



# 3

## **DEVELOPMENT OF ESP COURSEWARE VIA THE WWW: A COMPUTATIONAL LINGUISTICS ANALYSIS APPROACH**

**SARIMAH SHAMSUDIN  
MASPUTERIAH HAMZAH  
WAN FARA ADLINA WAN MANSOR  
ZAIRUS NORSIAH AZAHAR  
MOHAMMAD ISMAIL**



### **INTRODUCTION**

English Language programmes that are being designed to meet specific needs of the learners would benefit not only the learners themselves but also their stakeholders. An example of an English Language programme that has been designed to meet the English for Specific Purpose (ESP) needs of Civil Engineering (CE) students of Universiti Teknologi Malaysia (UTM) is the programme called English for Civil Engineering or ECE. This programme was a three-semester programme that has been planned through very close collaboration between the faculty of Civil Engineering (FCE) and the Department of Modern Languages (DML).

The ECE programme ‘rests on the principle that language learning is and should be part of the mainstream CE programme (DML of UTM, 1996). Its ultimate aim was to be able to assist CE students in using English for executing their academic work

assigned by their faculty, the FCE. In general, the ECE programme was mainly intended to 'serve as the foundation of the long term objective of enhancing the marketability of CE graduates by integrating language training into the overall training plans of the FCE' (DML of UTM, 1996). The ECE programme offered three different courses which are UHB1312, UHB1322, and UHB2332, to the CE students. Although it is subdivided into three courses, that is, one course per semester, the courses should be regarded as a continuous and coherent training scheme. Hence, in general, the ECE programme intends to assist the CE students in extracting information for written and oral communication for both academic and professional purposes.

In relation to the second course of the programme, UHB1322, in every phase of the course, the students are required to perform two parallel and complementary types of activities – the core activities and the support activities (also called pedagogic activities). The core activities are activities which would train the CE students to extract information from texts related to Civil Engineering Materials (one of the core topics in Civil Engineering), that are written in English, and consequently, make notes from them. These notes would then be used to write essays and reports related to Civil Engineering Materials (CEM). As for the pedagogic activities of the UHB1322 course, they are activities which lead to and support the core activities of the course. All pedagogic activities such as the reading of texts, vocabulary/glossary work and language exercises, will be related to the topic on Civil Engineering Materials.

In relation to the above pedagogic activities, they are very time-consuming as they require several classroom sessions with the language instructor. They also require frequent referencing to an extremely limited number of monolingual and bilingual Civil Engineering dictionaries as well as English CEM textbooks in the UTM library. In order for the CE students to perform the pedagogic activities much more efficiently and efficaciously, a unique self-

access courseware that can be accessed via the WWW (World Wide Web, Web or Internet) using the computational linguistic analysis approach, has been designed for the ECE programme in UTM, which will be referred to as 'UNITEKMA ECE Courseware'. This chapter aims to describe the process of producing the prototype of the above courseware in which its lessons focus both on content words and language structures taken from texts on Civil Engineering Materials. The following sections of this chapter will discuss related literature to the design of the courseware.

## **CORPUS**

In the context of modern linguistics, a corpus (plural form: corpora) is regarded as: 'a collection of texts, of the written or spoken word, which is stored and processed on computer for the purpose of linguistic research' (Renouf in Sinclair 1987:1). In general, corpora can provide language instructors and analysts with insights into typical features of their linguistic structures and language use (Renouf in Sinclair 1987; Skelton & Willis 1993). Nevertheless, McEnery and Wilson (1996) refine the usage of corpus further by suggesting that corpora are sources of empirical data which can serve two major purposes.

Firstly, corpora can be used to develop computational tools for processing natural language. A few examples of those computational tools are part-of-speech taggers, automated lexicon generators, automated parsing systems and machine translation systems (McEnery & Wilson 1996). Secondly, corpora can also be used for studying language itself, such as in speech research, lexical and grammatical studies, the study of semantics, pragmatics and discourse analysis and the teaching of languages and linguistics (McEnery and Wilson 1996). Hence, in relation to language teaching, a more specific

usage of corpora, particularly the ‘specialised corpora’ (such as the Guangzhou Petroleum Corpus and the HKUST Computer Science Corpus mentioned earlier) is for teaching English for specific purposes whereby according to McEnery and Wilson (1996:105) ‘such corpora can be used to provide many kinds of domain-specific material for language learning, including quantitative accounts of vocabulary and usage which address the specific needs of students in a particular domain more directly than those taken from more general language corpora’.

As mentioned earlier in this chapter, the ESP needs of the CE students in UTM is to use English for executing their academic work, assigned by their faculty, the FCE. For example, they have to extract information from their content-area textbooks written in English, for academic purposes. Hence in order to assist the students in achieving this objective, the web-based ECE courseware or ‘UNITEKMA ECE Courseware’ was developed to use items (common terminology and samples of their common phrases and sentences) obtained through the creation and computational linguistics analysis of a ‘specialised corpus’ taken from authentic text of the students’ faculty-assigned CEM textbooks. In addition, items from those texts are also chosen because the CE students have already acquired the schemata (as defined by Carrell, Devine and Eskey 1988) for those texts in their first language through their CEM (subject-area) course. Hence, the schema theory will be discussed further in the next sub-section.

## **CORPUS CREATION**

In order to develop the ‘UNITEKMA ECE Courseware’ which is based on corpus analysis, a specialised corpus has to be created first. However, in general, the process of creating a corpus is divided into three stages. First is the design stage which involves the identification

or selection of texts based on set criteria. Next, is the acquisition of these texts, which might involve request of permission for copyright, and the final stage is the processing of those texts into computer-readable form (Renouf in Sinclair 1987; Sinclair 1991; Roe 1995).

During the design stage of corpus creation, the first thing to consider is its aim (Sinclair 1991:13). In the case of a specialised corpus, if it is intended for teaching English for Specific Purposes (ESP), its aims are clearly defined that is ‘to provide many kinds of domain – specific materials for language learning’ (McEnery and Wilson 1996:105) and to assist the learners in achieving the specific objective in learning English.

After the aims for creating a particular corpus has been determined, the next step at this stage is the identification or selection of the corpus text based on set criteria. In relation to the specialised corpus since it is a ‘domain-specific’ corpus, the text for the corpus should obviously be selected from the specified domain of the learners’ target discourse community.

The next major stage for corpus creation is the acquisition stage, that is, the collection of all relevant texts which have been identified for creating the corpus. However, as a note of caution, this particular stage might involve requesting permission to reproduce (in computer-readable form) the selected text for the corpus, if the publisher holds a copyright on the text.

The final stage of corpus creation is the processing stage, which involves two major steps. Firstly, if the selected text for the corpus cannot be acquired initially, in computer readable form such as those obtained from books, magazines or journals, they have to be scanned or keyboarded. Nevertheless, sometimes both methods are being used at this stage. Secondly, is to edit the text for errors.

The process of scanning involves the use of a computer and OCR or Optical Character Recognition software and an optical scanning machine (commonly known as a scanner) for capturing the image on the text that is being scanned and converting it into

computer readable form. Subsequently, the converted version is saved onto the hard disk of the computer and the floppy disk in the form of text files. As for keyboarding, this process only requires a computer and word processing software. It is simply executed by 'keying in' the content of the selected text onto the computer, using word processing software such as Microsoft Word or WordPerfect and then saving it onto the hard disk of the computer and a floppy disk as text files. Hence, after the corpus has been developed through the process of scanning and/or keyboarding, it will be compared with the original text and edited for errors. This is normally done both manually and using a 'spelling checker' (available on most word processing software).

In general, the above procedure briefly explains the steps that should be followed in order to process selected text for a particular corpus that are not readily available in computer-readable form. However, if the selected texts for the corpus are originally in computer-readable form, such as those which are available on the WWW, they are simply copied and then saved onto the hard disk of the computer or a floppy disk as text files. Hence, three of the most common Internet browsers that are used to surf the WWW and download information are, Netscape Navigator, Mosaic and Internet Explorer.

## **COMPUTATIONAL LINGUISTICS ANALYSIS OF THE CORPUS**

When the specialised corpus has been created, the next step in the development of the 'UNITEKMA ECE Courseware' based on corpus analysis is to select the most suitable computational tool for executing a computational linguistics analysis of the corpus. Nonetheless, the computational tools chosen will depend entirely

on the objectives of the self-access material that is intended for the learners. Therefore, since the 'UNITEKMA ECE Courseware' will be providing the CE students of UTM with a component on glossary, 'quantitative accounts of vocabulary and usage which address' (McEnery and Wilson 1996:105) the ESP needs of the students will have to be acquired from the computational linguistics analysis of the specialised corpus. Hence, the most suitable type of computing tool which is capable of accomplishing the above task is a concordance software such as 'MonoConc'.

## **SCHEMA THEORY**

The other reason for selecting (through corpus creation and analysis) items from the CE students' faculty-assigned CEM textbooks for the 'UNITEKMA ECE Courseware' that is integrated into the ECE programme in UTM is because the students have already obtained the schemata (as defined by Carrell et. al 1988) for those texts in their first language, through their CEM (subject area) course. Subsequently, the selected items for the courseware are organised in the form of hypertext, in which the schema theory underpins the general structure of the hypertext.

The schema theory assumes that the human memory system consists of numerous 'packets of knowledge' (schemata) and that 'each packet specifies a configuration of other packets (sub-schemata) which represent the constituents of the schema' (Rumelhart and Norman 1985:37). In relation to the understanding of language, Greene (1986:35) points out that the schema theory 'represent the general knowledge which aids the understanding of conversations and texts, as well as of real-life events'. Simply put, schema theory is knowledge that has been stored in memory which assists in the interpretation of new information by enabling the new information to be a part of the stored knowledge (Anderson and Pearson 1988).

In second language reading (SLR), this knowledge is referred to by Carrell (1988:2) as ‘background knowledge’ or schemata. She further suggests that there are two different types of schemata that are useful for SLR: ‘content schemata’ (‘background knowledge of the content area of the text’) and ‘formal schemata’ (‘background knowledge of the rhetorical organisational structure of the text’) (Carrell 1988:104). Based on various research done on SLR, it has been suggested that both types of schemata are essential to ensure efficient and effective reading as well as comprehension of texts written in the second language (Carrell and Eisterhold 1983; Carrell and Wallace 1983; Carrell 1983, 1984, 1985). Thus, with reference to the ‘UNITEKMA ECE Courseware’ proposed for the ECE programme in UTM, items for the courseware are selected from the CE students’ faculty-assigned CEM texts because they have already acquired the schemata of those texts in their first language through their CEM (subject-area) course. These schemata should then enable the students to comprehend all the selected items in the courseware effectively and efficiently, and consequently, assist them in performing the pedagogic activities, followed by the core activities of the ECE programme with minimal difficulty.

### **COMPUTER-ASSISTED LANGUAGE LEARNING (CALL) VIA THE WORLD WIDE WEB (WWW, WEB or INTERNET)**

The concept of CALL via the WWW refers to the initial idea that the corpus-based ‘UNITEKMA ECE Courseware’ designed in the form of ‘hypertext’, ‘Web content’ or ‘Web document’ (a collection of ‘web pages’ or on-line hyperlinked documents using HTML or hypertext markup language), will be installed onto a chosen ‘Web server’ (a program that sits on a computer connected to the Internet and waits for requests from ‘Web browsers’ for a copy of its ‘Web



content') connected to the WWW. Its location will therefore, be referred to the 'Web server's' 'home page address' (also known as uniform resource locator or URL; address where a file or any other resource is located on the WWW). Then, the hyperlinked 'UNITEKMA ECE Courseware' will be made accessible to the CE students of UTM, for language learning purposes, from the 'Web server', using computers that are linked to the WWW. However, retrieval of the hyperlinked "UNITEKMA ECE Courseware" can only be done by locating the personal 'home page' ('Web page' that act as a front door to the 'Web content') of the material on the WWW. By specifying the 'home page address' or location of the material on the WWW, students can access the content using a 'Web browser' (a program that can retrieve any sort of information from just about any computer connected to the Internet and display it in a recognisable format) that has been installed or connected onto the computers.

Generally, the process of designing or creating a 'Web document' or 'Web content' to be made available on the WWW involves five major stages (Horton, Taylor, Ignacio and Hoft. 1996; Tatters 1996). The first stage is called the analysis stage, followed by the designing stage. Next, is the development stage, followed by the implementation stage and finally the evaluation stage of the prototype 'Web document' or its first complete version on the WWW.

## **METHODOLOGY**

The process of creating the 'UNITEKMA ECE Courseware' involves five stages (ADDIE Model):

1. Analysing the ECE program and CE academic requirement.
2. Designing the conceptual framework based on computational

linguistic analysis of CECM ‘specialised corpus’ (Tribble & Jones, 1989; Skelton & Willis, 1993; Jackson & Dhir, 1988).

3. Developing the web-based courseware.
4. Implementing/piloting the courseware.
5. Evaluating the courseware based on students’ responses.

### **Studying the ECE Programme and Civil Engineering Academic Requirements**

The first step during the analysis stage of the ‘UNITEKMA ECE Courseware’ is establishing the aims or objectives of the courseware, based on the needs analysis of its intended users. However, since the courseware has been proposed to be introduced as a material for the ECE programme in UTM, its aims/objectives were based on the needs analysis (English for Specific Purposes or ESP needs of the CE students) that has been carried out by the DML of UTM for designing the ECE programme. In general, the ESP needs of the CE students is to use English in the course of their academic work. Therefore, the ‘UNITEKMA ECE Courseware’ has been proposed to enable the students to achieve familiarity and understanding of the content area (covered in their mainstream CE programme) and its language structure in the English Language and consequently assist them in executing their faculty-assigned academic work, in a relatively shorter amount of time. The ‘UNITEKMA ECE Courseware’ is generated by and built around the requirements of the CE students’ CEM course (compulsory course for CE students of the Faculty of Civil Engineering).

Identifying the tools needed for creating the courseware is the next important step. The three types of tools essential in the courseware development are the hardware, software and specialised corpora on Civil Engineering material (CEM).

The hardware used are an IBM compatible PC with 32Mb

of RAM and Pentium processor and an OCR (optical character recognition) Scanner, while the software used are: a concordance software called MonoConc, a 'Web browser' – Netscape Communicator, a 'Web page' designer – FrontPage and a word processing software – Microsoft Word.

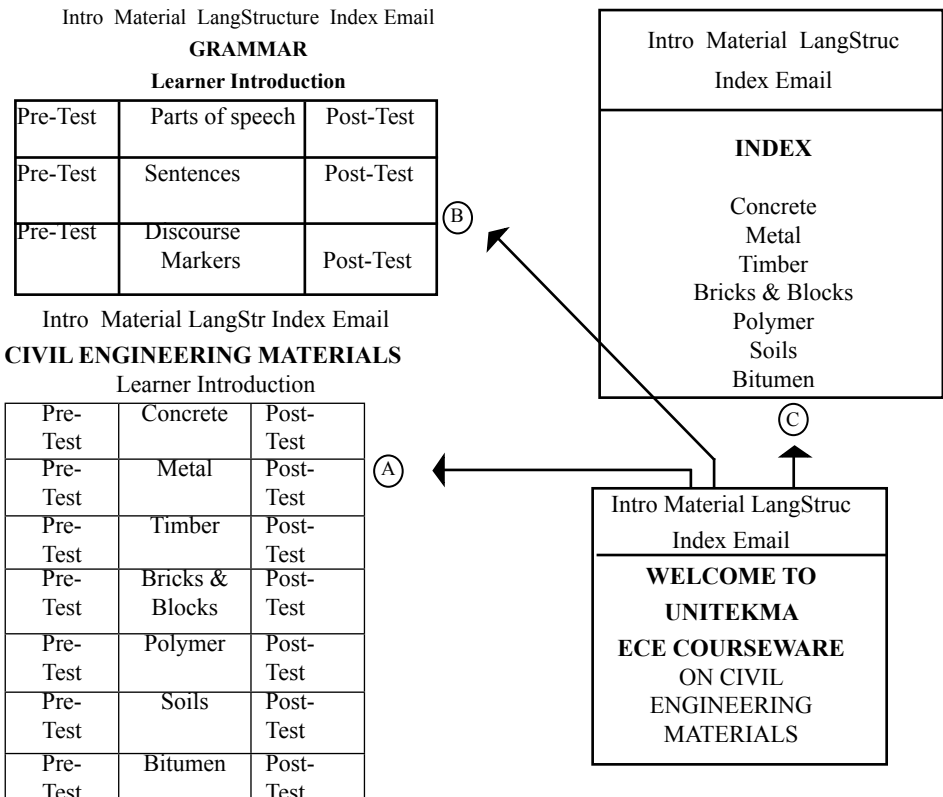
Skelton and Willis (1993) define corpora as specialised corpus that is made up of the language used by the target discourse community, and carefully selected so as to give a balanced and suitably proportional sample of typical speech events and written text in common use. The corpora on CEM is necessary for creating the 'UNITEKMA ECE Courseware' because it will be used mainly to develop the contents ('domain-specific' language learning material as defined by McEnery and Wilson 1996:105) of the courseware.

### **Designing the Structural Framework or Cluster of the 'UNITEKMA ECE COURSEWARE' based on Computational Linguistics Analysis of the CEM Specialized Corpora**

Since the courseware can be accessed on the World Wide Web (WWW) and is intended to assist the CE students in carrying out pedagogic activities, it will provide them with a 'Welcoming Page' 'hyperlinked' to four major sections of the courseware. They are the introductory section (containing the aims, starting level and contents of the courseware), material section, language structure section and index section of the courseware (see Figure 1 for the structural framework or clusters for the 'UNITEKMA ECE Courseware' Home page).

The material section is organised according to the seven different CEM namely Concrete, Metal, Timber, Bricks and Blocks, Polymer, Soils and Bitumen (see Figure 4 for structural framework or clusters for the material section of the courseware). Each material has two sub-sections – glossary and main topics. For the glossary part, a pre-test, a post-test and

three different types of exercises are provided and scores are given for attempting the tests and exercises which are in the form of matching and filling in the blanks. The glossary part also contains the glossary for all conceptual areas or key concepts of every CEM and common terminologies or lexical items for those key concepts that have been identified based on computational linguistics analysis of the CEM specialised corpora. The topics for the 'main topics' subsection have been identified by the subject specialist who teaches the CEM course in the FCE, according to importance. This subsection contains questions related to the selected topics and answers to the questions will be provided. In general the material section helps to familiarise CE students with the key concepts and topics in the area of CEM.



Int Mat LangStruc Indx Email		
<b>INTRODUCTION</b>		
TO		
UNITEKMA	ECE	COURSEWARE
AIM	LEVEL	CONTENT

**Figure 1** – Structural Framework or Clusters for  
‘UNITEKMA ECE Courseware’ Home page

## **Corpus Creation**

The process of creating the corpora on CEM involves three different stages. They are the design stage, acquisition stage and finally, the processing stage.

During the design stage of the CEM corpora, since it is a specialised corpora, the aim of creating the corpora is clearly defined, which is to provide authentic and ‘domain-specific’ material for the ‘UNITEKMA ECE Courseware’, that has been proposed. The texts for creating the corpora have been selected from one of the CEM reference books of the CEM course. The book chosen for creating the corpora on CEM is entitled ‘Civil Engineering Materials’ by Jackson and Dhir (1988). This book has been specifically chosen because there is no translation of the book in Bahasa Malaysia and it is a highly recommended reading material for the CEM course.

Before the process of creating the corpora on CEM can be done, permission for copyright to use the texts of the chosen CEM reference book for corpora creation has to be obtained first from its publisher. Fortunately, during the acquisition stage, permission has been granted by the publisher of the book to scan texts from

the book. The files are then saved as 'text files' (computer-readable form) for computational linguistic analysis purposes.

During the processing stage, which is the last stage of the process involved in creating the corpora on CEM, all the texts from the chosen book have been scanned using an OCR (optical character recognition) software and a scanner and then saved as 'text files'. After that, all the scanned 'textfiles' have been edited for errors using word processing software in comparison with the original texts and then saved onto the hard disk of a computer and several floppy diskettes as back-up copies. Hence, when all these procedures have been completed, the resulting specialised corpora on CEM will be in the form of 'text files'. They are the corpus on Concrete, Metal, Timber, Bricks and Blocks, Polymer, Soils and Bitumen.

### **Computational Linguistics Analysis of the Corpora**

According to Tribble and Jones (1989), concordancing is a form of lexical analysis that locates and gives information about all occurrences of a particular lexical item within a text. Concordance output enables learners to discover not only the meaning but also other information about the keyword like frequency, common collocations, grammatical features, multiple meanings and connotations (see Figures 2 and 3).

Count	Pct	Word
773	2.1456%	concrete
756	2.0984%	is
641	1.7792%	a
479	1.3296%	for
462	1.2824%	be
342	0.9493%	are
342	0.9493%	strength
298	0.8272%	with
286	0.7938%	cement
272	0.7550%	as
238	0.6606%	on
229	0.6356%	by
222	0.6162%	this
217	0.6023%	that
209	0.5801%	it
200	0.5551%	which
198	0.5496%	or
190	0.5274%	can
181	0.5024%	aggregate
159	0.4413%	water
154	0.4275%	may
153	0.4247%	from
150	0.4164%	its
141	0.3914%	used
139	0.3858%	test
132	0.3664%	mix
121	0.3359%	should
120	0.3331%	at
120	0.3331%	content
118	0.3275%	an
114	0.3164%	workability
110	0.3053%	also
103	0.2859%	given
98	0.2720%	figure

1 file in current corpus 36,027 words

Start | Exploring - S:356217a001 | MonoConc for Windo... | Microsoft Word - sjplcon.d

Figure 2 – Frequency List for the Corpus on Concrete

MonoConc for Windows - [Frequency Statistics - [concrete]]			
File Concordance Frequency Display Window Info			
2-Left	1-Left	1-Right	2-Right
57 of	122 of	65 is	27 the
27 strength	70 the	31 strength	23 a
19 the	36 in	23 in	18 is
16 in	17 hardened	22 and	15 in
15 for	13 fresh	18 mix	11 be
13 a	12 a	11 are	11 and
11 properties	9 on	10 The	11 to
11 and	7 given	9 as	10 its
8 to	7 for	9 which	9 are
6 use	6 prestressed	8 with	8 with
6 type	5 When	8 the	8 as
6 within	5 ready-mixed	8 can	8 not
5 or	5 which	7 In	6 for
5 reduction	5 compacted	7 made	6 of
5 effect	4 making	7 to	5 can
4 fully	4 that	7 has	5 The
4 at	4 reinforced	6 for	5 also
4 which	4 mass	6 work	5 this
4 characteristics	4 plastic	6 will	4 proportions
4 used	4 flowing	6 structures	4 both
4 workability	4 to	5 It	4 under
3 pores	4 A	5 surface	3 may
3 that	3 when	4 by	3 shrinkage
3 such	3 as	4 increases	3 general
3 mass	3 cement	3 cover	3 when
3 resistance	3 quality	3 during	3 design
3 ability	3 into	3 particularly	3 that
3 from	3 than	3 creep	3 although
3 before	3 saturated	3 having	3 strength
3 content	3 or	3 at	3 practice
3 durability	3 affect	3 durability	3 subjected
3 poor	3 precast	3 workability	3 properties
3 quality	3 plain	3 This	
3 masses		3 These	
		3 subjected	
		3 Its	

1 file in current corpus 36,027 words

Start | Exploring - S356217a001 | MonoConc for Windo... | Microsoft Word - freqcon...

**Figure 3** – Synoptic Profile for the Key Concept ‘Concrete’  
(Frequency of words which occur before and after the  
key concept ‘Concrete’)



Based on the analysis done using MonoConc, about ten of the most common non-grammar words for each type of material in the order of their frequency are as follows:

- Corpus on Concrete - concrete, strength, cement, aggregate, water, test, mix, content, workability, ratio (see Figure 2 – frequency list).
- Corpus on Metal – steel, stress, metal, material, carbon, strength, alloy, temperature, welding, corrosion.
- Corpus on Timber – timber, section, strength, moisture, cell.
- Corpus on Bricks and Blocks – brick, block, strength, concrete, clay, water, mortar, resistance, calcium, masonry.
- Corpus on Polymer – materials, polymer, fiber, properties, stress, creep.
- Corpus on Soils – water, stress, clay, flow.
- Corpus on Bitumen – mix, test, temperature, aggregate.

All these words have therefore, been used to represent the major conceptual areas or key concepts of the ‘UNITEKMA ECE Courseware’, in particular, its material section. Hence, this is why the structural framework or clusters for the material section of the ‘UNITEKMA ECE Courseware’ is as illustrated in Figure 4 whereas Figure 5 indicates the structural framework or clusters for the language structure and index section of the courseware (Figures 4 and 5 are the extension of the structural framework shown in Figure 1).

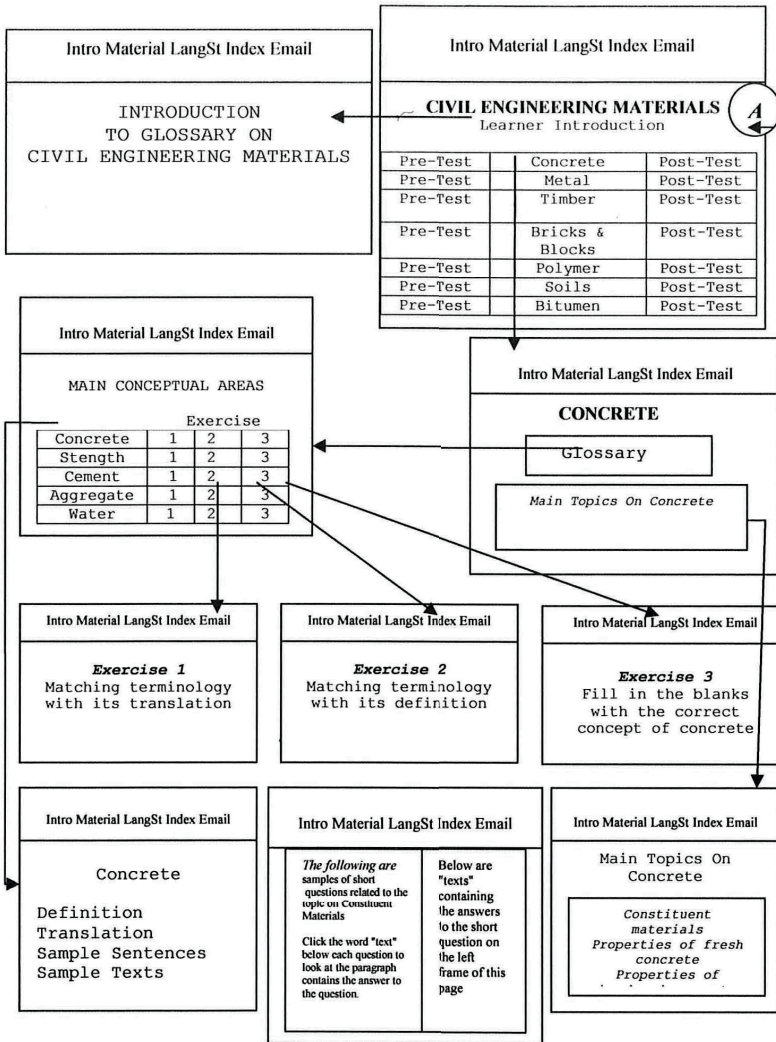


Figure 4 – Structural Framework for the Material Section

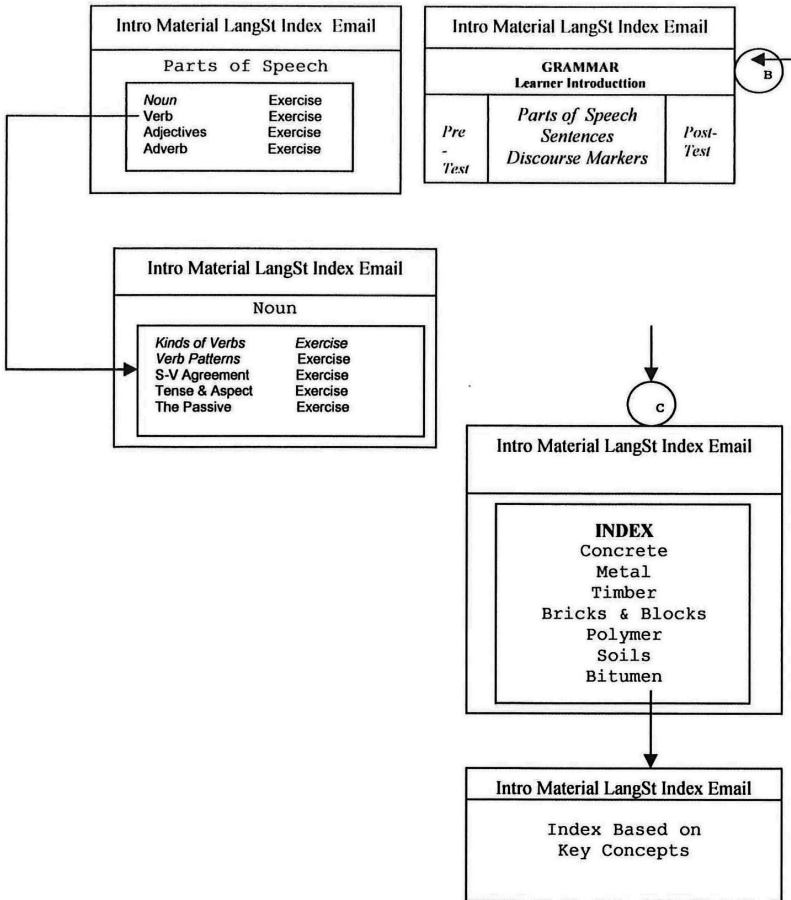


Figure 5 – Structural Framework for the Language Structure and Index Section

## **Developing the Web-Based Courseware**

The ‘UNITEKMA ECE Courseware’ is created with reference to the design of its structural framework or clusters as shown in Figure 1, 4 and 5, using the software FrontPage. Its home page or main page is a ‘Welcoming page’ consisting of four major sections (see Figure 1), which are the introduction page, material section, language structure section and index section of the courseware.

Every major section has a top frame which contains five buttons linking to all the major sections of the courseware and a link to an e-mail address the user could use to send comments on the courseware itself (see Appendix A: Introduction Section and Appendix B: Material Section). In the introduction page, students are introduced to the aim/objectives, level and content of the courseware (see Appendix A). Help is also provided on how learners could fully utilise the courseware.

The material section is organized according to the seven different CEM mentioned earlier and has a pre-test and a post-test for each CEM (see Appendix B). The purpose of the tests is for the students to evaluate their understanding of the key concepts of CEM. Each material has two sub-sections linked to it – glossary and ‘main topics’. For the glossary part, three different types of exercises are provided and scores are given for attempting the exercises (see Appendix C). The type of exercises is in the form of matching. The glossary part also contains the glossary for all conceptual areas or key concepts of every CEM mentioned earlier and common terminology or lexical items for those key concepts. The topics for the ‘main topics’ sub-section have been identified by the subject specialist who teaches the CEM course in FCE, according to importance (see Appendix D). This sub-section contains questions related to the selected topics and answers to the questions are also provided as shown in Appendix E. In general, the material section helps CE students to become familiarised with the key concepts and key topics in the area of CEM.

The language structure section contains three sub-sections – Parts of Speech, Sentences and Discourse Markers (see Appendix F). Similar to the material section, this section also contains pre-tests and post-tests for each of its sub-section. Tutorial is also provided for each sub-section of the language structure section of the courseware. Finally, the last section of the courseware is the index section which has a link to the glossary part of the material section of the ‘UNITEKMA ECE Courseware’, categorised according to the seven chosen CEM materials (as shown in Appendix G).

### **Implementing/Piloting the Courseware**

When either the prototype or first complete version of the ‘hyperlinked’ ‘Web document’ (in the form HTML files) has been developed, it has to go through the next stage, which is the implementation/pilot stage. At this stage, the ‘Web document’ has to be installed onto a computer that has a ‘Web browser’ and then tested locally on the computer using the Web browser’s local mode (Horton et. al 1996). This is done by viewing the ‘Web pages’ of the ‘Web document’ thoroughly in order to make sure that they are being displayed correctly and that all the ‘hyperlinked’ in the ‘Web pages’ function as expected. Otherwise, the ‘Web pages’ have to be adjusted and readjusted until they perform effectively as they should.

After this is done, the ‘Web document’ will have to be installed or ‘mounted’ onto a chosen ‘web server’, thus, giving it a ‘home page address’ that refers to the ‘address’ of the ‘Web server’ in addition to its own file directory and filename (Horton et. al. 1996). So, generally, the ‘home page address’ will be in the following format:

http://address of ‘web server’/file directory/filename of ‘Web document’.html.

The last step of the implementation stage is to test the 'Web document' again, by accessing it in the same way as the targeted users would on the WWW, using the document's 'home page address' (Horton et. al 1996). This is done in order to check that the 'Web pages' of the 'Web document' and its 'hyperlinks' are being displayed as expected and that the 'Web pages' are integrated onto the 'Web server' appropriately. With regards to this research project, the above process of implementing and piloting the prototype 'UNITEKMA ECE Courseware' has been conducted and the necessary changes have been made to ensure that all the links in the courseware are active and that all 'Web pages' of the courseware are linked according to its designed structural framework.

### **Evaluating the Courseware Based on Students Responses**

The final stage of the creation of the 'UNITEKMA ECE Courseware' is the evaluation stage. In this last stage, the targeted users of the 'UNITEKMA ECE Courseware', which consist of CE students in UTM who are taking the UHB1322 course of the ECE programme, have been invited to 'browse' the document and 'navigate' through it on the WWW, using all the 'hyperlinks' that are available on its 'Web pages'.

### **CONCLUSION**

In conclusion, computational linguistics analysis has helped to identify the main conceptual areas or key concepts and common terminologies or lexical items in the context of CEM, thus, facilitating students in understanding the content area of CEM. Computational linguistics analysis also provides authentic and 'domain-specific'

language learning materials for the production of the Internet-based Civil Engineering material, which is, the 'UNITEKMA ECE Courseware'.

The Internet-based civil engineering materials are useful to support the language learning process of the Civil Engineering students since they can access learning at their own time and pace. What is important is that the internet-based materials provide a platform for CE students to become familiarised with the key concepts and lexical items related to their field of specialisation, which they need to acquire in their content subjects.

## REFERENCES

- Aitkenhead AM & Slack JM (eds) 1985 *Issues in Cognitive Modeling*  
London: Lawrence Erlbaum Associates Ltd Publishers
- Anderson R C & Pearson P D 1988 'A Schema-Theoretic View of Basic Processes in Reading Comprehension' in Carrell P L, Devine J & Eskey D E (eds) *Interactive Approaches to Second Language Reading* Cambridge: CUP
- Carrell P L & Eisterhold J C 1983 'Schema Theory and ESL Reading Pedagogy' *TESOL Quarterly* 17/4: 553-573
- Carrell P L & Wallace B 1983 'Background Knowledge: Context and Familiarity in Reading Comprehension' in Clarke M & Handscombe J (eds) *On TESOL '82* Washington, D.C.: TESOL
- Carrell P L 1983 'Some Issues in Studying the Role of Schemata, or Background Knowledge, in Second Language Comprehension' *Reading in a Foreign Language* 1/2: 81-92
- Carrell P L 1984 'Evidence of a Formal Schemata in Second Language Comprehension' *Language Learning* 34/2: 87-112
- Carrell P L 1985 'Facilitating ESL Reading Comprehension by

- Teaching Text Structure' *TESOL Quarterly* 19/4: 727-752
- Carrell P L 1988 'Some Causes of Text-Boundedness and Schema Interference in ESL Reading' in Carrell P L, Devine J & Eskey D E (eds) *Interactive Approaches to Second Language Reading* Cambridge: CUP
- Carrell P L, Devine J & Eskey D E (eds) 1988 *Interactive Approaches to Second Language Reading* Cambridge: CUP
- Department of Modern Languages UTM 1996 *A Framework for Collaborative Roles and Responsibilities in Language Development for Academic Purposes between the Department of Modern Languages and the Faculty of Civil Engineering, Universiti Teknologi Malaysia* University Technology of Malaysia: Faculty of Management & Human Resource Development (unpublished)
- Greene J 1986 *Language Understanding: A Cognitive Approach* Milton Keynes: Open University Press
- Horton W, Taylor L, Ignacio A & Hoft N L 1996 *The Web Page Design Cookbook: All the Ingredients You Need to Create 5-Star Web Pages* New York: John Wiley & Sons Inc
- Jackson N & Dhir R K 1988 *Civil Engineering Materials* (4th edition) London: Macmillan
- McEnery T & Wilson A 1996 *Corpus Linguistics* Edinburgh: Edinburgh University Press
- Renouf A 1987 'Corpus Development' in Sinclair J M (ed) *Looking Up: An Account of the COBUILD Project in Lexical Computing* London: Collins Sons & Co Ltd
- Roe P 1995 *Lexical Studies Component* Aston University: Language Studies Unit
- Rumelhart D E & Norman D A 1985 'Representation of Knowledge' in Aikenhead A M & Slack J M (eds) *Issues in Cognitive Modeling* London: Lawrence Erlbaum Associates Ltd Publishers
- Sinclair J 1991 *Corpus Concordance Collocation* Oxford: OUP
- Skelton J & Willis J 1993 *Course and Syllabus Design Component*



Aston University: language Studies Unit  
Tatters W 1996 *Teach Yourself Netscape 2 Web Publishing in a Week*  
Indianapolis, IN: Sams.net Publishing  
Tribble C & Jones G 1990 *Concordances in the Classroom: Resource  
Book for Teachers* Essex: Longman Group UK Ltd

### Dictionaries:

- Hawkins J M 1991 *Kamus Dwibahasa Oxfjord Fajar: Inggeris-Melayu, Melayu-Inggeris* Selangor, Malaysia: Penerbit Fajar Bakti Sdn Bhd
- Ramli M 1990 *Kamus Kejuruteraan Bangunan* Kementerian Pendidikan Malaysia, Kuala Lumpur, Malaysia: Dewan Bahasa & Pustaka
- Salleh M H et. al. (eds) 1994 *Istilah Kejuruteraan: Bahasa Inggeris Bahasa Malaysia* Kementerian Pendidikan Malaysia, Kuala Lumpur, Malaysia: Dewan Bahasa & Pustaka
- Schwarz C, Davidson G, Seaton A & Tebbit V (eds) 1988 *Chambers English Dictionary* Edinburg: W & R Chambers Ltd
- Scott J S 1991 *The Penguin Dictionary of Civil Engineering* (4th edition) London: Penguins Books Ltd
- Sidang Pengarang 1992 *Kamus Teknikal & Kejuruteraan dengan Istilah Kejuruteraan* Kuala Lumpur, Malaysia: Golden Books Centre Sdn Bhd
- Sinclair J et. al. (eds) 1987 *Collins Cobuild English Language Dictionary* London: William Collins Sons & Co Ltd
- Vazirani V N & Chandola S P 1992 *Dictionary of Civil Engineering* New Delhi, India: Mohan Pramlani Publishers for Oxford & IBH Publishing Co.

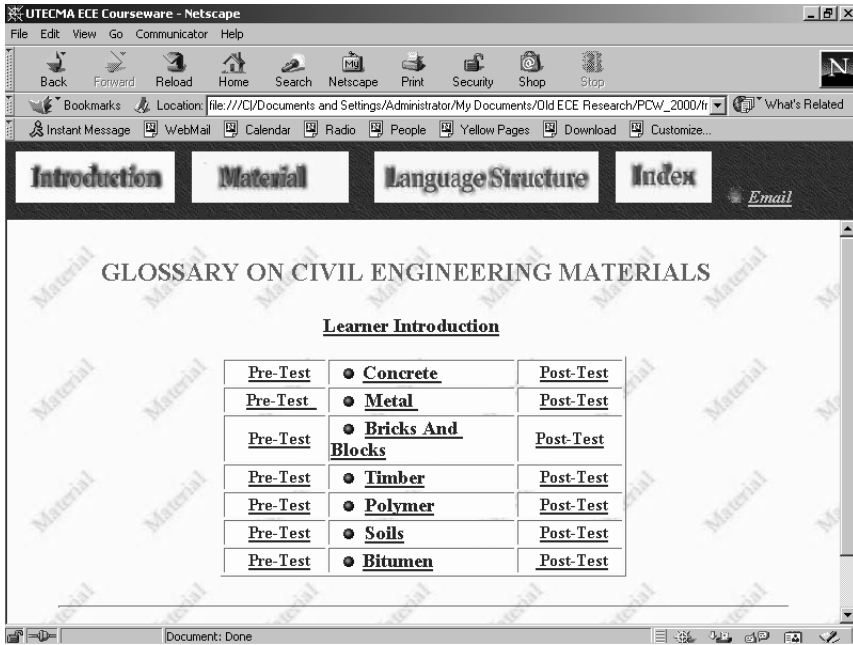
## APPENDIX A

### 'Web page' of the Introduction Section



## APPENDIX B

### 'Web page' of the Material Section



## APPENDIX C

## 'Web page' for the Material 'Concrete'

UITEMA ECE Courseware - Netscape

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Shop Stop

Bookmarks Location: file:///C:/Documents and Settings/Administrator/My Documents/Old ECE Research/PCW\_2000/tr What's Related

Instant Message WebMail Calendar Radio People Yellow Pages Download Customize...

**Introduction** **Material** **Language Structure** **Index** [Email](#)

## MAIN CONCEPTUAL AREAS FOR THE MATERIAL

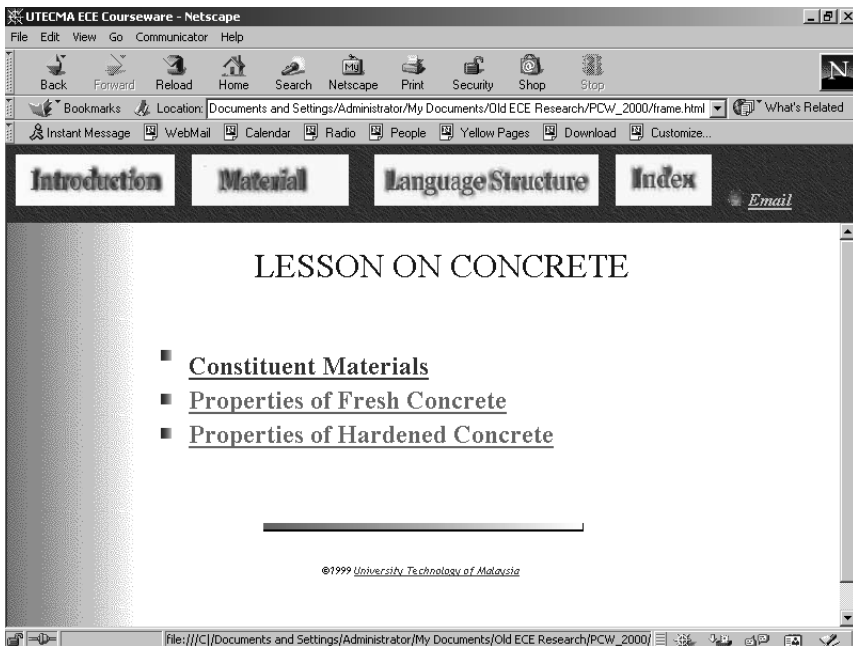
### CONCRETE

Conceptual Word	Exercise			Frequency
<u>Concrete / Concretes</u>	1	2	3	730 / 37
<u>Strength / Strengths</u>	1	2	3	335 / 22
<u>Cement / Cements</u>	1	2	3	279 / 43
<u>Aggregate / Aggregates</u>	1	2	3	172 / 66
<u>Water</u>				152
<u>Test / Tests</u>	1	2	3	137 / 59
<u>Mix / Mixes</u>				128 / 46
<u>Content</u>				120
<u>Workability</u>	1	2	3	107
<u>Ratio</u>				91

Document: Done

## APPENDIX D

‘Web page’ containing important topics on theMaterial  
‘Concrete



## APPENDIX E

### ‘Web page’ containing questions and answers to the questions related to the selected topic on ‘Constituent Materials’ of the Material ‘Concrete’

UTECEMA ECE Courseware - Netscape

File Edit View Go Communicator Help

Introduction Material Language Structure Index [Email](#)

The Following are samples of short questions related to the topic on Constituent Materials

Note :

Click the word "text" below each question to look at the text which contains the answer to the question. (The "text" will appear on the right page of this frame)

- Briefly describe, what is concrete ?  
[Text](#)
- Concrete known as a brittle material. How is the problems being overcome

### CONSTITUENT MATERIALS

Below are "text" containing the answers to the short questions on the left frame of this page.

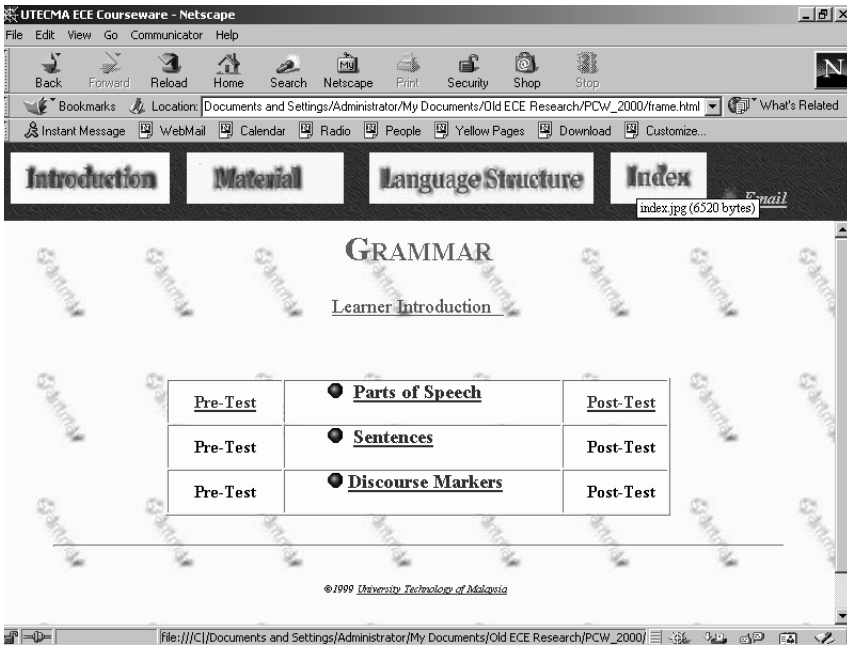
Text for Question number 1

Concrete is a man-made composite the major constituent of which is natural aggregate, such as gravel and sand or crushed rock. Alternatively artificial aggregates, for example, blast-furnace slag, expanded clay, broken brick and steel shot, may be used where appropriate. The other principal constituent of concrete is the binding medium used to bind the aggregate particles together to form a hard composite material. The most commonly used binding medium is the product formed by a chemical reaction between cement and water. Other binding mediums are used on a much smaller scale for special concretes in which the cement and water of normal concretes are replaced either wholly or in part by epoxide or polyester resins. These polymer concretes known as resin-based or resin-additive concretes respectively are costly and generally not suitable for use where fire-resistant properties are required but they are useful for repair work and other special applications. Resin-based concretes have been used, for example, for precast chemical-resistant pipes and lightweight drainage channels. This Part deals only with normal concretes

Document: Done

## APPENDIX F

### ‘Web page’ of the Language Structure Section



## APPENDIX G

### ‘Web page’ of the Index Section

