme
- jellied, jelly, jellies, jersey, juices, juicy, juicier, juiciest, junket, junkets
- kettle, kettles, kidney, kidneys, kitchen, kitchens, kiwi, kiwis, knead, kneading, kneaded, knife, knives, kosher
- laxative, laxatives, leavings, leftover, leftovers, lemon, lemons, lemonade, lift, lifts, lifting, lifted, liking, ling, lip, lips, liquorice, litchi, liver, livers, lotus, lunch, lunches, luncheon, luncheons
- mackerel, malnutrition, manna, maple, marinade, marmalade, marmalades, marrow, mash, mashing, mashed, mast, masts, mature, matures, maturing, matured, mayonnaise, meal, meals, meat, meats, mediterranean, mellow, mellowest, menu, menus, mess, microwave, microwaves, mild, milder, slight, great, milk, milky, miscellaneous, assorted, mix, mixes, mixing, mixed, mixture, combination, mousse,

mulberry, mulberries, mull, mulling, mulled, mullet, mullets, mushroom, mushrooms, mushy, mutton

- nosh, nourish, nourishes, nourishing, nourished, nut, nuts, nutmeg, nutritional, nutritive, nutritious, nutty, oat, oats, odour, odours, oesophagus, gullet
- oily, oilier, oiliest, okra, olive, olives, omelette, omnivorous, onion, onions, orange, oranges, organic, oven, ovens, overdone, underdone, overflowing, owe, owes, owing, owed, oyster, oysters
- palatable, unpalatable, palate, palates, pan, pans, saucepan, pancake, pancakes, crepe, pare, paring, pared, parmesan, parsley, parsnip, parsnips, partake, partakes, partaking, partook, partaken, pasta, pastas, paste, pastes, pea, peas, peachy, peanut, peanuts, pear, pears, peel, pepper, perishable, pheasant, pheasants, pick, picks, picking, picked, choose, pickled, pickles, picnic, pig, pigs, hog, pinch, pinches, pinching, pinched, pink, pinks, pinker, pint, pints, piquant, pizza, pizzas, plaice, plank, planks, plantain, please, plenty, lots, plum, plums, plunge, plunges, plunging, plunged, poach, poaches, poaching, poached, pop, pork, porridge, pot, pots, potato, potatoes, potpourri, pottage, prepare, prepares, preparing, prepared, preserve, preserves, preserving, preserved, proper, protein, proteins, pudding, puddings, pulp, pulses, pumpkin, pungent, powerful, delicate, puree
- quicker, quickest
- raid, raids, raiding, raided, raise, raises, raising, raised, rancid, range, ranges, rare, rarer, rarest, uncommon, common, ration, rations, raw, refresh, refreshes, refreshing, refreshed, refreshment, regimen, regime, regular, irregular, relish, relishing, relished, enjoy, remove, removes, removing, removed, render, renders, rendering, rendered, make, repast, replete, restaurant, restaurants, rhubarb, rib, ribs, rice, rices, rind, ripe, riser, risers, roast, roasting, roasted, rocket, rockets, spacecraft, roe, roes, roll, rolls, rolling, rolled, rot, rots, rotting, rotted, rubbery, rump, rye
- salad, salads, salmon, salt, salty, saltier, saltiest, sandwich, sandwiches, sandwiched, satiate, satiates, satiated, saturated, sauce, sauces, sausage, sausages, save, saves, saving, saved, savour, savours, savouring, savoured, savoury, scales, scallop, scallops, scatter, scatters, scattering, scattered, scone, scones, scorch, scorches, scorching, scorched, scramble, scrambles, scrambling, scrambled, clamber, scraps, seafood, seasoning, seasonings, section, sections, seed, seeds, semolina, separately, together, server, servers, serving, servings, sesame, setting, settings, shake, shaking, shook, shaken, shank, shanks, sherbet, sherry, cherries, shin, shins, shortening, sickly, simmer, simmers, simmering, simmered, simple, simpler, simplest, sirloin, skewer, skewers, skim, skims, skimming, skimmed, slice, slices, slim, slimmer, slimmest, slender, smack, smacks, smacking, smacked, smell, smells, smelt, smelts, smelting, smelted, smoky, smokier, smokiest, snack, snacks, soda, sodden, soaked, sole, only, soup, soups, spatula, spatulas, speciality, specialities, field, spice, spices, spiced, spicy, spicier, spiciest, spinach, spoil, spoils, spoiling, spoilt, spoon, spoons, spread, spreads, spreading, sprout, sprouts, sprouting, sprouted, squash, squashes, squashing, squashed, stale, staple, starve, starves,

- starving, starved, steam, steamer, steamers, stew, stews, stick, sticks, twig, stomach, stomachs, tummy, stringy, stringier, stringiest, strip, strips, stuffing, subsistence, substantial, significant, insubstantial, substitute, substitutes, substituting, substituted, succulent, suck, sucks, sucking, sucked, suet, sugar, sugary, suggest, suggests, suggesting, suggested, sup, supping, supped, supper, suppers, sustenance, swede, sweet, sweeter, sweetest, sweetish
- tab, tabs, takeaway, takeout, tallow, tart, tarts, taste, tasteless, taster, tastier, tastiest, tea, teas, teaspoon, teaspoons, thaw, thaws, thawing, thawed, thirsty, thirstier, thirstiest, tin, tomato, tomatoes, tongue, tongues, topping, touch, touches, touching, touched, trash, trifle, trough, troughs, tuber, tubers, tuna, tunas, turbot, turnip, turnips
 - uncooked, unctuous, underfed, underneath, beneath, undressed, distasteful, unripe, unsound, sound, untouched, unused, unwholesome, unhealthy, utensil, utensils
 - variation, variations, veal, vegetable, vegetables, vegetarian, vegetarianism, venison, vinegar
 - wait, waits, waiting, waited, waiter, waiters, waitress, waitresses, watery, feeble, weak, weaker, weakest, strong, wedge, wedges, wedging, wedged, wheat, wheats, whey, whip, whips, whipping, whipped, beat, whiting, wholewheat, wild, wilder, wildest, stormy, wrapper, wrappers
 - yoghurt, yogurt

E.2 Sea Vocabulary

The thematic vocabulary for *sea* as selected by the system, consisting of 1435 words, follows.

- abandon, abandons, abandoning, abandoned, aboard, abroad, overseas, address, addresses, admiral, admirals, admiralty, adriatic, adrift, afloat, aft, fore, aground, airlift, alabama, alluvium, amazon, amazons, amidships, ammonite, amphibian, amphibians, amphibious, anchor, anchors, anchorage, angle, angles, angler, anglers, angling, aquarium, aquaria, aquatic, archipelago, armada, armour, arrange, arranges, arranging, arranged, ascension, ashore, assemble, assembles, assembling, assembled, gather, astern, atlantic, atmosphere, attract, attracts, attracting, attracted, attrition, auk, auks, average, averages, mean
- backwater, backwaters, bait, baits, ballast, bank, banks, bar, bars, barge, barges, barrel, barrels, barrier, barriers, obstacle, base, bases, bottom, top, bash, bashes, thrash, basin, basins, bass, basses, bathe, bathes, bathing, bathed, batter, batters, battering, battered, bay, bays, beach, beaches, seashore, beacon, beacons, beam, beams, beaming, beamed, beat, beats, beating, beaten, becalmed, stranded, belt, belts, bering, berth, bilge, bilges, bite, bites, biting, bit, bitten, bivalve, blast, blasts, bleak, bleaker, gloomy, bright, blennioid, blockade, blockades, blow, blows, blowing, blew, blown, bluff, bluffs, board, boards, boat, boats, boatman, boatmen, bone, bones, bonny, lovely, boom, booms, slump, border, borders, frontier, bore, bores, boring, bored, borneo, botanical, bottoms, boulder, boulders, bound, bounds, bounding, bounded, leap, bow, bows, bowing, bowed, bowl, bowls, brace, braces, bracing, braced, breach, breaches, breached, violate, breakwater, bridge, bridges, brine, brit, brits, broadside, buccaneer, buccaneers, bucket, buckets, bump, bumps, bumping, bumped, bunker, bunkers, buoy, buoys, burden, burdens
- cabin, cabins, cable, calm, calmer, canal, canals, canister, canisters, canoe, canoes, cape, capes, captain, captains, carangid, cargo, cargoes, consignment, caribbean, carrier, carriers, catastrophic, disastrous, catch, catches, catching, caught, cave, caves, cetacean, cetaceans, channel, channels, station, chart, charts, diagram, charter, charters, check, checks, checking, checked, choppy, rough, chuck, chucks, chucked, throw, circulation, circulations, claw, claws, clean, cleaner, cleanest, dirty, clearance, coarse, coarser, coarsest, fine, coast, coasts, coastal, coaster, coastguard, coastguards, coastline, coastlines, cockle, cockles, cod, commando, commandoes, commission, commissions, commissioning, commissioned, competing, conflicting, conch, conger, container, containers, receptacle, contaminate, contaminates, contaminating, contaminated, convey, conveys, conveying, conveyed, communicate, convoy, convoys, coral, corals, cork, cormorant, cormorants, counter, counters, cove, coves, crab, crabs, crack, cracks, cracking, cracked, craft, crash, crashes, accident, crawl, crawls, crawling, crawled, creek, creeks, crew, crews, cross, crosses, crossing, crossed, crossings, cruise, cruises, crust, crusts, crustacean, crustaceans, cuba, curlew, curlews, curly, curlier, curliest, straight, current, currents, cutter, cutters, cuttlefish, cycle, cycles, cycling, cycled, cyprinodont
- deck, decks, deepen, deepens, deepening, deepened, deposition, depositions, depth, depths, derrick, desert, deserts, detachment, dip, dips, dipping, dipped, dunk,

- discharge, discharges, discharging, discharged, displacement, distance, distances, ditch, ditches, dive, dives, diving, dived, diver, divers, diverge, diverges, diverging, diverged, dock, docks, dolphin, dolphins, drag, drags, dragging, dragged, dredge, dredging, dredged, drift, drifts, drifting, drifted, drip, drips, dripping, dripped, drop, drops, dropping, dropped, drown, drowns, drowning, drowned, drum, drums, dry, dryer, driest, dryers, dump, dumps, dumping, dumped, dune, dunes, dust, dyke, dykes
- ebb, ebbs, ebbing, ebbled, embankment, embankments, engineer, engineers, environment, environments, erode, erodes, eroding, eroded, erosion, estuary, estuaries, everywhere, excess, excesses, explore, explores, exploring, explored, exposed
 - falls, fare, fares, fathom, fathoms, feather, feathers, fender, fenders, guard, ferry, ferries, fiji, filter, filters, filtering, filtered, fin, fins, finish, finishes, finishing, finished, firm, firms, company, fisher, fishing, flag, flags, flagship, flash, flashes, flask, flasks, flatfish, fleet, fleets, flipper, flippers, float, floats, floating, floated, flood, floods, floor, floors, flotsam, flow, flows, flowing, flowed, fluke, flukes, flushing, flux, instability, fly, flies, foam, froth, fog, fogs, follow, follows, following, followed, forget, forgets, forgetting, forgot, forgotten, remember, fork, forks, foul, foulest, disgusting, founder, founders, freight, fresh, new, freshwater, funnel, funnels
 - gadoid, gaff, galley, galleys, gig, gigs, show, gill, gills, glint, glints, glinting, glinted, glisten, globe, planet, gourd, gourds, grain, grains, gross, grossest, grunt, grunts, grunting, grunted, guards, guarding, guarded, guide, guides, gulf, gulfs, gull, gulls, gush, gushes, gushing, gushed, guts, haddock
 - haiti, handle, handles, harbour, harbours, harpoon, harpoons, hatch, hatches, hatching, hatched, haul, hauls, hauling, hauled, haven, havens, refuge, hawaii, hazard, hazards, headland, headlands, height, heights, helm, herring, herrings, hind, rear, hire, hires, hiring, hired, homeward, hook, hooks, hop, hops, hopping, hopped, horn, horns, hovercraft, hudson, hull, hulls, hydraulic
 - inch, inches, incoming, outgoing, infest, infests, infesting, infested, inland, inlet, inlets, inshore, offshore, intercept, intercepts, intercepting, intercepted, stop, interval, intervals, gap, island, islands, isle, isles
 - jack, jacks, jamaica, japan, jellyfish, jet, jets, jetty, jump, jumps, jumping, jumped, junk, rubbish, jut, juts, jutting, juttled
 - keel, knot, knots
 - lagoon, lagoons, lake, lakes, landlocked, lash, lashes, eyelash, launch, launches, launching, launched, laver, lead, leads, leading, led, lee, liberty, liberties, freedom, lifeboat, lifeboats, lift, lifts, lifting, lifted, lighter, lighters, lighthouse, lighthouses, limestone, limestones, liner, ling, links, load, loads, loading, loaded, lobster, lobsters, loch, lochs, log, logs, lookout, lough, lugworm
 - mackerel, madagascar, mainland, marine, marines, maritime, mast, masts, master, masters, mate, mates, friend, mauritius, measure, measures, measuring, measured, assess, melt, melts, melting, melted, migrate, migrates, migrating, migrated, mine, mineral, minerals, mississippi, mixture, combination, mobile, model, models, monitor, monitors, monitoring, monitored, mooring, moorings, motor, motors, engine, mullet, mullets, murray

- naples, nautical, naval, navigable, navigate, navigates, navigating, navigated, navigation, navigational, navigator, navy, navies, neat, neater, needle, needles, net, netting, newfoundland
- oar, oars, oblique, indirect, ocean, oceanic, ooze, oozes, oozing, oozed, orkneys, outlet, outlets, outward, return, overboard, oyster, oysters
- pacific, peaceable, belligerent, packet, packets, pack, pad, pads, paddle, paddles, panama, passport, passports, pearl, pearls, pebbly, pelican, pelicans, penguin, penguins, peninsula, peninsulas, percoid, petroleum, pier, piers, pile, piles, pilot, pilots, pipe, pipes, piracy, pirate, pirates, pisces, pitch, pitches, pivot, pivots, plaice, plane, planes, plenty, lots, plot, plots, conspiracy, plunge, plunges, plunging, plunged, ply, plies, plying, plied, pod, pods, pollute, pollutes, polluting, polluted, polo, polychaete, pond, ponds, pool, pools, post, pound, pounds, pour, pours, pouring, poured, prepare, prepares, preparing, prepared, preserve, preserves, preserving, preserved, press, presses, pressing, pressed, promontory, promontories, propeller, propellers, prow, pull, pulls, pulling, pulled, pump, pumps
- quartermaster
- radiator, radiators, raf, raft, rafts, range, ranges, rapid, fast, raw, ray, rays, reach, reaches, reaching, reached, reef, reefs, reopen, reopens, reopening, reopened, report, reports, reporting, reported, reserve, reserves, reserving, reserved, rich, richer, richest, poor, ride, rides, riding, rode, ridden, ring, rings, ringing, rang, rung, phone, ripple, ripples, rise, rises, rising, rose, risen, rivers, roar, roars, roaring, roared, roe, roes, roller, rollers, rougher, roughest, smooth, route, routes, rover, row, rows, line, rudder, ruff, ruffle, ruffles, ruffling, ruffled, rumple
- safe, safer, safest, dangerous, sail, sails, sailing, sailings, sailor, sailors, saline, salmon, salt, salty, saltier, saltiest, salvage, sand, sands, sanded, sandbank, sandbanks, sandpiper, save, saves, saving, saved, scale, scales, scaling, scaled, climb, scallop, scallops, scan, scans, scanning, scanned, schoolmaster, schoolmasters, sciaenid, scombroid, screw, screws, seabird, seabirds, seafood, seal, seals, sealing, sealed, seaport, seaports, seashores, seaside, season, seasons, experienced, seaweed, seaweeds, sediment, sediments, sergeant, serve, serves, serving, served, shake, shaking, shook, shaken, shallows, shark, sharks, shell, shells, shellfish, sheltered, shipment, shipments, shipping, shipwreck, shipwrecked, shoal, shoals, shore, shores, shoreline, shrimp, shrimps, sighting, sightings, sink, sinks, siren, sirens, site, sites, skate, skates, ski, skis, skim, skims, skimming, skimmed, skipper, skippers, sledge, sledges, sled, sleeper, sleepers, slush, smelt, smelts, smelting, smelted, smoother, smoothest, lumpy, snorkel, snorkels, snorkelling, soak, soaks, soaking, soaked, soluble, insoluble, sonar, spade, spades, sparid, spat, spats, quarrel, spin, spins, spinning, spun, spinner, spinners, spit, saliva, splash, splashes, splashing, splashed, split, splits, splitting, sponge, sponges, spot, spots, spout, spouts, spouting, spurt, spring, springs, sprinkle, sprinkles, sprinkling, sprinkled, squall, squalls, storm, stack, stacks, stacking, stacked, stage, stages, phase, starboard, port, start, starts, starting, started, begin, stations, stationed, position, staunch, steadfast, steam, steamer, steamers, steer, steers, steering, steered, stem, stems, stemming, stemmed, steward, stewards, stock, stocks, storms, stormy, stormier, stormiest, straighter, straightest, stretch,

stretches, stretching, stretched, extend, stretcher, strike, strikes, striker, strikers, strip, strips, stroke, strokes, stroking, stroked, caress, stuff, sturgeon, submarine, submarines, sucker, suckers, suetz, sunken, surf, surfing, surge, surges, swamp, swamps, swan, swans, sway, sways, swaying, swayed, sweep, sweeps, sweeping, swept, swell, swells, swelling, swelled, swollen, increase, swim, swims, swimming, swam, swum, swing, swings, swinging, swung, swirl, swirls, swirling, swirled

- tack, tacks, tackle, tackles, tackling, tackled, taiwan, tan, suntan, tanker, tankers, teleost, tender, tenderer, tenderest, terrapin, terrapins, terrestrial, thrashes, thrashing, thrashed, hammer, tidal, tide, tides, tinker, tinkers, tinkering, tinkered, ton, tons, tonnage, torpedo, torpedoes, tortoise, tortoises, tortoiseshell, toss, tosses, tossing, tossed, trace, traces, tracing, traced, trader, traders, traffic, trail, trails, track, transatlantic, transfer, transfers, transferring, transferred, transport, transportation, travel, travels, travelling, travelled, trawl, trawls, trawling, trawled, trawler, trawlers, trinidad, trough, troughs, trout, trouts, tug, tugs, tugging, tugged, yank, turbulent, turtle, turtles
- undercurrent, undercurrents, undertow, undertows, underwater, uniform, uniforms, uninhabited, deserted, unload, unloads, unloading, unloaded
- viking, vikings, voyage, voyages, journey
- waft, wafts, wafting, wafted, wait, waits, waiting, waited, wake, wakes, waking, woke, woken, warm, warmer, warmest, cool, wash, washes, washing, washed, watch, watches, watching, watched, observe, waterfront, waterline, waterproof, waters, waterway, waterways, wave, waves, waving, waved, weir, wet, wets, wetting, wetted, whale, whalebone, wharf, whip, whips, whipping, whipped, whirlpool, whirlpools, whirlwind, whirlwinds, whitefish, whiting, wick, wing, winged, wrap, wreck, wrecks, wrecking, wrecked
- yacht

E.3 Agriculture vocabulary

The thematic vocabulary for *agriculture* as selected by the system, consisting of 1516 words, follows.

- absorb, absorbs, absorbing, absorbed, acanthus, acorn, acorns, agrarian, agricultural, agriculture, airlift, alfalfa, lucerne, alluvium, almond, almonds, amoeba, amoebas, amphibian, amphibians, amphibious, annual, yearly, anthrax, apple, apples, apricot, apricots, april, aquatic, arable, archangel, archangels, ash, ashes, aubergine, aubergines, auction, auctions, august, imposing, autonomy, independence, autumn, avocado, avocados, axe, axes, axis
- bag, bags, bail, balm, balsam, bamboo, bamboos, banana, bare, bark, barks, barking, barked, barley, barn, barns, barrel, barrels, barren, fertile, barter, bartering, bartered, trade, base, bases, bottom, top, basil, basin, basins, bath, baths, batter, batters, battering, battered, bean, beans, bear, bears, bearing, bore, borne, carry, beat, beats, beating, beaten, beech, beet, beetroot, belt, belts, ben, bent, berry, berries, biennial, biological, birch, birches, bite, bites, biting, bit, bitten, blackberry, blackberries, blade, blades, block, blocks, bloom, blooms, flower, blossom, blossoms, blow, blows, blowing, blew, blown, bolt, bolts, bond, bonds, border, borders, frontier, borderland, bores, boring, bored, botany, bottoms, boulevard, boulevards, avenue, bound, bounds, bounding, bounded, leap, bowl, bowls, box, boxes, bran, branch, branches, brassica, brassicas, breed, breeds, strain, breeding, briar, briars, broccoli, broker, brokers, broth, bruise, bruises, brush, brushes, bucket, buckets, buckwheat, bud, buds, budget, budgets, buffalo, bison, bulb, bulbs, bulk, bull, bulls, bur, burn, burns, burning, burnt, burr, burrs, bush, bushes, shrub, bushel, bushels, bushy, bushier, bushiest, buy, buys, buying, bought, buyer, buyers, seller
- cabbage, cabbages, cacao, calf, calves, calibrate, calibrating, calibrated, calm, calmer, canal, canals, cane, canes, capacity, capacities, caper, capers, capsicum, capsicums, capsule, capsules, caraway, card, cards, carnation, carnations, carnivorous, carob, carobs, carrier, carriers, carrot, carrots, cart, carts, wagon, carton, cartons, cassava, manioc, cassia, catkin, catkins, cauliflower, cauliflowers, cedar, cedars, celery, cement, cereal, cereals, chaff, channel, channels, cheap, cheaper, cheapest, chemist, chemists, pharmacy, cherry, cherries, chestnut, chestnuts, chicory, chill, chills, chilling, chilled, heat, chrysanthemum, chrysanthemums, citrus, clone, clones, clove, cloves, clover, clovers, clump, clumps, cluster, cocoa, coffee, combine, combines, combining, combined, commerce, commercial, commission, commissions, commissioning, commissioned, community, competition, competitive, composite, compost, concession, concessions, cone, cones, conifer, conifers, conserve, conserves, conserving, conserved, save, consumption, container, containers, receptacle, core, cores, coriander, corn, corner, corners, corridor, corridors, cost, costs, costing, costed, cotton, cottons, county, counties, coupon, coupons, voucher, cow, cows, cream, credit, creeper, creepers, crop, crops, cross, crosses, crossing, crossed, crossbreed, crumble, crumbles, crumbling, crumbled, cube, cubes, cucumber, cucumbers, cultivate, cultivates, cultivating, cultivated, refined, cultivation, cultivator, culture, customer, customers, customs, cutting, cuttings, clipping

- daffodil, daffodils, dahlia, dairy, daisy, daisies, dam, dams, dandelion, date, dates, december, deciduous, evergreen, deliver, delivers, delivering, delivered, deposit, deposits, desert, deserts, develop, develops, developing, developed, developer, developers, development, dig, digs, digging, dug, dilute, dilutes, diluting, diluted, dip, dips, dipping, dipped, dunk, discount, discounts, district, districts, division, split, dock, docks, domain, domains, dominion, double, drag, drags, dragging, dragged, drain, drains, draining, drained, draw, draws, drawing, drew, drawn, sketch, drill, drills, drop, drops, dropping, dropped, drover, drovers, drum, drums, dry, dryer, driest, dryers, dumfries, dump, dumps, dumping, dumped, duty
- economy, economies, embargo, embargoes, ban, embryo, embryos, emmer, environment, environments, erosion, evergreens, evolution, excavate, excavates, excavating, excavated, exchange, exchanges, exchanging, exchanged, swap, excise, excises, executive, executives, export, exports, exporting, exported, import
- factory, factories, fair, fairer, fairest, fallow, farm, farms, farming, farmyard, farmyards, fat, fatter, fattest, overweight, thin, fathom, fathoms, feather, feathers, february, feed, feeds, feeding, fed, female, male, fence, fences, fennel, rich, fertilize, fertilizes, fertilizing, fertilized, inseminate, fertilizer, fertilizers, fig, figs, finance, finances, financing, financed, firm, firms, company, flail, flails, float, floats, floating, floated, flood, floods, flora, flour, flours, floured, flowering, blossoming, flush, flushes, flushing, flushed, fly, flies, foam, froth, fog, fogs, fold, folds, folding, folded, forage, forages, foraging, forest, forests, forestry, fork, forks, fountain, fountains, freeze, freezes, freezing, froze, frozen, freight, fresh, new, fret, frets, fretting, fretted, worry, frog, frogs, frontiers, fruitful, successful, furrow, furrows
- galloway, garden, gardens, gauge, gauges, gauged, generation, generations, gentle, gentler, gentlest, geography, geranium, geraniums, ginger, ginseng, glut, surplus, gluten, goat, goats, gooseberry, gooseberries, gouge, gouges, gouging, gouged, gourd, gourds, governor, governors, graft, grafts, grain, grains, grainy, gram, grams, granary, granaries, grape, grapes, grapefruit, grass, grasses, grassland, grasslands, grassy, grassier, graze, grazes, grazing, grazed, greenhouse, greenhouses, grove, groves, grow, grows, growing, grew, grown, growth, gum, gums, gymnosperm
- hack, hacks, hacking, hacked, harrow, harrows, harvest, harvester, haul, hauls, hauling, hauled, haulage, heath, heaths, moor, hemlock, herb, herbs, herbage, herdsman, herdsmen, hereford, hide, hides, hiding, hid, hidden, conceal, hoe, holding, holdings, investment, homestead, hopper, hoppers, horn, horns, hose, hoses, hosepipe, hull, hulls, humus, hyacinth, hyacinths, hybrid, hybrids
- ice, immersed, implement, implements, implementing, implemented, imports, importing, imported, impregnate, impregnates, impregnating, impregnated, infertile, intensive, inundate, inundated, swamp, invest, invests, investing, invested, invoice, invoices, bill, irrigate, irrigated
- january, jersey, juice, juices, july, june, jungle, jungles, juniper, junipers, kernel
- kernels, knot, knots
- lake, lakes, lamp, lamps, lands, landscape, landscapes, latex, laurel, laurels, layer, layers, leafy, leasehold, leguminous, lemon, lemons, lichen, lichens, lift, lifts, lift-

ing, lifted, lily, lilies, lime, limes, lincoln, linden, litchi, livestock, locust, locusts, log, logs, logging, loss, losses, lotus, lumberjack, lumberjacks, lush, verdant

- magnolia, magnolias, maize, male, mallow, malt, mandarin, mandarins, mandate, mandates, manure, maple, march, marches, marches, marching, marched, marginal, marigold, market, markets, marketplace, mash, mashing, mashed, mate, mates, friend, measure, measures, measuring, measured, assess, measurement, measurements, melon, melons, merchant, merchants, meter, meters, middleman, middlemen, milk, mill, mills, miller, millet, millstone, millstones, ministry, ministries, mint, mistletoe, mix, mixes, mixing, mixed, moor, moors, moss, mosses, motor, motors, engine, muck, mud, mulberry, mulberries, mustard
- native, natives, needle, needles, negotiate, negotiates, negotiating, negotiated, nettle, nettles, november, nut, nuts, nutmeg
- oat, oats, october, offset, offsets, offsetting, olive, olives, onion, onions, orange, oranges, orchard, orchards, organ, organs, organic, outlet, outlets, ovary, ovaries, ovule
- package, packages, paddy, paddies, palm, palms, pan, pans, saucepan, papaya, papayas, pawpaw, papyrus, parsley, parsnip, parsnips, pastoral, pasture, pastures, patch, patches, pea, peas, peach, peaches, peanut, peanuts, pear, pears, peddle, peddles, peddling, peddled, pepper, perennial, pest, pests, photosynthesis, pick, picks, picking, picked, choose, pinch, pinches, pinching, pinched, pine, pines, pineapple, pineapples, pink, pinks, pinker, pioneer, pioneers, pit, pits, stone, pith, plain, plantain, plantation, plantations, plot, plots, conspiracy, plough, ploughs, pluck, plucks, plucking, plucked, plug, plugs, plum, plums, plumbing, ply, plies, plying, plied, pod, pods, pole, pollen, pollution, contamination, pomegranate, pomegranates, poppy, poppies, pore, pores, pot, pots, potato, potatoes, pound, pounds, prairie, prairies, preserve, preserves, preserving, preserved, price, prices, prime, producer, producers, productive, prolific, propagate, propagates, propagating, propagated, disseminate, province, provinces, area, prune, prunes, pulp, pulse, pulses, pump, pumps, pumpkin, purchase, purchases, purchasing, purchased
- race, races, radical, fundamental, rain, rains, raining, rained, raise, raises, raising, raised, rake, rakes, ranch, ranches, rancher, ranchers, range, ranges, raspberry, raspberries, rate, rating, rated, merit, reap, reaps, reaping, reaped, receptacles, reclaim, reclaiming, reclaimed, regulation, regulations, rule, release, releases, releasing, released, repeat, repeats, repeating, repeated, reproduce, reproduces, reproducing, reproduced, copy, reserve, reserves, reserving, reserved, retail, revenue, revenues, rice, rices, richer, richest, poor, ride, rides, riding, rode, ridden, rind, ripe, rise, rises, rising, rose, risen, root, roots, rosaceous, rosemary, rotation, rough, rougher, roughest, smooth, rubber, rubiaceous, rugged, rules, runner, runners, running, rust, rye
- sage, sages, sale, sales, salt, sap, saps, sapping, sapped, diminish, scale, scales, scaling, scaled, climb, scoop, scoops, scooping, scooped, scour, scours, scouring, scoured, search, scrub, scrubs, scrubbing, scrubbed, scrubby, scythe, scythes, season, seasons, experienced, seasonal, sedge, sedges, seed, seeds, seizure, seizures, sell, sells, selling, sold, semolina, september, serve, serves, serving, served, shallow, shallower, shallowest, deep, shave, shaves, shaving, shaved, sheaf, sheaves,

shear, shears, shearing, shorn, shed, sheds, sheep, sheepdog, sheepdogs, shelf, shelves, shift, shifts, shifting, shifted, shoot, shoots, shooting, shot, shop, shops, store, shovel, shovels, shower, showers, shrubs, sickle, sickles, silage, slick, slicker, slip, slips, slipping, slipped, slide, smudge, smudges, smut, snag, snags, hitch, soap, soaps, sod, sods, soil, soils, sorghum, sounding, sow, sows, sowing, sowed, sown, plant, species, spinach, sponge, sponges, spore, spores, spray, sprays, spread, spreads, spreading, spring, springs, springboard, sprout, sprouts, sprouting, sprouted, spruce, spruces, square, squares, squash, squashes, squashing, squashed, stack, stacks, stacking, stacked, pack, stalk, stalks, stem, stamen, stamens, stamp, stamps, standard, standards, starch, stems, stemming, stemmed, sterile, stick, sticks, twig, sting, stings, stinging, stung, stock, stocks, stockbroker, stockbrokers, straw, strawberry, strawberries, stretch, stretches, stretching, stretched, extend, strike, strikes, string, strings, stringing, strung, strip, strips, stub, stubs, stubble, stud, studs, stump, stumps, submerged, subsoil, sucker, suckers, summer, summers, sunflower, sunflowers, supply, supplies, supplying, supplied, swamps, sweep, sweeps, sweeping, swept, swill, swills, swilling, swilled, sycamore, sycamores, syrup, tangerine

- tariff, tariffs, tax, taxes, tea, teas, teasel, temperature, temperatures, tender, tenderer, tenderest, terrestrial, territory, territories, thicket, thickets, thistle, thistles, thread, threads, tick, ticks, tobacco, tomato, tomatoes, tool, tools, topography, tractor, tractors, trader, traders, trailer, train, trains, transfer, transfers, transferring, transferred, transport, transportation, trash, tread, treads, treading, trod, trodden, trolley, trolleys, trough, troughs, truck, trucks, lorry, trunk, trunks, trust, trusts, trusting, trusted, tuber, tubers, tuft, tufts, tulip, tulips, turnip, turnips, turpentine, twigs
- undercut, undergrowth, unfruitful, unload, unloads, unloading, unloaded
- van, vans, vanilla, vegetable, vegetables, vegetated, vegetation, vegetative, verge, brink, vine, vines, grapevine, vineyard, vineyards, vintage, vintages, violet, violets
- wagon, wagons, walnut, walnuts, warehouse, warehouses, wares, wash, washes, washing, washed, waste, wastes, wasting, wasted, waterfront, waterproof, waterside, watery, feeble, weak, weaker, weakest, strong, weather, weed, weeds, weedy, weight, weights, wheat, wheats, wholesale, wholewheat, wick, wild, wilder, wildest, stormy, wilderness, wildernesses, willow, willows, winter, winters, wither, withers, withering, withered, woodland, woodlands, wool
- xylem
- yard, yards, yew, yews, yield

E.4 Freire vocabulary

The thematic vocabulary for *Freire's Generative Themes for the town of Ceilandia* as selected by the system, consisting of 1633 words, follows.

- abandonment, absentee, absentees, accept, accepts, accepting, accepted, active, sedentary, adopt, adopts, adopting, adopted, advance, advances, advancing, advanced, development, advert, adverts, advertise, advertises, advertising, advertised, affirm, affirms, affirming, affirmed, assert, afraid, frightened, aggregate, agree, agrees, agreeing, agreed, concur, disagree, ahead, behind, airwaves, alignment, alignments, affiliation, allocation, allocations, allow, allows, allowing, allowed, forbid, altar, altars, amen, amusement, anglican, appliance, appliances, application, applications, appropriate, inappropriate, appropriation, appropriations, archangel, archangels, arid, ark, arrears, arrival, arrivals, departure, ascetic, aspect, aspects, assign, assigns, assigning, assigned, assist, assists, assisting, assisted, help, asylum, asylums, attached, attachment, attachments, attack, attacks, attacking, attacked, attrition, auction, auctions, augustine, awning, awnings, canopy, axe, axes
- backwards, forwards, backyard, backyards, bacon, ballot, ballots, band, bands, bank, banks, baptism, baptisms, baptize, baptizes, baptizing, baptized, bar, bars, bargain, bargains, barren, fertile, basic, fundamental, basin, basins, batter, batters, battering, battered, bear, bears, bearing, bore, borne, carry, benefit, benefits, bet, bets, betting, bible, bill, bills, birth, births, bite, bites, biting, bit, bitten, bitter, bitterest, blanch, blanches, blanching, blanched, bland, blander, blast, blasts, blend, blends, blending, blended, blessed, blind, block, blocks, blow, blows, blowing, blew, blown, board, boards, bond, bonds, boom, booms, slump, booth, booths, cubicle, bosom, bosoms, chest, bottle, bottles, bounce, bounces, bouncing, bounced, rebound, bowl, bowls, box, boxes, brave, braver, bravest, courageous, cowardly, breath, breaths, breeze, breezes, brighten, brightens, brightening, brightened, brisk, brisker, energetic, bruise, bruises, brush, brushes, burn, burns, burning, burnt, burst, bursts, bursting, bury, buries, burying, buried, bus, buses, butcher, butchers
- cable, cake, cakes, calm, calmer, canal, canals, cancel, cancels, cancelling, cancelled, cap, caps, capacity, capacities, capture, captures, capturing, captured, carriage, carriages, cart, carts, wagon, cat, cats, catastrophic, disastrous, catch, catches, catching, caught, cement, chamber, chambers, chance, chances, channel, channels, station, chapel, chapels, char, chars, charring, charred, charity, charities, chase, chases, chasing, chased, pursue, check, checks, checking, checked, cheer, cheers, cheering, cheered, chill, chills, chilling, chilled, heat, christen, christening, christened, christianity, chuck, chucks, chucked, throw, churchill, cistern, cisterns, tank, claim, claims, claiming, claimed, maintain, clean, cleaner, cleanest, dirty, cloud, clouds, club, clubs, coach, coaches, trainer, collect, collects, collecting, collected, gather, collector, collectors, combine, combines, combining, combined, comfort, discomfort, comfortable, uncomfortable, communion, communism, capitalism, community, compete, competes, competing, competed, concede, concedes, conceding, conceded, conceive, conceives, conceiving, conceived, concession, concessions, conduct, conducts, conducting, conducted, congress, congresses, conservative, tory, console, consoling, consoled, constituent, constituents,

contact, contacts, content, contents, continue, continues, continuing, continued, contrast, contrasts, conversion, conversions, cook, cooks, cooking, cooked, cool, cooler, coolest, warm, cosy, homely, council, councils, county, counties, credit, crisp, crisper, soggy, croft, crop, crops, crush, crushes, crushing, crushed, cup, cups, customary, usual

- debate, debates, discussion, decider, declare, declares, declaring, declared, announce, defence, defences, protection, defender, defenders, defensive, deliver, delivers, delivering, delivered, deluge, deluges, flood, democracy, democratic, demonstrate, demonstrates, demonstrating, demonstrated, den, dens, lair, deny, denies, denying, denied, repudiate, admit, deposit, deposits, designate, designates, designating, designated, desire, desires, desperate, dice, direct, indirect, dirtier, grubby, dispensation, dispense, dispenses, dispensing, dispensed, dispute, disputes, distance, distances, distant, faraway, nearby, distribution, dispersal, divine, division, split, dock, docks, doctor, doctors, dog, dogs, mess, domain, domains, domestic, home, foreign, double, draft, drafts, drain, drains, draining, drained, draw, draws, drawing, drew, drawn, sketch, dream, dreams, dress, dresses, dried, dehydrated, drive, drives, driving, drove, driven, drizzle, dryer, dryers, duck, ducks, dump, dumps, dumping, dumped, dust, dustbin, dustbins, dusty, dustier, dustiest, duty
- easter, elder, election, elections, electronics, enrich, enriches, enriching, enriched, enslave, enslaved, enter, enters, entering, entered, entertain, entertains, entertaining, entertained, establishment, creation, estate, estates, eviction, evictions, exchange, exchanges, exchanging, exchanged, swap, executive, executives, expense, expenses, exposure, extend, extends, extending, extended, stretch
- fade, fades, fading, faded, fail, fails, failing, failed, succeed, fair, fairer, fairest, faithful, familiar, farm, farms, fast, faster, fastest, quick, slow, favourable, favourite, fee, fees, feed, feeds, feeding, fed, fetid, stinking, feud, feuds, vendetta, fight, fights, fighting, fought, figure, figures, form, final, last, finger, fingers, firm, firms, company, fit, fits, fitting, fitted, flood, floods, flooding, fold, folds, folding, folded, folk, people, folks, forlorn, formation, foster, foul, foulest, disgusting, frame, frames, freeze, freezes, freezing, froze, frozen, fresh, new
- gamble, gambles, risk, gentile, gentiles, gentle, gentler, gentlest, glorious, magnificent, gospel, gospels, grace, grand, grander, grandest, majestic, grandfather, grandfathers, grandmother, grandmothers, grant, grants, grass, grasses, grey, grip, grips, gripping, gripped, groom, grooms, guard, guards, guarding, guarded, guest, guests
- hall, halls, hat, hats, hate, hates, hating, hated, love, haul, hauls, hauling, hauled, hazard, hazards, haze, hazy, clear, heats, heating, heated, hell, hells, history, holder, holders, hole, holes, holiday, holidays, homeless, plain, homestead, honour, honours, honouring, honoured, honourable, hopeful, host, hosts, household, households, housekeeping, housewife, housing, accommodation, hovel, hovels, shack, hurdle, hurdles, obstacle
- immediate, instant, improve, improves, improving, improved, incompetent, competent, incredible, fantastic, independent, separate, infidel, infidels, influence, institution, institutions, insulate, insulates, insulating, insulated, shield, insure,

insures, insuring, insured, integrate, integrates, integrating, integrated, interrupt, interrupts, interrupting, interrupted, introduce, introduces, introducing, introduced, inundate, inundated, swamp, invest, invests, investing, invested, investment, irrigate, irrigated, issue, issues

- jam, jams, join, joins, joining, joined
- kitchen, kitchens, kite, kites, knife, knives
- label, labels, labour, labours, landlady, landladies, lap, laps, lash, lashes, eyelash, layman, laymen, lead, leads, leading, led, leg, legs, lift, lifts, lifting, lifted, litter, rubbish, local, lodge, lodges, loss, losses, lure, lures, lured, trick,
- mackerel, mail, post, maintenance, majority, minority, manage, manages, managing, managed, run, mandate, mandates, map, maps, marital, market, markets, marriage, marriages, master, masters, mat, mats, match, matches, game, maternal, mere, merest, mild, milder, slight, great, milk, ministry, ministries, minute, minutes, minuted, miserable, mission, missions, mistake, mistakes, error, mobile, model, models, modest, unassuming, moisture, moral, ethical, mormon, mortgage, mortgages, motion, movement, mp, mps, muddy
- nay, indeed, neat, neater, neck, necks, negative, positive, net, netting, network, networks, nominate, nominating, nominated, notice, notices, noticing, noticed, nurse, nurses, nursery, nurseries, nut, nuts
- offset, offsets, offsetting, opening, first, opposition, oratory, rhetoric, overdue, overwhelming, overpowering, owe, owes, owing, owed
- pack, packs, packing, packed, pad, pads, paddock, pain, pains, paint, paints, palace, palaces, parade, parades, parcel, parcels, parent, parents, parish, parishes, park, parks, parody, parodies, pasture, pastures, patrimony, pepper, permanent, temporary, persecute, persecutes, persecuting, persecuted, persistent, pet, pets, pick, picks, picking, picked, choose, pig, pigs, hog, pin, pins, placement, plastic, plastics, plate, plates, platform, platforms, plop, ploy, plumb, plumbs, plumbing, plumbed, fathom, plunge, plunges, plunging, plunged, policy, policies, polls, pool, pools, popular, unpopular, possibility, possibilities, practise, practises, practising, practised, presence, absence, preserve, preserves, preserving, preserved, press, presses, pressing, pressed, price, prices, priesthood, primary, main, prime, privilege, privileges, pro, pros, professional, amateur, profession, professions, progressive, project, projects, scheme, propose, proposes, proposing, proposed, suggest, protestantism, proud, prouder, proudest, ashamed, provision, puff, puffs, puffing, puffed
- quarrel, quarrels
- race, races, rain, rains, raining, rained, raise, raises, raising, raised, rapid, rate, rating, rated, merit, raw, reclaim, reclaiming, reclaimed, recognize, recognizes, recognizing, recognized, reconcile, reconciles, reconciling, reconciled, referendum, reflect, reflects, reflecting, reflected, show, refuge, register, registers, regular, irregular, reject, rejects, rejecting, rejected, relative, relatives, relations, relief, religion, remain, remains, remaining, remained, removal, remove, removes, removing, removed, renew, renews, renewing, renewed, resume, repair, repairs, repairing, repaired, repeat, repeats, repeating, repeated, report, reports, reporting,

reported, representative, representatives, delegate, republican, requisition, requisitioned, commandeer, reserve, reserves, reserving, reserved, reshuffle, reshuffles, reorganize, residence, residences, resident, residents, retreat, retreats, retreating, retreated, reunite, reunites, reuniting, reunited, ride, rides, riding, rode, ridden, rigid, flexible, ring, rings, ringing, rang, rung, phone, rise, rises, rising, rose, risen, romp, romping, romped, roof, roofs, root, roots, rot, rots, rotting, rotted, rough, rougher, roughest, smooth, row, rows, line, rubber, rude, ruder, impolite, polite, rule, rules, running, rush, rushes, rushing, rushed

- safety, sale, sales, salt, salvation, satisfaction, save, saves, saving, saved, saviour, saviours, scale, scales, scaling, scaled, climb, scene, scenes, schedule, schedules, timetable, schoolmaster, schoolmasters, scratch, scratches, scratching, scratched, screen, scrutiny, season, seasons, experienced, seat, seats, section, sections, secure, secures, securing, secured, seed, seeds, senior, seniors, serve, serves, serving, served, settle, settles, settling, settled, steady, settlement, settlements, agreement, severe, severest, shame, share, shares, sheet, sheets, shelter, shelters, sheltered, shine, shines, shining, shone, shoot, shoots, shooting, shot, shop, shops, store, shoulder, shoulders, shouldering, shouldered, sick, ill, simple, simpler, simplest, singer, singers, sink, sinks, sister, sisters, site, sites, slate, slates, slating, slated, snap, snaps, snapping, snapped, snow, snows, snowing, snowed, solid, liquid, soup, soups, sour, spin, spins, spinning, spun, spiral, spirals, spiritual, temporal, spit, saliva, split, splits, splitting, spoil, spoils, spoiling, spoilt, sponge, sponges, sponsor, sponsors, sponsoring, sponsored, finance, spoon, spoons, spot, spots, spring, springs, sprout, sprouts, sprouting, sprouted, squabble, squabbles, squabbling, squabbled, square, squares, squash, squashes, squashing, squashed, squeeze, squeezes, squeezing, squeezed, stamp, stamps, star, stars, stations, stationed, position, stay, stays, staying, stayed, steam, steep, steeper, steepest, step, steps, stick, sticks, twig, stoppage, stoppages, strike, storm, storms, stormy, stormier, stormiest, straight, straighter, straightest, straw, stream, streams, brook, street, streets, strengthen, strengthens, strengthening, strengthened, fortify, stretches, stretching, stretched, strikes, stuff, succeeds, succeeding, succeeded, suffrage, suitable, supply, supplies, supplying, supplied, suppress, suppresses, suppressing, suppressed, surplus, surprise, surprises, surrogate, substitute, survey, surveyor, surveyors, swamps, sway, sways, swaying, swayed, sweet, sweeter, sweetest, swell, swells, swelling, swelled, swollen, increase, swing, swings, swinging, swung, switch, switches
- tail, tails, tailing, tailed, shadow, taint, taints, tainting, tainted, tap, taps, tart, tarts, tea, teas, telephone, tender, tenderer, tenderest, tenement, tenements, test, tests, testing, tested, testament, testaments, testimony, thank, thaw, thaws, thawing, thawed, throws, throwing, threw, thrown, thumb, thumbs, ticket, tickets, tie, ties, tying, tied, tight, tighter, loose, timetables, tinker, tinkers, tinkering, tinkered, tip, tips, touch, touches, touching, touched, trade, trail, trails, track, tramp, tramps, trap, traps, trash, travel, travels, travelling, travelled, treaty, treaties, tricks, trickle, trickles, trickling, trickled, truck, trucks, lorry
- unattended, unconscious, unfair, unfit, unhappy, happy, unison, unofficial, official, untidy, tidy, update, urgent, useless, useful, usurp, usurps, usurping, usurped
- vault, vaults, venerate, venerating, venerated, revere, vet, vets, village, villages, violent, aggressive, vote, votes

- wait, waits, waiting, waited, warmer, warmest, wastage, waste, wastes, wasting, wasted, waterproof, watershed, wax, waxes, wealthy, wealthier, wealthiest, weather, wet, wets, wetting, wetted, whip, whips, whipping, whipped, whiting, wide, wider, widest, narrow, wild, wilder, wildest, win, wins, winning, won, lose, wing, winged, winner, winners, loser, worker, workers, workhouse, poorhouse, worth, wreck, wrecks, wrecking, wrecked

E.5 Common words vocabulary

The vocabulary of *common words*, consisting of 1406 words, follows.

- a, able, can, about, above, across, act, acts, acting, acted, add, adds, adding, added, extra, after, before, again, against, age, ages, ago, air, all, almost, nearly, alone, along, already, also, always, never, among, amidst, amount, amounts, an, ancient, and, angry, another, answer, answers, answering, answered, ant, ants, any, anything, are, area, areas, arm, arms, arming, armed, armaments, army, armies, around, art, as, ask, asks, asking, asked, at, attention, attentions, away
- back, backwards, forwards, bad, worse, worst, good, ball, balls, be, being, been, became, because, become, becomes, becoming, get, bed, beds, beds, began, begin, begins, beginning, begun, start, stop, beginnings, end, behaviour, conduct, behind, beige, believe, believes, believing, believed, think, belong, belongs, belonging, belonged, below, beside, best, better, betters, bettering, bettered, between, big, bigger, biggest, large, small, black, blacks, blacker, blackest, blood, blue, blues, bluer, bluest, down, body, bodies, book, books, booking, booked, born, both, bottom, bottoms, top, box, boxes, boy, boys, bread, break, breaks, breaking, broke, broken, bring, brings, bringing, brought, britain, british, brother, brothers, brown, browns, brownest, build, builds, building, built, construct, but, by, bye
- cheerio, call, calls, calling, called, came, cannot, capital, care, cares, caring, cared, case, cases, cast, casts, cause, causes, effect, central, centre, centres, century, centuries, certain, sure, change, changes, charge, charges, charging, charged, chief, chiefs, child, children, christ, church, churches, city, cities, class, classes, clear, clearer, clearest, close, closes, closing, closed, shut, open, cold, colder, coldest, colour, colours, come, comes, coming, command, commands, commanding, commanded, common, commoner, commonest, rare, condition, consider, considers, considering, considered, consist, consists, consisting, consisted, contain, contains, containing, contained, control, could, count, counts, counts, counting, counted, country, countries, course, cut, cuts, cutting
- dad, dark, darker, darkest, light, daughter, daughters, day, days, dead, alive, deal, death, deaths, deep, deeper, deepest, shallow, degree, degrees, did, die, different, difficult, easy, direction, directions, disease, diseases, divide, divides, dividing, divided, split, do, does, doing, done, door, doors, down, up, drink, drinks, drinking, drank, drunk, during
- e, each, early, late, earth, easily, east, easier, easiest, tense, eat, eats, eating, ate, eaten, effects, effecting, effected, eight, eights, eighteen, either, eleven, else, england, english, enough, enter, enters, entering, entered, equal, error, errors, especially, particularly, etc, europe, even, evening, evenings, event, events, ever, every, everybody, everyone, everything, nothing, everywhere, evil, example, examples, express, expresses, expressing, expressed, additional, eye, eyes
- face, faces, fact, fail, fails, failing, failed, succeed, fair, fairer, fairest, fall, falls, falling, fell, fallen, false, family, families, far, near, fast, faster, fastest, quick, slow, father, fathers, fear, fears, feel, feels, feeling, felt, feelings, emotion, opinions, feet, few, fifteen, fifth, fifty, fifties, fill, fills, filling, filled, empty, film, films, find, finds,

finding, found, discover, fine, finer, finest, fire, first, last, fish, fishes, five, fives, fix, fixes, fixing, fixed, fasten, set, varying, flat, flats, apartment, flesh, flower, flowers, foot, for, force, forces, forcing, forced, form, forms, former, forth, out, forty, forties, forward, founds, founding, founded, four, fours, fourteen, fourth, france, free, friday, fridays, friend, friends, ally, from, front, fronts, fruit, fruits, full, fuller, fullest, further, future, past

- game, games, gave, gets, getting, gotten, girl, girls, give, gives, giving, given, particular, glass, go, goes, going, went, gone, god, going, gold, best, goods, got, grade, grades, grading, graded, gray, great, greater, greatest, greek, green, greens, greener, greenest, ground, grounds, group, groups
- had, hair, hairs, half, hand, hands, happen, happens, happening, happened, hard, harder, hardest, soft, has, have, having, he, head, heads, hear, hears, hearing, heard, heart, hearts, heaven, paradise, heavy, heavier, heaviest, held, hello, help, helps, helping, helped, her, here, there, herself, high, higher, highest, low, him, himself, his, hold, holds, holding, holy, holier, holiest, home, homes, hope, hopes, hoping, hoped, hot, hotter, hottest, hour, hours, house, houses, how, however, nevertheless, human, hundred, hundreds, husband, husbands
- i, idea, ideas, if, ife, ill, sick, important, in, include, includes, including, included, excluding, indeed, input, inputs, inside, insides, outside, instance, instead, interesting, into, in, is, it, its, itself
- jesus, job, jobs, john, join, joins, joining, joined, just, justice
- keep, keeps, keeping, kept, kind, kinds, sort, king, kings, kingdom, kingdoms, kings, knew, know, knows, knowing, known, knowledge
- land, language, languages, larger, largest, later, law, laws, code, lay, lays, laying, laid, leaf, leaves, least, leave, leaving, left, length, lengths, less, more, let, lets, letting, letter, letters, level, levels, lex, lie, lies, lying, lain, life, light, lights, lighting, lighted, darkness, like, likely, probable, unlikely, line, link, links, connection, liquid, liquids, little, live, lives, living, lived, long, look, looks, looking, looked, lord, lords, lose, loses, losing, lost, lot, lots, love, loves, loving, loved, low, lower, lowest, upper
- machine, machines, made, main, mainly, primarily, make, makes, making, man, men, manner, way, many, mark, marks, marking, marked, mass, matter, matters, may, mays, might, maybe, perhaps, me, mean, means, meaning, meant, signify, meet, meets, meeting, met, meetings, metal, metals, middle, centre, may, mile, miles, loads, mind, minds, mine, minister, ministers, miss, misses, misses, missing, missed, hit, moment, moments, money, month, months, more, morning, mornings, most, mother, mothers, mouth, mouths, move, moves, moving, moved, movement, movements, touching, mr, mrs, much, music, must, my, myself
- name, names, narrow, narrower, narrowest, wide, nation, nations, natural, nature, need, needs, needing, needed, neither, new, newer, newest, next, nice, nicer, nicest, night, nights, nine, nines, nineteen, no, yes, none, nor, north, not, note, notes, message, now, number, numbers
- of, off, offer, offers, offering, offered, office, offices, often, frequently, rarely, oil, oils, old, older, oldest, elderly, young, on, once, one, ones, only, onto, opens, opening, opened, or, order, other, others, our, outsides, over, own

- paper, papers, part, parts, specific, party, parties, pass, passes, passing, passed, pay, pays, paying, paid, peace, people, definitely, period, periods, person, persons, phrase, phrases, piece, pieces, bit, place, places, plant, plants, play, plays, playing, played, please, plural, singular, point, points, police, poor, poorer, poorest, rich, port, ports, position, positions, location, possible, power, prep, homework, present, current, problem, problems, process, processes, produce, produces, producing, produced, public, pull, pulls, pulling, pulled, purpose, purposes, aim, push, pushes, pushing, pushed, put, puts, putting
- question, questions, quicker, quickest, quite
- rather, read, reads, reading, ready, real, genuine, imaginary, really, reason, receive, receives, receiving, received, red, reds, redder, refer, refers, referring, referred, region, regions, relate, relates, relating, related, remember, remembers, remembering, remembered, remind, reminds, reminding, reminded, rest, result, results, return, returns, returning, returned, right, correct, wrong, road, roads, rock, room, rooms, round, run, runs, running, ran, drip
- said, same, sat, saw, say, says, saying, sayings, school, schools, sea, seas, second, seconds, see, sees, seeing, seen, seem, seems, seeming, seemed, send, sends, sending, sent, sense, senses, service, services, set, sets, seven, sevens, seventeen, several, shall, shape, shapes, sharp, sharper, sharpest, blunt, she, ship, ships, short, shorts, shorter, shortest, long, should, show, shows, showing, showed, shown, side, sides, sight, vision, sign, signs, symbol, silver, similar, sin, sins, since, sing, sings, singing, sang, sung, single, singles, sir, sirs, sister, sisters, sit, sits, sitting, situation, situations, six, sixes, sixteen, sixteens, size, sizes, skin, skins, slower, slowest, smaller, smallest, smile, smiles, smiling, smiled, so, social, society, softer, softest, rough, some, someone, something, sometimes, somewhere, son, sons, soon, soul, souls, spirit, sound, sounds, south, space, spaces, speak, speaks, speaking, spoke, spoken, special, ordinary, speech, spell, spells, spelling, spelt, spellings, spirit, spirits, spirited, soul, spokes, stand, stands, standing, stood, state, states, still, stone, stones, stops, stopping, stopped, story, stories, tale, strength, strong, stronger, strongest, weak, study, studies, studying, studied, subject, subjects, topic, such, suddenly, sun, surer, doubtful
- table, tables, take, takes, taking, took, taken, talk, talks, talking, talked, tell, tells, telling, told, ten, tens, than, thank, that, the, thee, their, them, themselves, then, therefore, consequently, thereof, these, they, thick, thicker, thickest, thin, thinner, thinnest, thing, things, thinks, thinking, thought, third, thirteen, thirty, thirties, this, those, thou, though, although, thousand, thousands, three, threes, through, thus, thy, till, until, time, times, to, together, too, tops, towards, town, towns, tree, trees, true, truth, try, tries, trying, tried, turn, turns, turning, turned, twelve, twenty, twenties, two, twos, type, types
- under, understand, understands, understanding, understood, unit, units, unto, upon, on, upside, upstairs, downstairs, us, use, uses, using, used, usual, usually
- value, various, vehicle, vehicles, very, view, views, opinion
- walk, walks, walking, walked, wall, walls, want, wants, wanting, wanted, was, water, waters, watering, watered, ways, we, weaker, weakest, well, were, west, what, when, where, whether, which, while, white, whites, whiter, who, whole,

whom, whose, why, wife, wives, will, wind, winds, with, within, without, woman, women, wood, woods, word, words, work, works, working, worked, world, would, write, writes, writing, wrote, written

- year, years, yellow, yellows, no, yet, you, young, younger, youngest

E.6 Phono-Graphix vocabulary

The vocabulary for the Phono-Graphix Method, consisting of 1152 words, follows.

- a, about, absent, absolute, accident, acorn, acrobat, act, active, address, advance, adventure, again, ahoy, aim, all, alp, amnesia, amp, and, animal, ant, anxious, any, appear, apple, are, asia, asp, at, ate, atom, atrocious, august, autumn, awful
- baby, back, bad, bag, bait, balance, ball, barge, bash, bat, beach, bean, bear, became, bed, beep, before, bell, ben, bend, bet, beyond, bib, bid, big, bin, bird, bishop, bit, bitter, bitterly, black, blind, blip, blizzard, block, blot, blouse, blue, boast, boat, boil, booklet, boom, boot, bought, bountiful, box, boy, brain, bread, break, breeze, brick, bridge, brief, brighten, brown, bruise, brute, bud, budge, budget, bug, bulb, bump, bun, burglary, burn, burp, bush, but, but, buzz, buzzard
- cake, calculate, camel, camp, can, candle, candy, cane, cannibal, canoe, cap, cape, capital, captain, capture, car, carpenter, carpeting, carry, cashier, castle, cat, cattle, caught, celebrate, cent, centigrade, centre, certain, certainly, chalk, chap, character, chat, chemical, chicken, chief, child, chill, chimney, chimp, chin, chip, chlorine, choice, choose, chop, christmas, chuck, chug, churn, cinder, circus, city, clam, claw, clay, clean, clock, closet, clown, club, coal, coast, cod, code, coffee, coil, collar, collision, colour, come, computer, concert, cone, conscious, consequence, containment, control, cook, cop, costume, cot, cottage, could, count, country, cow, crayon, cream, crew, cricket, criminal, crisp, crocodile, crook, crop, crow, crude, cruise, crumble, cry, crypt, cub, cube, cucumber, cue, curl, curve, customer, cut, cute, cylinder
- dad, damp, dan, dancer, dawn, day, dead, deaf, debt, december, deep, delicious, deliver, dementia, dent, dentist, deposit, descriptive, desert, desk, destroyer, detective, detergent, develop, did, die, dig, dim, dime, dimension, dinner, dirt, dirty, disappear, discover, dish, dive, do, dock, dog, doll, dollar, done, dove, down, dread, dream, drip, drop, drought, drum, duck, dug, dumb, during, dynamite
- each, eagle, early, earn, earth, ease, east, eat, eclipse, eight, election, electric, electricity, elephant, elf, elk, elm, embarrass, embroider, end, energy, engineer, enough, enter, escapade, establish, eve, example, excuse, exercise, expensive, expert, explanation, explosive, expressway
- facia, fact, fake, fancy, fantasy, fat, fate, father, fawn, feather, feel, feet, fellow, fern, ferocious, feverish, few, fib, fictitious, fiddle, field, fin, fine, fingerprint, first, fish, fitted, five, flag, flake, flew, flirt, float, flow, flower, flute, fly, foal, foe, foil, food, foot, for, forest, forgetful, forward, fought, found, fountain, four, fox, fraud, freesia, freeze, freezer, friend, frog, from, frown, fry, fudge, fuel, full, fun, funny, fur, furniture, future, fuzz
- gadget, galaxy, gamble, garage, garbage, gardener, gate, gather, generous, gentle, gesture, get, ghetto, ghost, gift, gin, ginger, girl, give, glad, glass, glow, glue, gnat, gnome, go, goat, golf, gopher, gosh, gossip, got, governor, gown, grab, graduate, graf, graph, grass, gratitude, gravity, great, grey, grief, grin, grog, grouch, ground, group, grow, gym

- had, hag, haggles, hammer, hand, handicapped, handle, hanger, happen, happily, happy, harbour, harm, hate, haul, have, hawk, he, head, heard, heat, height, held, helmet, help, her, herd, hesitate, hid, hide, high, hill, him, hind, his, hiss, hit, hive, hollow, home, hook, hoot, hop, horrify, hospital, host, hot, hour, house, how, howl, hue, hug, hum, hump, hurricane, hurt, husband
- i, immersion, imp, in, indian, inertia, infant, ink, innocent, insect, inspire, instead, is, it
- jack, jam, jan, jaw, jeep, jerk, jet, jewel, jig, job, judge, jug, juice, jump, jungle, just
- kangaroo, keep, kept, key, kid, kiss, kit, kite, knee, knew, knot, know, known
- label, ladder, lake, lamp, landed, lantern, lap, lard, large, late, law, lawn, lead, leap, leapt, learn, leave, led, leisure, less, lesson, let, lettuce, lick, lie, lift, light, lightning, lip, little, loaf, lobster, lock, loft, log, lollipop, look, loose, loot, lose, lost, lot, loud, love, loyal, luck, lucky, luggage, lumber, lump, lurk
- machine, mad, made, magazine, magic, magnetic, main, make, malicious, mammal, man, mane, map, maple, mat, mat, match, math, may, me, meagre, mean, meant, measure, meat, mechanic, meet, men, mercy, mess, messy, met, metre, microphone, microscope, middle, mild, miles, militia, milk, mind, mineral, minister, mint, mission, mitt, momentous, monday, monkey, monster, month, monument, moon, moose, mop, mope, most, moth, mountain, mouse, much, mud, muffin, mug, mule, multiple, multiply, mum, museum, music, musical, musician, muskrat, must, mustard, mut, mute, mutt, my
- nap, nation, national, naval, navigate, neck, need, nerve, net, new, nice, night, nip, no, nod, noon, not, note, noun, now, nudge, numb, numbers, nursery, nut, nutritious
- observer, obstacle, occupy, off, office, oil, oink, olive, on, operate, orange, out, over, overflow, owl, oyster
- packed, pain, panther, paper, parade, parakeet, parallel, parrot, pass, past, pasture, pat, peak, peanut, penny, persia, pet, petite, petticoat, pew, pewter, pharmacy, phone, phonics, physician, pie, pile, pillow, pin, pine, pip, pipe, pit, plant, plantation, play, please, pleasure, plenty, plug, plum, plus, pneumonia, pocket-book, poison, poke, pole, polish, pond, pop, posture, pot, pottery, pout, pout, powder, pray, precocious, president, prevent, prey, priest, principal, principle, print, printer, product, projector, prune, pry, publicity, pudding, puddle, puff, pump, pumpkin, pupil, purple, put, putt
- quarrel, quarter, queen, quick
- rabbit, rag, rain, rake, ran, rap, rapid, rat, rate, raw, ray, read, real, receive, reckless, recommend, recorder, recover, rectangle, red, reflect, refrigerator, register, repair, reporter, reptile, request, rhino, rick, rid, ride, rift, rig, right, rip, ripe, ripple, risk, river, roast, roast, rob, robe, robin, rock, rocket, rod, rode, roller, room, roost, root, rough, round, royal, rubber, rudder, rude, rug, run, runner, rust

- sack, sad, safety, said, sail, sal, sale, sam, sample, sand, sat, satisfied, saturate, saturday, saw, say, scared, science, scold, scoop, sea, seat, see, seem, seen, sensible, serve, service, set, sham, shampoo, shape, she, shed, sheet, shell, shield, shin, ship, shirt, shoe, shook, shoot, shop, shot, should, show, shut, shy, sick, sigh, silly, simmer, simple, simply, sip, sir, sis, sit, skip, skirt, sky, sled, sleigh, slot, small, smart, smile, smit, snail, snake, sneak, sneeze, snicker, snob, snooze, snow, so, sob, sock, soft, soil, some, soon, sop, sot, sound, soup, spider, spin, spoil, spoon, spot, spread, spy, stable, stain, stale, star, steak, steal, steam, stew, stick, stir, stob, stock, stop, store, straight, street, stripe, stroke, structure, struggle, stuck, stuff, stump, sugar, suit, summer, sun, sunny, super, surround, suspension, swag, swat, sweat, sweet, swig, swish, symbolic
- tab, table, tail, talent, talk, tap, tape, target, taught, teacher, team, telephone, television, tell, ten, tend, tender, tent, that, the, them, then, there, they, thin, things, this, thought, thread, threatening, throat, through, throw, thug, thunder, thursday, tick, tie, tight, tile, time, tip, to, toast, today, toe, toil, told, tolerant, tom, tone, took, top, torment, tornado, tot, tote, touch, tough, tow, tractor, trampoline, transform, transparent, transport, trap, travel, tray, tread, treasure, treat, tree, tremendous, trip, trombone, troop, troy, truck, true, trust, tub, tuck, tuesday, tug, tune, tunnel, turn, twig, twin, twine
- umbrella, under, uneasy, up, use
- vacation, valley, vanish, vapour, variation, vein, very, vet, vine, vision, visit, voice
- waist, wait, walk, wall, want, waste, water, way, we, web, web, weigh, weight, wet, what, when, whether, whip, whisker, white, whiz, whole, whose, why, wig, wiggle, wild, will, win, wine, winner, winter, wish, wit, with, wither, wonderful, wood, work, world, worm, worried, worry, worse, worth, would, wrap, wreck, wrench, wrinkle, write
- xylophone
- yawn, yes, your, yum
- zip

E.7 Jolly Phonics vocabulary

The Jolly Phonics vocabulary, consisting of 749 words, follows.

- aid, aim, also, am, an, ant, any, art, as, august, autumn
- back, bad, bag, bake, ball, band, bang, bar, bark, barn, barrel, bat, because, bed, bee, belt, bench, bend, bendy, bent, best, big, bin, bird, bit, bite, bleed, blew, blister, block, blot, blunt, boat, body, bog, boil, bone, book, boot, born, bother, box, boy, bran, brand, brew, brick, bright, bring, broom, brown, brush, bud, bug, bulb, bump, bun, bunch, burn, burnt, burst, bus, but, butter, buzz
- cake, came, camp, can, cap, car, caravan, card, cart, cash, cat, cave, chain, chalk, chap, charm, chart, chat, check, chest, chick, chill, chimp, chin, chip, chop, chum, church, clang, clap, claw, clever, cling, clip, close, cloth, cloud, clown, club, coach, coal, coat, cod, coil, coin, cook, cork, corn, cot, could, count, cow, coy, crab, crash, cream, creep, crisp, crook, crop, crowd, crush, crust, cry, cube, cue, cup, curl, cut
- dad, daddy, damp, date, day, deck, den, did, die, dig, digger, dip, dirt, dish, dive, diver, doctor, dog, dot, down, drag, drain, draw, dream, drink, drip, drop, drown, drug, drum, duck, due, dug, duke, dump
- eat, eel, egg, elephant, elf, employ, end, enjoy
- faint, fan, farm, fat, fault, fed, feed, feet, few, fight, fill, film, fin, first, fish, fist, fit, five, fix, fizz, flag, flame, flap, flash, flat, flew, flex, flip, float, fluff, fly, foam, fog, foil, food, fool, for, forget, fork, fort, found, fox, fraud, free, fresh, frill, frog, from, front, frost, frown, fry, fun, funny, fuss
- game, gap, gas, gate, get, gift, girl, give, glad, glue, goal, goat, golf, good, got, grab, gran, grand, greed, green, grill, grim, grip, grit, groin, ground, gruff, gulp, gun
- had, ham, hand, handy, hang, hard, harsh, hat, haunt, hay, held, help, hem, hen, hid, high, hill, hilly, him, hip, his, hiss, hit, hog, hoist, home, hook, hop, hope, horn, hot, how, hug, hum, hung, hunt, hush, hut
- ice, if, imp, in, indian, ink, is, it
- jail, jam, jar, jaw, jelly, jet, job, jog, join, joint, jolly, joy, jug, jump, jumper, june, just
- keep, kept, kerb, kettle, kick, kind, king, kiss
- lad, lamp, land, landed, lane, lap, late, later, laugh, law, leaf, leg, lend, lent, let, lick, lid, lie, life, lift, light, like, limp, line, lip, list, lit, litter, loaf, loan, lock, loft, log, lolly, long, look, lost, lot, loud, luck, lump, lunch, lung
- mad, main, man, map, march, mat, meet, melt, men, mend, met, mice, milk, mill, mint, miss, mist, misty, mix, moan, moist, mole, moo, moon, moth, mother, mouse, mouth, much, mud, mug, mum, mummy, must
- nail, nap, neck, nest, net, new, next, nice, night, nip, noon, north, not, now, nurse, nut

- oak, oat, oil, ointment, omelette, on, or, orange, order, out, owl, ox
- pack, paid, pain, paint, pan, pant, park, part, pat, pause, peck, peep, peg, pen, people, pest, pet, pie, pig, pillow, pin, pinch, pink, pip, pipe, pit, plan, plate, play, plot, plug, plum, plump, pod, point, pond, pop, porch, pork, port, pot, pram, pray, print, prop, proud, pub, pudding, purse, push
- queen, quick, quilt, quin
- rail, rain, ram, ran, rash, rat, raw, read, red, rich, ride, ring, rip, road, rob, rock, rod, roof, root, rope, rot, round, rub, rude, rug, run, running, runny, rush
- sack, sad, sail, sand, sang, sat, save, saw, scarf, scream, scrub, seat, see, seed, seeds, set, shall, shark, sharp, shawl, shed, sheep, sheet, shelf, shell, shine, ship, shirt, shock, shook, shop, short, shot, should, shout, show, shrimp, shut, shy, sift, silk, silly, sister, sit, six, sixth, skate, skin, sky, slab, slap, sleep, slept, slim, slip, slope, slow, slug, smart, smash, smelly, smog, smoke, snail, snap, sniff, snip, snow, snug, soak, soap, soft, soil, song, soon, south, sparrow, speak, speed, spell, spend, spill, spin, spit, split, spoil, spoon, sport, spot, spun, stamp, stand, star, start, stay, step, stern, stew, stick, sting, stir, stole, stone, stood, stool, stop, stork, storm, straw, stream, street, string, strong, stuck, such, suck, sudden, summer, sun, sunny, supper, swam, sweep, sweet, swim, swing, swum, swung
- tail, tan, tap, tart, team, teeth, ten, term, test, thank, that, them, then, thick, thimble, thin, thing, think, thinner, third, thirsty, this, those, three, thump, tidy, tie, tip, toast, toil, tooth, top, torch, toy, train, trap, tree, trim, trip, tub, tube, tuesday, tug, turn, turnip, twig, twin, twist
- ugly, under, up, us
- van, very, vet
- waist, walk, wave, wax, web, wednesday, weed, week, weep, well, went, west, wet, wife, will, win, wind, windy, wing, winter, wish, with, wood, wool, would
- yap, yard, yawn, yell, yet
- zebra, zip, zoo

E.8 Tricky words vocabulary

The *Tricky words vocabulary* for the Jolly Phonics Method, consisting of 58 words, follows.

- any, are
- be, because, before, by
- come, could
- do, does, down
- four
- give, go
- have, he, here
- i
- like, little, live
- made, many, me, more, my
- no
- old, one, other
- put
- right
- said, saw, she, should, so, some
- the, their, there, they, to, two
- want, was, we, were, what, when, where, which, who, why, would
- you

E.9 THRASS vocabulary

The vocabulary for the THRASS Method, consisting of 268 words, follows.

- a, and, ant, any, august
- baby, ball, banana, beach, bed, bell, bird, blaze, blend, blind, block, blow, boat, bold, book, box, brain, bread, brick, bridge, brown, brush, bull, bus
- cage, car, cat, chair, cheese, chef, cherry, circus, city, claw, cliff, climb, cloud, clown, coffee, coin, cold, collar, cow, crab, crane, crisp, crown, crumb
- december, deer, desk, dinner, doctor, dog, dolphin, door, draw, dream, dress, drink, drive, drum, duck
- ear, egg, emu
- feather, february, fern, fish, fizz, flag, flame, flea, float, floor, fly, fold, fork, fossil, freeze, friday, fridge, fringe, frog, frost, froth, fur
- garden, gate, giant, glass, glider, glitter, globe, glove, glue, gold, grain, grape, grass, green, grip
- hair, hammer, hand, he, hippo, horse, house
- i, ice, in, ink, is, it
- jam, january, july, june
- key, king, kite, kitten, knee
- ladder, lamb, laser, leg, letter, light, lion
- march, may, me, measure, monday, moon, moor, mouse
- nest, net, nose, note, november / october, of
- panda, plant, plate, pleat, plug, plum, pond, pony, pram, prawn, press, prince, prize
- queen, quilt / rabbit, rain, rocket
- sand, saturday, sauce, saw, school, screw, september, shark, shirt, shrub, skate, skip, skull, sleep, sleeve, slice, slide, slug, small, smash, smell, smile, smoke, snack, snail, snake, snatch, sneeze, snooze, snow, sold, sound, spade, spear, spoon, sport, spot, spring, square, stand, star, station, step, stir, straw, stream, street, string, strong, sun, sunday, swallow, swamp, swan, swim, swing, switch
- tap, tape, teacher, that, the, thread, throat, throne, thumb, thursday, tiger, tin, to, toast, toy, train, tray, treasure, tree, trip, troop, trot, trunk, tuesday, tusk, twig, twilight, twin, twine, twist
- voice
- was, watch, water, wednesday, wheel, worm, wrist
- yawn / zebra, zip

Appendix F

Grammars

F.1 Knowledge Tree Definition Language - KTDL

$\langle \text{KTDL program} \rangle ::= \langle \text{KTDL clause} \rangle \mid$
 $\langle \text{KTDL program} \rangle \langle \text{KTDL clause} \rangle$

$\langle \text{KTDL clause} \rangle ::= \langle \text{hierarchy number} \rangle \langle \text{nodename} \rangle$
 $\langle \text{skills sequence} \rangle \langle \text{commands} \rangle$

$\langle \text{hierarchy number} \rangle ::= \langle \text{doted number} \rangle -$

$\langle \text{doted number} \rangle ::= \langle \text{digit} \rangle \mid$
 $\langle \text{digit} \rangle \langle \text{doted digits} \rangle$

$\langle \text{doted digits} \rangle ::= . \langle \text{digit} \rangle \mid$
 $\langle \text{doted digits} \rangle . \langle \text{digit} \rangle$

$\langle \text{digit} \rangle ::= 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9$

$\langle \text{nodename} \rangle ::= \langle \text{sentence} \rangle$

$\langle \text{sentence} \rangle ::= /* \text{ sequence of caracteres without dots and colons } */$

$\langle \text{skill sequence} \rangle ::= : \langle \text{skills} \rangle . \quad | \quad .$

$\langle \text{skills} \rangle ::= \langle \text{skill} \rangle \mid$

<skills> , <skill>

<skill> ::= <sentence>

<commands> ::= <empty> |
 <command> . |
 <commands> <command> .

<empty> ::= /* nothing at all */

<command> ::= &exclude |
 &source=<sourcename> |
 <phonetic command>

<sourcename> ::= /* sequence of up to ten letters */

<phonetic command> ::= <column feature> |
 <phonetic command> <column feature>

<column feature> ::= V | C | X | <phoneme> | <letter> | \$ | * |
 [<column specifiers>] |
 { <column specifiers> }

<phoneme> ::= /* any of the representations of the 44 english phonemes */

<letter> ::= /* any of the lower case english alphabet symbols */

<column specifiers> ::= <column feature> |
 <column specifiers>,<column feature>

F.2 Teaching Strategy Definition Language - TSDL

<TSDL program> ::= <TSDL clause> |
 <TSDL clause> <TSDL program>

<TSDL clause> ::= <if command> | <action command>

<if command> ::= if <boolean expression> then <TSDL clause> . |
 if <boolean expression> then <TSDL clause>
 else <TSDL clause> .

<boolean expression> ::= <boolean primary> |
 <boolean expression> and <boolean expression> |
 <boolean expression> or <boolean expression> |
 not <boolean expression> |
 (<boolean expression>)

<boolean primary> ::= <activity attribute> (<nodename>) <relationship>
 <degree> |
 <degree> <relationship> <degree> |
 <skill> (<nodename>) <relationship> <degree> |
 <activity> (<nodename> | anyknowledge) visited |
 <activity> (<nodename> | anyknowledge)
 scored (<degree>)

<activity attribute> ::= average | learning | exemplifying | exercising | testing

<relationship> ::= is | isnt | greater | less | notgreater | notless

<degree> ::= poor | enough | good | excelent | <value> | <variable> |
 <system variable>

<value> ::= <integer> | - <integer>

<integer> ::= <digit> | <integer> <digit>

<activity> ::= lesson | example | exercise | test

Appendix G

Object Specifications

The Authoring Tool for describing the activities in the prototype of the system was implemented through electronic forms that allow the specification of the steps of the activity. So far, as already mentioned in Chapter 7, there are twelve different objects. In this appendix the forms for the objects and their fields are discussed.

G.1 Start-Activity Object

Every activity must start with the declaration of the Start-Activity Object. The attributes of the activity are:

Start-Activity Object

Id Number

Name

Banner

Song

Duration

Subject

Activity-type

Complexity Level

Skill

Colour

The *Id Number* is a number to identify uniquely each step of the activity. It is present not only in the Start-Activity Object but in every object, and is used to indicate the sequence of execution of the steps. The choice of non-contiguous values can be useful in the definition of order number of the steps, to make it easier to insert steps (we should, for example, enumerate the steps with an interval of 10 units, that is, 10, 20, 30...).

The *Name*, *Banner* and *Song* fields are the attributes that identify the activity for the student: its name will be said by the computer while a picture indicated in

Banner appears on the screen and a background song is played in the beginning of the activity.¹ The interval of time for displaying the first screen is defined in *Duration*. The *Subject*, in hierarchical dotted numbers, defines the location of the activity in the Knowledge Tree. The type of activity (lesson, example, exercise or test) and its complexity level (from 1 to 5) are declared in the fields *Type*, and *Complexity* respectively. The skills associated with the activity are declared in the field *Skill*. The field *Colour* defines the background colour of the first screen of the activity. An example of the Start-Activity form is shown in Figure G.1.

The screenshot shows a window titled "Literacy Teacher Authoring Tool". Inside the window, the text "Enter Start-Activity Fields:" is centered. Below this text, there are two columns of input fields. The left column contains: "Id Number" with value "10", "Banner" with value "fatcat.gif", "Duration" with value "8", "Type" with value "exercise", and "Skill" with value "Code Knowledge". The right column contains: "Name" with value "Identifying Phonemes - Fat Cat", "Song" with value "phon.au", "Subject" with value ".1.1", "Complexity" with value "1", and "Back Color" with value "yellow". At the bottom left of the form area is a button labeled "Give up".

Figure G.1: Start-Activity Object

G.2 End-Activity Object

The End-Activity object is used to declare the stop point of an activity. The object allows the definition of a *Banner* and a *song* to be showed and played as soon as the activity ends. The form is shown in Figure G.2.

¹Songs are not played yet in the current version of the system.

Enter End-Activity Fields:

Id Number	<input type="text" value="160"/>	Banner	<input type="text" value="fatcat.gif"/>
Song	<input type="text" value="pho.au"/>	Duration	<input type="text" value="8"/>

Figure G.2: End-Activity Object

G.3 Input Object

This object is conceived to allow the student to input data either through the keyboard, microphone or hand-writing device. More than one kind of input can be indicated at once. In this case the order of execution of the input operations should be defined. An example of the *Input Object* form is shown in Figure G.3.

The fields *Id Number* and *Count variable* are located in the first block of the form. *Id Number* indicates the sequence of the step within the activity. The field *Count variable* appears not only in this object, but in most of them, to name and enable a counter variable for the number of times this step will be executed during the activity. The counter will receive the name indicated and it can be later referred to by other objects as an *emphactivity variable*. This will be useful for counting numbers of tries, number of success and numbers of fails, for instance. If the field *Count variable* is left blank, there will be no counting for this step.

There are three possible ways for prompting the student: *voice-prompt*, *text-prompt* and *buzz-prompt* (the prompt will indicate to the student that some kind of input should be made at this point). The *Timbre* of the voice must be declared (*male* or *female*). If the *speech file* is declared the *student* utterance will be recorded and kept in this file. If the *speech variable* is declared, the coded form

Literacy Teacher Authoring Tool

Enter Input Object Fields:

----- General -----

Id Number Count variable

----- Prompt Section -----

Voice prompt Text prompt

Voice timbre Prompt buzz

----- Speech Section -----

Speech file Speech variable

Speech duration Speech order

----- Handwriting Section -----

Hand variable HandW duration

HandW order

----- Keyboard Section -----

Keyboard variable Keyboard duration

Keyboard delay Keyboard order

Figure G.3: Input Object

of the speech will be kept in this variable. The parameter *speech duration* defines for how long the microphone will remain activated. If *hand variable* is declared then the interpretation of the hand writing input will be available in this variable (this facility is not implemented yet and its details will emerge in the future). If *Keyboard variable* is declared then the keyboard is activated for typing data in, and the input data will be placed in that variable. The *keyboard duration* defines for how long the keyboard will remain activated. The variables labelled related with order will be used in the cases where there are more than one operation at once, to sequence the operations.

Record-Speech operation is used to allow the recording of an utterance, ended by a click in the mouse, giving origin to a recorded file. The parameters are the *file-name* and the *maximum recording time*. The file originated in this operation can be used later either for reproduction or for more detailed analysis by the program (for instance, applying speech recognition to be accepted as answers to questions posed by the system).

G.4 Output Object

This object will allow some output to be produced either as speech, written material on screen, song or picture. It is possible to declare different forms of output in the same step, in which case the order to execute these operations must be declared. The form is divided in sections, and the lack of information in one or more sections implies that the type of output of that section will not be produced in this step. A *Counter variable* can be provided for the object, so it will register how many times the step has been executed in the activity.

In the *Song section* the optional background *song* to be played during the execution of the step is defined. In the *Speech Section* the optional *speech* to be said in the step is described. It can come either from a pre-recorded *speech file*, or from a text string to be spoken or even from a text file to be converted into speech and then spoken. The *duration* parameter determines the maximum speech time, the *timbre* defines if it is to be said by a male or female voice, and the *order* determines when it is going to be said, in relation to the other output elements. The *speech image* parameter defines an image to be shown while the speech is going on.

The fields related to the *Write Section* are: *Text to write*, intended to contain the string to be written (enclosed in double quotes) or the name of the set with the picture sounds to be placed on the screen; *Duration*, to indicate how long this screen image should last; *Position*, to indicate where the text is to be placed in the screen; *Colour*, to define the colour of the letters; *Background Colour* and *Shape*, to define a shell for the text; and the *text order*.

Only part of the Speech Section and Write Section are implemented in the prototype of the system. The *Animation Section* and the *Picture Section* field interpretation is straightforward. An example of the form is shown in Figure G.4.

G.5 Show-Scene Object

*Show scene*² is a particular case of output, which, by its complexity, is considered separately from the other forms of output. This object is a very basic “comics composer”, which will be able to represent either one, two or three *characters* in a *scenery*, each one with an optional *speech balloon* over their head. The characters can be a *Man*, a *Woman*, a *Boy* or a *Girl*, either *Sad*, *Serious* or *Happy*, either turned to the *Left*, to the *Right* or looking *Ahead*, either *Blond*, *Brown* or *Dark*, and either *Black* or *White* (for the skin). The background scenery

²Not implemented yet.

Literacy Teacher Authoring Tool

Enter Output Object Fields:

----- General -----

Id Number Count var

----- Song Section -----

Song file Song Time

Song order

----- Speech Section -----

Speech file Text to speak

Speech duration Speech order

TTS file Text timbre

Speech image

----- Write Section -----

Text to write Duration

Font Size Text Position

Text Color Background Color

Background Shape Text order

----- Animation Section -----

Animation file Animation time

Animation order

----- Picture Section -----

Picture file Picture position

Picture size Picture order

Figure G.4: Output Object

can be chosen among: *Nature, City, Home, Work, School*. For the composition of a scene one must supply the parameters: *name, size of the scene, position, background, number of characters*, and, for each character, the *type, the colour of the hair, the colour of the skin, the humour, the position, and the balloon contents* (optionally the contents of the balloon can be said). An example of the form for the Show-Scene Object is shown in Figure G.5.

G.6 Yes-or-No Object

This object poses the student a “yes-or-no” question. Its parameters are the *Question, the Size of the prompt, the Position, the Colour of the letters, the Background colour, the Buzzer to be used, the maximum Waiting time and the Default answer* (assumed if the waiting time expires). The answer is given through the mouse. After its execution the reserved variable *Yes* will have either the value *true* or *false*, and can be further consulted. The optional fields *Yes Branch* and *No Branch* indicate the next step to be executed depending on the answer. The form is exemplified in Figure G.6.

Literacy Teacher Authoring Tool

Enter Show Scene Object Fields:

Id Number	<input type="text" value="08Q"/>	Position	<input type="text" value="centre"/>
Cartoon size	<input type="text" value="normal"/>	Background color	<input type="text" value="yellow"/>
Background scene	<input type="text" value="nature"/>	Numb of characters	<input type="text" value="2"/>
First character	<input type="text" value="man serious left dark white"/>	First balloon	<input type="text" value="Let's try it"/>
Second character	<input type="text" value="girl happy right blond white"/>	Second balloon	<input type="text" value="Ok"/>
Third character	<input type="text"/>	Third balloon	<input type="text"/>
Count var	<input type="text"/>		
<input type="button" value="Give up"/>			

Figure G.5: Show-Scene Object

Literacy Teacher Authoring Tool

Enter Yes-or-No Fields:

Id Number	<input type="text" value="140"/>	Question	<input type="text" value="Would you like to try it again?"/>
Prompt size	<input type="text" value="normal"/>	Position	<input type="text" value="least"/>
Letter color	<input type="text" value="red"/>	Background color	<input type="text" value="yellow"/>
Timbre	<input type="text" value="female"/>	Prompt buzz	<input type="text" value="ring"/>
Waiting time	<input type="text" value="20"/>	Yes branch	<input type="text" value="30"/>
Yes counter	<input type="text"/>	No branch	<input type="text"/>
No counter	<input type="text"/>	Default answer	<input type="text" value="no"/>
<input type="button" value="Give up"/>			

Figure G.6: Yes-or-No Object

G.7 Choice Object

This object is used to place a set of *buttons* on the computer screen, from which the student is required to choose one. In the definitive version the set of buttons will be represented by icons, tags and speech. If the mouse rests more than one second over a button, one tag should appear beside it and its content should be spoken aloud to the user. If the student clicks on a button it will be considered as his or her choice and its value will be returned in the reserved variable *BarElement*. The operation is considered done when the student clicks the *done button* or the *timeout* expires, in which case the variable *BarTimeout* is made *true*. The parameters of Choice Object are:

1. the *list of elements* to be buttoned represented by a *set identifier*.
2. the *list of icons* related with the *list of elements*.
3. the *size* of the representation of the element bar: *Small, Medium* or *Large*.
4. the *maximum time* allowed for choosing the elements.
5. the *text* to be said when the buttons appear.
6. the *timbre* of the voice: male or female.
7. the *position* of the elements.
8. the *colour* of the letters.
9. the name of the *picture* to be shown when the step is being executed.
10. the kind of the *buzzer* to be used: *Ring, Bell, Bip*.

An example of the form for the *Choice object* is presented in Figure G.7.

G.8 If Object

This object is used to cause a jump to one of two possible continuing steps in the activity depending on the evaluation of a *logical expression*. No input or output is required or produced. The logical expression must be formulated with reserved or user defined variables following specific construction rules. The *Yes Branch, Yes Counter, No Branch* and *No Counter* work just like the fields of the same name in the Yes-or-No Object. If the *expression* is empty the value *Yes* is assumed, so

Literacy Teacher Authoring Tool

Enter Choice Fields:

Id Number	<input type="text" value="90"/>	Element list	<input type="text" value="rbuttons"/>
Icon list	<input type="text" value="standard"/>	Pict list	<input type="text" value="I"/>
Function	<input type="text" value="normal"/>	Time out	<input type="text" value="40"/>
What to say	<input checked="" type="checkbox"/> Which of these sounds is containe	Timbre	<input type="text" value="male"/>
Position	<input type="text" value="centre"/>	Letter color	<input type="text" value="blue"/>
Big Button	<input type="text" value="myword"/>	Buzz type	<input type="text" value="bip"/>
Count var	<input type="text" value="I"/>		
<input type="button" value="Give up"/>			

Figure G.7: Choice Object

Literacy Teacher Authoring Tool

Enter IF Fields:

Id Number	<input type="text" value="90"/>	Expression	<input type="text" value="Phonemein BarElement myword"/>
Yes branch	<input type="text" value="J 20"/>	Yes counter	<input type="text" value="ok"/>
No branch	<input type="text" value="I"/>	No counter	<input type="text" value="notok"/>
<input type="button" value="Give up"/>			

Figure G.8: If Object

the *activity* will proceed to the step indicated at the *Yes Branch*. An example of the form is shown in Figure G.8.

The *logical expression* available so far in the If Object are:

- Yes=True or Yes=False

To test the value of the variable Yes. This variable is automatically set at the execution of the Yes-or-No object.

- Phonemein—BarElement—{variable}

To verify whether the phoneme just clicked by the student is in the word contained in the indicated variable.

- SubString—BarElement—{variable}

To verify whether the letter just clicked by the student is in the word contained in the indicated variable.

- {variable} {relation} {variable}

Variable can be either any variable declared in the program or the reserved word BarElement, representing the element last clicked. *Relation* can be “=” or “<”. The result is *true* or *false* depending on the evaluation of the expression.

G.9 Jump Object

The normal execution of the steps of an activity is following the sequence of declaration of the steps. The *Jump* object is used to alter the sequence of execution of the steps of the activity, causing the unconditional jump to the the step which number is indicated in the field *Jump to*. Figure G.9 shows an example of the Jump Object.

G.10 Set-Var Object

This object allows the definition of working *variables* and the attribution of variable values using *expressions* involving *pre-defined functions*. In the example of Figure G.10, the variable *myword* will receive the value of the first word of the list *wlist*, that will lose its first element thereafter. The system will assist the user in the use of pre-defined functions through a help button. The result of the *expression* can be either a number or an alphanumeric string.

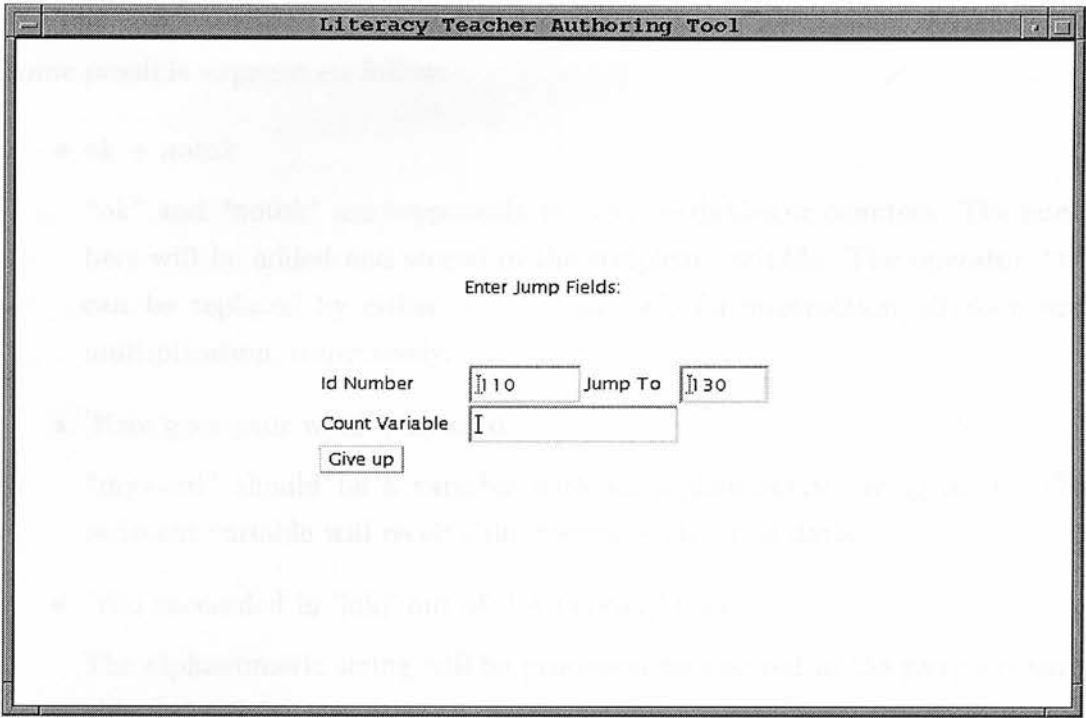


Figure G.9: Jump Object



Figure G.10: SetVar Object

The system will check the validity of the proposed expression. Examples of some possible expressions follow:

- `ok + notok`

“ok” and “notok” are supposedly numeric variables or counters. The numbers will be added and stored in the recipient variable. The operator “+” can be replaced by either “-”, “/” or “*”, for subtraction, division and multiplication, respectively.

- `'Here goes your word'+myword`

“myword” should be a variable with an alphanumeric string in it. The recipient variable will receive the complete string of data.

- `'You succeeded in '|ok|' out of '|ok+notok|' tries'`

The alphanumeric string will be processed and stored in the recipient variable.

- `TakeSubString|myword|start|quant`

The recipient variable be provide with a number of letters indicated in *quant* from the variable “myword”, starting in the position indicated in *start*.

- `TakeFirst|setname`

The recipient variable will receive the first element of the set indicated. The element will be removed from the set.

- `getsuperword(< superset >)`

The recipient variable will have the name of the original variable plus the suffix *Sup*, and will contain a word of the indicated *superset* that phonologically contains the original word.

G.11 Make-Set Object

The system contains some basic *sets* needed for the literacy teaching process, as the *vocabulary*, the *phonemes* used in the language, the *letters*, the *vowels*, the *consonants*, the *numbers*, and so on. The *object Make-Set* is used to generate different new sets from the basic ones. The form for the object is exemplified below. The fields *Step Id* and *Counter variable* have their usual meaning. *SetName* gives the name of the set to be created in the step. The field *TotElements* defines the maximum number of elements the set can have. The field *Source* indicates

the origin set for the operation. In order to generate a set with this command, it can be used with any of the basic sets available for the user (*Words*, *VowelsUp*, *ConsonantsUp*, etc), any combination of these, sets already constructed in the activity or even composed sets, i.e., sets of elements within parenthesis, separated by commas. The *Selection Method* is the method to be used to select the elements (for example: *PickUp* would select the elements randomly, the absence of declaration will select in order of appearance). The form is illustrated in Figure G.11.

Enter Make Set Fields:

Id Number	20	Setname	\$list
Total of elements	10	Source set	Words
Selection	\$SelectWords:Kind 1.1.1		Count var
			1

Figure G.11: MakeSet Object

The options of expressions for the Make-Set object are:

- **SelectWords:Condition**

This option is used for selecting words with some particular characteristics out of an original set to the destination set. The condition can be:

- “BarElement in ListElement”

Means that a word must contain the letter present in the *BarElement* in order to be in the resultant set. (*BarElement* is the button clicked in a previous Match or Choice object.)

- “BarPhoneme in ListElement”

Means that a word must contain the phoneme present in the *BarElement* in order to be in the resultant set.

- “Kind:< *kind* >”

To indicate the kind of phoneme that must be in the words or the complexity of the words that should figure in the resultant set. The kind of phoneme is either the hierarchical number of the curriculum item or a integer number from 1 to 5 to represent the complexity of the word.

- pickUp

To select a random sample of *quantity* elements from the source set and store it in the resultant set.

- random

To generate a resultant set which has the same elements of the original one, but in a random order.

- takesetphoneme(< *word* >:Start:Quantity)

To generate a resultant set which contains a set of *quantity* phonemes from the *source* variable, starting from the position *start*. If the value of *Start* is -1 then selection of the phonemes is random.

- Replace:< *element1* >:< *element2* >

To generate a resultant set equal to the source set except for all the elements with the value equal to the one in *element1* will have the value of *element2*.

- Exclude:< *element* >

The resultant set will have all the elements of the source set except the ones with the value of *element*.

- SetSuperSet

For each word in the original set (the one in the field *set name*) the *source* set will be examined word by word. If a word is contained in other words of the *source* set (e.g. *word* is contained in *sword*) it will be kept in the resultant set, which will have the name of the original set plus the string *Res*. The *super-words*, that is, the ones that contains the original word will be appended to the super-set, with the end *Sup*. Put another way, for the

original set *mywords* the resultant set with the words which have super-words will be called *mywordsRes* and the set of super-words will be named *mywordsSup*.

- SuperPhonemes:< *subword* >:< *superword* >

This expression will generate a set of phonemes to be named as indicated in *set name* as explained below. The *subword* is a word that is part of another bigger word named the *superword*; the resultant set will represent an *equation of phonemes*, explaining the formation of the bigger word using the smaller one as a part of it. For example, if the *subword* is *ring* and the *superword* is *bring*, the result would be a set with the following characters:
b + r i n g = b r i n g.

- getspell:< *word* >:< *amount* >

This expression will generate a set of words that are spelled with alternative spellings for the proposed *word*. The first element of the set is the right spelling, and the quantity of alternatives required in *amount* is provided. The name of the resultant set must come in the field *Set Name* of the object.

G.12 Match Object

This object is intended to provide a way to interactively match two *lists*. When executed it will present a list with unordered values and the student will try to put it in the right order in the empty list below the first one. An example of the form to be filled to declare the object is shown in Figure G.12.

The description of the form fields follows:

1. *Button Type* - to determine whether the *buttons* are *phonetic* or based in *letter names*.
2. *Word* - to identify the word and the corresponding *picture* to be shown.
3. *Button List* - a *list* with the elements to appear on the row of *buttons*.
4. *Matching List* - a *list* to be considered as the elements to be matched.
5. *Object Prompt* - a sentence to be said when the object is first entered and each time the interrogation mark is pressed.
6. *Action Prompt* - a sentence to be said before a *button* should be pressed.

Literacy Teacher Authoring Tool

Enter Match Fields:

Id Number	65	Button Type	Phonetic
Word/picture	myword	Button list	phoset
Match list	phosetout	Object prompt	What are the sounds of "+myword+" ?
Action prompt	What is this sound?"	Fail Msg	this sound is not here – try another"
Success Msg	It's all right!"	Timeout	80
Count var	1	Last msg	Ok, you matched everything"

Give up

Figure G.12: Match Object

7. *Fail message* - a sentence to be said in case of failure to match elements.
8. *Success message* - a sentence to be said when two elements are successfully matched.
9. *Timeout* - the total waiting time before quitting the object if there is no move.
10. *Count* - to associate a *count variable* to count how many times the object has been entered during the activity.

Appendix H

Teaching Strategy for Short Phonix

The complete TSDL code (Teaching Strategy Definition Language) for Short Phonix follows:

```
if learning (Fat Cat Sounds) greater than enough
  then disable lesson(Fat Cat Sounds).
if exercising(Fat Cat Sounds) greater than enough
  then disable exercise(Fat Cat Sounds).
if testing(Fat Cat Sounds) greater than enough
  then disable lesson(Fat Cat Sounds),
    disable example (Fat Cat Sounds),
    disable exercise(Fat Cat Sounds).
if testing(Fat Cat Sounds) greater than good
  then disable test(Fat Cat Sounds).
if testing(Fat Cat Sounds) less than enough then
  enable exercise(Fat Cat Sounds),
  enable lesson(Bug On Jug Sounds),
  enable example(Bug On Jug Sounds),
  enable test(Bug On Jug Sounds).
if exercising (Fat Cat Sounds) less than enough
  then enable example(Fat Cat Sounds),
    enable lesson(Fat Cat Sounds).

if learning (Bug On Jug Sounds) greater than enough
  then disable lesson(Bug On Jug Sounds).
if exercising(Bug On Jug Sounds) greater than enough
  then disable exercise(Bug On Jug Sounds).
```

if testing(Bug On Jug Sounds) greater than enough
 then disable lesson(Bug On Jug Sounds),
 disable example (Bug On Jug Sounds),
 disable exercise(Bug On Jug Sounds).
 if testing(Bug On Jug Sounds) greater than good
 then disable test(Bug On Jug Sounds).
 if testing(Bug On Jug Sounds) less than enough
 then enable exercise(Bug On Jug Sounds),
 enable test(Bug On Jug Sounds),
 enable exercise(Fat Cat Sounds).
 if exercising (Bug On Jug Sounds) less than enough
 then enable example(Bug On Jug Sounds).
 if learning (Ben Bon Sounds) greater than enough
 then disable lesson(Ben Bon Sounds).
 if exercising(Ben Bon Sounds) greater than enough
 then disable exercise(Ben Bon Sounds).
 if testing(Ben Bon Sounds) greater than enough
 then disable lesson(Ben Bon Sounds),
 disable example (Ben Bon Sounds),
 disable exercise(Ben Bon Sounds).
 if testing(Ben Bon Sounds) greater than good
 then disable test(Ben Bon Sounds).
 if testing(Ben Bon Sounds) less than enough
 then enable exercise(Ben Bon Sounds),
 enable test(Ben Bon Sounds),
 enable test(Bug On Jug Sounds),
 enable exercise(Bug On Jug Sounds).
 if exercising (Ben Bon Sounds) less than enough
 then enable example(Ben Bon Sounds),
 enable exercise(Ben Bon Sounds),
 enable lesson(Ben Bon Sounds),
 enable test(Fat Cat Sounds),
 enable exercise(Fat Cat Sounds),
 enable test(Ben Bon Sounds).
 if learning (V C C Words) greater than enough
 then disable lesson(V C C Words).

if exercising(V C C Words) greater than enough
then disable exercise(V C C Words).

if testing(V C C Words) greater than enough
then disable lesson(V C C Words),
disable example (V C C Words),
disable exercise(V C C Words).

if testing(V C C Words) greater than good
then disable test(V C C Words).

if testing(V C C Words) less than enough
then enable exercise(V C C Words),
enable test(V C C Words),
enable test(Ben Bon Sounds),
enable exercise(Ben Bon Sounds).

if exercising (V C C Words) less than enough
then enable example(V C C Words),
enable exercise(V C C Words),
enable lesson(V C C Words),
enable test(Bug On Jug Sounds),
enable exercise(Bug On Jug Sounds),
enable test(V C C Words).

if learning (C V C C Words) greater than enough
then disable lesson(C V C C Words).

if exercising(C V C C Words) greater than enough
then disable exercise(C V C C Words).

if testing(C V C C Words) greater than enough
then disable lesson(C V C C Words),
disable example (C V C C Words),
disable exercise(C V C C Words).

if testing(C V C C Words) greater than good
then disable test(C V C C Words).

if testing(C V C C Words) less than enough
then enable exercise(C V C C Words),
enable test(C V C C Words),
enable test(V C C Words),
enable exercise(V C C Words).

if exercising (C V C C Words) less than enough

then enable example(C V C C Words),
enable exercise(C V C C Words),
enable lesson(C V C C Words),
enable test(C V C C Words).

if learning (C C V C Words) greater than enough
then disable lesson(C C V C Words).

if exercising(C C V C Words) greater than enough
then disable exercise(C C V C Words).

if testing(C C V C Words) greater than enough
then disable lesson(C C V C Words),
disable example (C C V C Words).

if testing(C C V C Words) less than enough
then enable exercise(C C V C Words),
enable test(C C V C Words),
enable test(V C C Words),
enable exercise(V C C Words).

if exercising (C C V C Words) less than enough
then enable example(C C V C Words),
enable exercise(C C V C Words),
enable lesson(C C V C Words),
enable test(V C C Words),
enable exercise(V C C Words),
enable test(C C V C Words).

suspend winner(15).

if exercising (C C V C Words) greater than enough or
testing (C C V C Words) greater than enough
then suspend winner(5).

expertise(Bug On Jug Sounds)=90.

expertise(Auditory Processing)=80.

expertise(Fat Cat Sounds, Blending)=70.

Appendix I

Tools

I.1 Vocabulary Tools

In order to obtain and deal with the vocabularies required in this research, a total of 68 Java programs have been written, totaling 16233 lines of code. All these programs are available in the directory `lts/voc` of the project. The relation of the main programs with a brief description of their functions and interfaces follows.

- *analyse1.java* - program used to interpret the results of the vocabulary survey. It analyses the result of the survey for a particular vocabulary and a specific population (aliens or native speakers).

Format:

```
java analyse1 {survey file} {word file name} {native / alien}
```

- *ConvCollins.java* - this program was used to convert the Collins English Dictionary into the format of the Collins Cobuild Dictionary.

Format:

```
java ConvCollins {ced file} {output dictionary}
```

- *ConvMCR.java* - this program was used to convert the Medical Council Research Dictionary into the format of the Collins Cobuild Dictionary.

Format:

```
java ConvMCR {mcr file} {output dictionary}
```

- *Dialg.java* - it is a Java Applet available in the Internet through the URL <http://fintray.dai.ed.ac.uk:8811/words.html> , which provides the interface for the survey on the thematic vocabularies, as described in Appendix B.

- *Freq.java* - this program provides the analysis of frequency of a dictionary, in terms of minimum, average, maximum frequencies, and other information useful to decide about the frequencies of cut for vocabularies.

Format:

```
java Freq {frequency file} {maximum words}
```

- *FreqWords.java* - program used to add the frequencies of the words in the text file indicated by the second parameter to the frequency file in the first parameter, giving the frequency file of the third parameter. The expected maximum of words must come in the forth parameter.

Format:

```
java FreqWords {old freq file} {text file} {new freq file} {max words}
```

- *Ktree.java* - This is the compiler of the knowledge tree. It analyses the knowledge tree and distributes the thematic vocabulary according to the corresponding curriculum.

Format:

```
java Ktree {vocabulary file}
```

Note: the vocabulary file must have the .pho extension and must have been generated by either the PickupOW.java program or the PickupW.java program.

- *Lookword.java* - This program is the responsible for generating a dictionary with the phonemic coding of the words from an uncoded one, by consulting the Festival System for each word of the input dictionary.

Format:

```
java Lookword {uncoded dictionary} {coded dictionary}
```

- *ManIE.java* - This program allows the manual inclusion and exclusion of words in a particularly dictionary.

Format:

```
java ManIE {input dictionary} {base dictionary} {command file}
           {output dictionary}
```

- *Merge.java* - program used to merge two input dictionaries producing a single one with entries from both input dictionaries.

Format:

```
java Merge {input dict 1} {input dict 2} {output dict}
```

- *MergeFreq.java* - program used to include the information about the frequency of occurrence of the words into the dictionary.

Format:

```
java MergeFreq {dict file} {frequency file} {new dict} {maxwords}
```

- *PickupOW.java* - program designed to select the words according to constrictions on frequency, length of the word in number of letters and in number of phonemes, and age of acquisition. It produces the vocabulary in the format accepted by the knowledge tree compiler.

Format:

```
java PickupOW {infile} {outfile} {maxletters} {maxpho} {minfrq}
               {maxaoa}
```

- *PickupW.java* - program similar to the previous one, except for it additionally inserts the derivative words into the vocabulary produced. The format for calling is identical to the previous one.
- *Selector.java* - program used to promote the thematic selection, producing a new version of the input dictionary only with the words selected.

Format:

```
java Selector {input dictionary} {output dictionary} {selection file}
              {sample rate} {required seeds} {strict/relax} {lower/all}
              {with/without} {notwords/onlyseeds} {andwords/ignoreand}
```

Note: *sample rate* is the rate of words that should be picked up in the selection process: 1 means pickup all the input words, 2 means pickup one in each two, and so on; *required seeds* is the number of seeds that must appear in the description of a word so it is considered eligible for appearing in the output dictionary; *strict/relax* is related to the allowance for repeated seeds to be counted as a new seed; *lower/all* is related to permitting the use words with the first letter in uppercase; *notwords/onlyseeds* is related to the use of the “not-words” in the selection; *andwords/ignoreand* is related to the verification of the “and-words” in the selection process.

I.2 Authoring Tool

The development of the Authoring Tool involved the coding of 10,425 lines of Java programs, distributed in 42 source files.

The main interface is the Author.java program, which can be called from the command line, either in Unix, Linux or PC platform with the following format:

```
java Author <source file name>
```

Source file name is the name of an activity of a programme for teaching to read. If it does not exist yet the program will create it and keep it in the file name provided to be used or edited later. It is recommended that the name of the source file should be composed of two letters identifying the programme in use (e.g. “pg” for Phono-Graphix, “jp” for Jolly Phonix, “tr” for THRASS) separated by a dash from a sequential number to identify the activity. For example, an activity could be named “jp-1”. In this case the source file for this activity will be named “jp-1.src” and the code file with the code to be run by the lesson interpreter will be named “jp-1.cod”.

The Author.java program is interactive and has some basic facilities to help the development of the activities. The code file is obtained when the operator (the expert teacher) commands the “compilation” of the activity. The compilation checks the syntax of the parameters of the objects and the coherence of the jump, if and yes-or-no objects used in the activities, but cannot guarantee that the activity is consistent and does the job it is supposed to do. The designer of the activity must test it with the lesson interpreter, described in next section.

I.3 Lesson Interpretation

The 45 programs developed to interpret the activities of the system compose a Java Applet and consist of 11,121 lines of code.

They can be started in some recent Internet browsers, but for efficiency and compatibility with the version used for the development (JDK 1.2) it is recommendable to run it through the AppletViewer JDK1.2.01 program, provided by Sun Microsystems.

The interface for triggering the Applet with the AppletViewer program is
appletviewer tmaplet.html

The programs and everything else needed for running them are available in the directory /hame/lts/develop at the computer complex in the South Bridge site of the Division of Informatics.

The program is conceived to interact with illiterate people, giving all the main instructions verbally through the Festival System speech synthesiser.

I.4 Software Summary

In total, for the implementation of the system related to this research, a total of 37,779 lines of Java code, distributed in 155 source files have been written.