UNIVERSITI TEKNOLOGI MARA

A CORPUS-BASED STUDY OF MORPHOLOGICAL PRODUCTIVITY OF ENGLISH LANGUAGE CHEMICAL ENGINEERING TEXTBOOKS

NORRIHAN BT SULAN

Thesis submitted in fulfilment of the requirement
for the degree of
Doctor of Philosophy

Faculty of Education

MAY 2011

Abstract

The effective teaching of word identification requires the acquisition and mastery of specific word identification skills. The aim of this study is to determine the patterns of complex words which, because of the large number of meanings signalled by word derivations, are the foundation for decomposition skills. It also aims to develop lexical knowledge of words with complex composition through the study of morphologically productive affixes in chemical engineering textbooks. It seeks to find the most productive morphological categories in the specialized corpus created and to find the density of complex words and their morphological patterns.

A corpus of Chemical Engineering Level 1 (CEL1) textbooks used at Universiti Malaysia Pahang (UMP), Malaysia was designed to capture a range of linguistic features in actual language use for more effective teaching and learning. This study focuses on theories on word recognition which proposed that reading consists of decoding and linguistic comprehension which are necessary for reading success. Baayen's Morphological Productivity Measurement is used to determine the most productive affixes which will be the base for materials development for teaching and learning processes.

The study looks at the frequent patterns of morphologically complex words which might enable language instructors to design reading materials based on actual language use. It was found that 75.57% of the affixed words in the corpus are single-affixed words while the most complex word has six affixes attached to it. The most frequent prefixes in the corpus are *un-, re-, de-,pre-* and *dis-* while for suffixes, the most frequent occurrence are *ion-,-ly, -er, -al, -able/ible.* For prefixes attached to technical words, *dia-, hydro-, poly-, iso-,* and *thermo-* are found to be the most frequently used.

The findings from this study are potentially beneficial for developing ESP materials to meet the linguistic needs of students in engineering and related disciplines. The findings have pedagogical implications for language teaching and learning enabling teaching materials to be produced with the information from the CEL1 corpus. Teaching materials can be designed using the complex word patterns and presented to allow multiple exposure, thus enhancing word recognition skills.

TABLE OF CONTENT

TITLE PAGE		
DECLARATION		ii
ABSTRACT		iii
ACKNOWLEDGEMENTS		iv
TABLE OF CONTENTS		vi
LIST OF TABLES		xi
LIST OF FIGURES		xii
LIST OF ABBREVIATIONS		xiii

CHAPTER 1: INTRODUCTION

1.1	Background	1
1.2	Role of Morphological Knowledge	7
1.3	Morphologically Complex Words	8
	1.3.1 Multi-morphemic Words	11
	1.3.2 Morphological Productivity	12
1.4	Overview of Corpus Linguistics	13
1.5	Corpora for Specialized Domain	16
1.6	WordSmith Tools Version 4.0	16
1.7	Conceptual Framework	18
1.8	Morphological Analysis	19
1.9	Morphological Instruction	19
1.10	Research Problem	20
1.11	Statement of Problem	22
1.12	Aims of the research	25
1.13	Research Questions	25
1.14	Scope of Study	25
1.15	Limitations of the study	29
1.16	Significance of study	26
1.17	Operational Definitions	29
1.18	Summary	30

CHAPTER 2 : LITERATURE REVIEW

2.1	Introduction	32
2.2	Language Proficiency and Specialized Vocabulary	33
2.3	Morphological Awareness for Word Recognition Skills	35
2.4	Morphology and English for Academic Purposes (EAP).	36
2.5	Morphological Knowledge and Vocabulary Expansion	37
2.6	Corpus-based Study from the Perspectives of Language Teaching and Learning	38
2.7	Derivational Morphemes	40
2.8	Inflectional affixes	44
2.9	Vocabulary Development	45
2.10	Corpus linguistics Approach	47
2.11	Studies in Malaysia	49
2.12	Statistics for corpus-based study	50
2.13	Advantages of Corpus-based research	51

2.14	Corpus building	51
2.15	Limitation of the corpus building	52
2.16	Past Studies	53
2.17	Corpora and English for Specific Purposes	59
2.18	Corpora Integration in Language Learning Environment	60
2.19	Implications of Morphological Productivity to Vocabulary Instruction	61
2.20	Academic English	63
2.21	Quantitative and Qualitative Approach in Corpus-based Study	65
2.22	Morphological awareness	66
2.23	Word Formation	67
2.24	Corpus based study on Morphological Productivity	72
2.25	Significance of study of Morphological productivity.	76
2.26	Morphological Awareness and Reading Ability	80
2.27	Prefixes in technical word-formation.	82
2.28	Small Corpora	83
2.29	The Significance of the Analysis of Frequent Words	86
2.30	The Significance of Vocabulary Instruction	89
2.31	Implication of Vocabulary Instruction Using Morphological knowledge	93
2.32	Corpus Resources: Written Texts	99
2.33	Theoretical Background and Implications of Research	100
2.34	Guidelines in Designing a Specialized Corpus	104
2.35	Some Examples of Specialized Corpus	106
2.36	Summary	107

CHAPTER 3 : RESEARCH METHODOLOGY

3.1	Introduction	108
3.2	Research Design	108
	3.2.1 Corpus Design of Chemical Engineering Level 1 (CEL1)	108
	3.2.2 Textbook as Corpus Resources	109
	3.2.3 General Information of the CEL1 Corpus	111
3.3	Procedure for data collection	113
	3.3.1 Building the corpora	113
	3.3.2 Stage1: Digitalizing the Corpus	114
	3.3.3 Stage 2: Cleaning the corpus	115
	3.3.4 Stage 3: Generating a Wordlist	116
3.4	Data Analysis Procedure	120
	3.4.1 Generating frequency List	120
3.5	Sampling Technique	124
	3.5.1 Size	124
	3.5.2 Representativeness	126
	3.5.3 Scope of study	127
	3.5.4 Multiple Affixation	128
	3.5.5 Affixed Technical Words	129
	3.5.6 Limitations	129

	3.5.7	Quantitative Measures	131
	3.5.8	Token-based Productivity	132
		3.5.8.1 Prefix Token based	132
		3.5.8.2 Suffix Token- frequency	134
	3.5.9	Type-based productivity	136
		3.5.9.1 Prefix Type-Frequency	137
		3.5.9.2 Suffix Type- Frequency	138
	3.5.10	Ranking of Morphological Productivity	139
		3.5.10.1 Hapaxes	141
3.6	Data Analysis		144
	3.6.1	Qualitative Measure	144
	3.6.2	Quantitative Productivity Analysis	145
3.7	Summary		146

CHAPTER 4 : FINDINGS

4.1	INTRODUCTION		148
4.2	Frequency Effects		151
	4.2.1	Type Frequency	152
	4.2.2	Type Frequency for Suffixed Words	153
	4.2.3	Adjective forming suffix	156
	4.2.4	Noun forming suffixes	157
	4.2.5	Verb Forming Suffixes	158
	4.2.6	Adverb Forming Suffixes	159
4.3	Type Frequency of Prefixed Words		160
4.4	Token Frequency		163
4.5	Token Frequency of Prefixed Technical Words		168
4.6	Token Frequency for Suffixed Technical Words		171
4.7	Productivity Rank		173
	4.7.1	The relationship between frequency and productivity	173
	4.7.2	Suffixes Productivity Rank	174
	4.7.3	Prefixes Productivity Rank	176
	4.7.4	Problematic aspects of Morphological Analysis	177
	4.7.5	Density of Complex Words	178
	4.7.6	Patterns of Complex Word	181
4.8	Compound Words		183
4. 9	Word Formation Patterns		185
4.10	Summary		189

CHAPTER 5: DISCUSSION

5.1	Introduction	190
5.2	CEL1 and its Implication to Material Development	190
5.3	Corpus and Word List Generation	191
5.4	CEL1 and Its Implication to Grammatical Competence	192
5.5	Implication to Morphological Knowledge	193
5.6	Contribution to Content Area Teaching	197
5.7	CEL1 Wordlist and Its Implication to Vocabulary Instruction	199

5.8	Word-formation Processes From CEL1 Analysis		204
5.9	Morphological Productivity and Its Impact on Material Development		205
5.10	Technical Prefixes		206
5.11	Complexity of Words in Content Area Reading Materials		209
5.12	Theoretical Implications		210
	5.12.1	Impact of Morphological Awareness and Reading Ability	210
	5.12.2	Implication of Multiple Exposure to Word Recognition	211
	5.12.3	Importance of Word Formation Knowledge to Language Teaching	212
	5.12.4	Specific Word Instruction	214
	5.12.5	The Significance of Frequency in Word Teaching	219
	5.12.6	Implications of Morphological Knowledge to Content Area Teaching	222
	5.12.7	Implications of Investigating Morphological Characteristics in a Corpus	225
	5.12.8	Implications for language learning and teaching	227
5.13	Summary		229

CHAPTER 6 : CONCLUSION

6.1	Introduction	230
6.2	Recommendations	233
6.3.	Future Research	240
6.4	Summary	241

References		242
Appendices		
Appendix A :	Stop List	274
Appendix B :	Morphological Productivity (prefixes)	275
Appendix C:	Morphological Productivity (suffixes)	276
Appendix D:	Example of Word List and Frequency Count	278
Appendix E:	Pictures, Graphs, captions, Formulas, Equations, Diagrams, Tables, Photos	280
Appendix F:	Type Frequency of Suffixes Arranged in Descending Order	281
Appendix G:	Type Frequency of Prefixes Arranged in Descending Order	283
Appendix H:	Token Frequency of Suffixes Arranged in Descending Order	285
Appendix I:	Token Frequency of Prefixes Arranged in Descending Order	286
Appendix J:	Prefixes Attached to Technical Words	287
Appendix K:	Token Frequency of Suffixed Technical Words in	288

	Descending Order	
Appendix L:	Examples of Compound Words in Descending Order	289
Appendix M:	The 20 Most Common Prefixes in Academic Texts, White et al. (1989)	290
Appendix N:	The Common Latin and Greek Prefixes, White et al. (1989)	291
Appendix O:	The 20 Most Common Suffixes in Academic Texts, White et al. (1989)	292
Appendix P:	Example of Words Arranged for Specialized Glossary (CEL1)	294
Appendix Q:	Lesson Plans	295

LIST OF TABLES

Table 3.1:	Titles of Textbooks	110
Table 3.2:	Statistics of the Chemical Engineering Corpus 1 (CEL1)	112
Table 3.3:	Example of Items in A stop list	119
Table 3.4:	Number of Tokens after stop list	122
Table 3.5:	Structure of CEL1	125
Table 3.6:	Formula for Measuring Morphological Productivity	141
Table 3.7:	Measure of Productivity (Prefix)	143
Table 3.8:	Measure of Productivity (Suffix)	144
Table 4.1:	Statistics of the Chemical Engineering Corpus 1 (CEL1)	150
Table 4.2:	Categories of suffixes from CEL1	155
Table 4.3:	Functions of Prefixes in CEL1	161
Table 4.4 :	Categories of Token-based Frequency of Suffixes in CEL1	164
Table 4.5:	Examples and Meanings of Token-based Frequency of Prefixes in CEL1	166
Table 4.6:	Functions of Technical Prefixes in CEL1	169
Table 4.7:	Categories of Suffixes for Technical Words in CEL1	172
Table 4.8:	Productivity Rank for Suffixes in CEL1	174
Table 4.9:	Productivity Rank for Prefixes	176
Table 4.10:	Composition of Affixed Words	179
Table 4.11:	Distribution of Complex Words in the Corpus.	180
Table 4.12:	Patterns of Complex Words in CEL1	182
Table 4.13:	Compound Words Sorted in Descending Order.	184
Table 4.14:	Most Frequent Patterns of Word-formation from CEL1	187

LIST OF FIGURES

Figure 1.1:	Conceptual Framework	17
Figure 3.1:	An Extract of CEL1 Corpus.	114
Figure 3.2:	Misspelled words	116
Figure 3.3:	Words, Frequency and Percentage	118
Figure 3.4:	Stop List	122
Figure 3.5:	Main stages in the creation of CEL1	123
Figure 3.6:	The Wordlist in Alphabetical Order	133
Figure 3.7:	Sorted Prefix Word List	134
Figure 3.8:	Wordlist Sorted in Reversed Order	135
Figure 3.9:	Word list with suffix *–ian*	136
Figure 3.10:	Example of Types and Tokens in CEL1 Word List	137
Figure 3.11:	Example of Prefix Type-Frequency	138
Figure 3.12 :	Example of Suffix Type-Frequency	139
Figure 3.13:	Setting the Frequency of Hapaxes	142
Figure 3.14:	Setting the Wordlist of Hapaxes	143
Figure 4.1:	The Breakup of the CEL1 Before and After Cleaning	150
Figure 4.2 :	The Most Frequent Word Derivation Process for Suffixes (Type)	159
Figure 4.3:	The Most Frequent Word Derivation Process For Suffixes (Token)	166
Figure 4.4:	The Most Frequent Word Derivation Process For Suffixed Technical Words (Tokens) in CEL1	173
Figure 4.5:	The Most Productive Suffixes	175
Figure 4.6:	The Most Productive Prefixes	177
Figure 4.7:	Distribution of Complex Words	179
Figure 4.8:	Composition of Word Category in CEL1	181
Figure 4.9 :	The frequency count of compound words in CEL1	185
Figure 4.10:	The ten most frequent word-formation patterns	188
Figure 5.1:	Levels of Suffixes in Relation to Types and Tokens	201
Figure 5.2:	The 20 Most Common Suffixes in Academic Texts	201
Figure 5.3:	Levels of Prefixes in Relation to Types and Tokens from CEL1	202
Figure 5.4:	The 20 Most Common Prefixes by White, Sowell and Yanagihara (1989)	203
Figure 5.5:	Common Greek Prefixes (White et al.)	207
Figure 5.6:	Common Latin Prefixes (White et al.)	207
Figure 5.7 :	Technical Prefixes in Relation to Tokens	208

LIST OF ABBREVIATIONS

CEL1	Chemical Engineering Level 1
ESP	English for Specific Purposes
ESL	English as a Second Language
EFL	English as a Foreign Language
NLP	Natural Language Processing
EAP	English for Academic Purposes
UWL	University Word List
LOB	Lancaster-Oslo-Bergen Corpus
OCR	Optical Character Recognition
COBUILD	Collins Birmingham University International Language Database
KMITL	King Mongkut's Institute of Technology Ladkrabang
ENGICOR	Environmental Engineering Corpus
BNC	British National Corpus
SEEC	Student Engineering English Corpus
DCL	Detailed Consistency List
WKT	Word Knowledge Task
JDEST	Jiaotong Daxue English of Science and Technology
HKUST	Hong Kong University of Science and Technology
GPEC	Guangzhou Petroleum English Corpus
MICASE	Michigan Corpus of Academic Spoken English
TIF	Tagged Image File
UMP	Universiti Malaysia Pahang
L2	Second language
TTR	Type-Token Ratio
STTR	Standardised Type-Token Ratio
WF	Word Formation

CHAPTER 1

INTRODUCTION

1.1 Background

English is one of the most powerful languages for networking globally and is equally important in academic and professional environment. Many students of higher learning institutions, where English is a second or foreign language, do not achieve their educational goals because they lack the proficiency to understand the texts that they read. Language instructors often cite these problems as lack of fluency or adequate control over the language, including inadequate vocabulary (MacGowan-Gilhooly, 1996). English language learners with limited vocabulary often have difficulty to understand complex reading materials which are often related to their core courses.

In trying to ensure that students know the meaning of basic words, and to provide enough input and reinforcement, language instructors are confronted with the need to design effective vocabulary instruction for students learning English and applying knowledge in their content area subjects. There are several issues to be considered and one of them is the determination of the actual words these students will encounter in specific academic settings taking into consideration the large deficits in English vocabulary.

Reading materials for higher education can pose a great problem for students who lack proficiency in the target language. Engineering students, for example, encounter numerous complex words in their field of study which require them to have word recognition skills. For students in Malaysian universities, most of the texts and reference

books are in English. Many students of such programmes have to learn and memorize extensive lists of complex words, jargons and terminology. Scientific words in these scientific registers consist of several components and knowing the common ones often appear in related disciplines, will enhance students' reading competency.

A crucial aspect in teaching English to university students is to develop their knowledge of technical terms and complex words. As students are generally aware that language plays an important role in their future professional career, careful selection of the specialized language to be taught should be emphasized to provide students with effective teaching and learning. At present, there have been analyses using specialized corpora containing a range of textbooks or other sources for a specialized domain which provide lists of vocabulary intended for language teaching. However, it is found that insufficient information of patterns of complex words present in scientific genres, especially engineering. Thus, a study of this area is significant as research claim that one difference between achieving and non-achieving students is their level of vocabulary development (Elley, 1988; Hart & Risley, 1995).

As language instructors rely on textbooks for vocabulary teaching, they need to consider the lexical items which are specific to certain genres compared to general English. The basic principles for word selection, according to Sinclair and Renouf (1988) are frequency, patterns of usage, and the typical combinations of words. This study therefore explores the frequency of word-formation of complex words as well as the patterns of usage based on a corpus created for this purpose. Many technical and complex words in engineering English can be decomposed into morphemes which can facilitate the learning of a new word through word-formation identification.

Word recognition skills are critical for students to develop their reading ability as independent reader especially in Malaysia, where English is taught as a second language (ESL). Therefore, knowledge of prefixes, suffixes and root words is significant in morphemic analysis which can assist students to predict meanings of words and can also help students unlock the meanings of words. Gruber (1986) noted that a child's knowledge of 20 prefixes and 14 roots can unlock the meaning of over 100,000 words and a set of 29 prefixes and 25 roots will give meanings to over 125,000 words. Students of higher education whose knowledge of affixes is greater than that cited by Gruber should be able to comprehend the functions of more words if they are taught to use affixes relevant to their discipline.

In English, some suffixes are commonly used, but there are some affixes which are more peculiar to medical science than to the engineering register. For example, in the medical sciences, the suffix *'it is'* is used to indicate disease are *"itis," bringing the* meaning of inflammation; *"oma,"* meaning tumor; and *"osis,"* meaning a condition, usually morbid. The suffixes are common in medical terminology but they may not be common suffixes in engineering or other science-based domains.

Chemistry and other science and technology domains encompass other territories in the science domain. It is a known fact that science and engineering require precision. The main function of specialized language is to make possible the transfer of knowledge and the communication between experts either of the same or different fields. The range of this domain is big, from chemistry to medicine and from geology to agricultural and biotechnology. Due to the emerging needs of new terms, new reactions or new concepts, language too needs to grow dynamically.

Contrary to the needs for a dynamic growth of language among university students and the ability of reading complex texts which is the fundamental skill for tertiary level students, many students have low reading ability often associated with the reason why students do not perform well which affects students' academic performance. Since English is central in the academic and professional lives of students, it is important that it is included in the curriculum in ways that are effective in developing their ability to use English effectively in actual communication situations as well as to excel academically while undergoing their study at tertiary level. At this level, students are normally required to use English in activities like large group discussions, class discussions, out of class projects, small group work and class participation. Limited vocabulary can prevent students from developing ideas and arguments effectively.

The selection of words to be taught in tertiary education relates to the learners' language achievement. This is one of the challenges for language instructors to decide on the lexical content where the choice of vocabulary is of utmost importance apart from other language skills. Therefore, detailed accounts of the language in use must be determined if teachers are to meet the needs of the learners at tertiary level. This means inclusion of a list of words identified to be representative of the language in context is essential. Learners' language needs have to be matched with the specialized vocabulary pertaining to specific language domain or genre. The acquisition of vocabulary is central in learning a second language. Language planners must be aware of the words or vocabulary to be integrated in the material development in order to improve and reinforce students' competency. Vocabulary is knowledge of words and word meaning that students encounter in content area textbook and related readings. In this context, vocabulary refers to words that students must know when attempting to read

difficult text with a good amount of comprehension. One of the most persistent findings in reading research is that the extent of students' vocabulary knowledge relates strongly to their reading comprehension and overall academic success (Baumann, Kameenui, & Ash, 2003,; Beck, McKeown, & Kucan, 2002).

Schmitt (1998) highlights the need to focus on individual words and to be able to measure the degree or depth of knowledge for each of these words, in order to reach a better understanding of vocabulary acquisition. Nagy and Scott (2000) mention several dimensions that describe the complexity of what it means to know a word. Word knowledge is incremental where readers need to have many exposures to a word in different contexts before they recognize it. Word knowledge is also multidimensional where words can have multiple meanings (e.g., *plant*: a living thing which grows in earth, factory in which such machines are used). This word can also have different functions when used in different contexts.

The inclusion of English in technical and engineering studies enhances the students' ability to access the extensive information available in English and provides them with opportunities to practice using English before they enter the job market. However, in a recent survey on oral communication skills among Malaysian graduates (Fatmawati et al., 2005), on the item of graduates' ability to present or deliver ideas and opinions clearly, it was found that only 4% of the participating human resources managers felt that graduates fulfilled the industries' requirement. From the survey, it was clear that there is a need to teach engineers to be proficient in their communication but language instructors are still grappling with the problem. Vocabulary learning is important to expand the word knowledge and is needed by learners to become successful readers of English.

One way to research the language of textbooks is to use a corpus. Science and technology-based texts are written in a very specific form which requires the readers to be competent and well-equipped with the necessary skills in dealing with complex words. On the other hand, the ability to use knowledge of root words and affixes to determine the meanings of unfamiliar, morphologically complex words is important to draw students' attention to the morphemes that form a word. The ability to analyse the meanings of the individual morphemes helps students to understand the meaning of the whole word. The ability to identify meanings of common affixes and combining them with meanings of familiar roots can assist in determining the meanings of many new words (Burns, Roe and Ross, 1999).

The goal of effective reading is to comprehend reading texts, and the ability to identify each individual word is a crucial to assist comprehension. Reading involves a highly interactive set of complex processes which requires the ability to recognize or decipher printed words (decoding), and those concerned with understanding the message (reading comprehension)T. o develop linguistic competency, students with reading difficulty need more than opportunities for incidental learning. Templeton & Morris (2000) have the opinion that focus on words with clear instruction that includes can strengthen the effects of vocabulary learning. Grabe (1991) informs that a vocabulary of high-frequency words (estimates range from 2,000 to 10,000 words) prepares students to learn English from context.

1.2 Role of Morphological Knowledge

Many new words are made up of simple word parts (prefixes, roots, and suffixes) which have their own functions . If students are familiar with the words, they are able to use the words even if they do not understand the exact definition. Pittman (2003) suggests that by gradually studying the most commonly occurring prefixes, roots, and suffixes, students can acquire a vocabulary even in it most complex form.

In English, of the 20,000 most commonly used words, 20% are formed with prefixes and among these words 15 prefixes comprise 82% of the prefixes used (Roe, Stoodt & Burns, 1987). Honig, Diamond and Gutlohn (2000) claim that four most frequent prefixes (*un, re, in/im/il/ir/* and *dis*) account for 97% of prefixed words in printed school English which support the notion that morphological knowledge increases vocabulary skills which, in turn, contributes to successful reading comprehension. In addition, knowledge of morphemes and their spellings boost recognition of word parts and, thus, are more efficient in decoding. Knowing the morphological relations between words will result in understanding how words are formed through the use of prefixes, suffixes, and roots, are better in reading comprehension.

Advanced reading materials often contain multi-syllabic words which consist of morphemes and students who lack morpheme awareness will have difficulty in comprehending the reading texts. The knowledge of morphology will assist students dealing with advanced reading texts.

1.3 Morphologically Complex Words

English vocabulary is steadily growing with science and technological advancement. Pitman (2003) mentioned that the huge amount of the new words used to express novel ideas, are complex words formed from simple word parts (prefixes, roots, and suffixes). Reading and vocabulary are very much related where the richer the vocabulary, the more efficient the reading process and the more one reads the richer the vocabulary becomes. For many L2 students in higher learning institutions, reading can pose a problem if they are not proficient enough and lack the skills of identifying multimorphemic words which challenge both reading and understanding reading texts. Some words are structurally complex. Morphologically complex words contain more than one morpheme and are sometimes referred to as polymorphemic words. For example, the word *recirculation* contains the prefix *re-,* and suffix *-ion*. Monomorphemic on the other hand, are words which consist of only one morpheme, for example, *tear, and lay.*

Nation (2001) opines that when second language learners encounter complex words with multiple affixes, there are bound to be some possibilities. They might be able to recognize the words and understand the words, they might not be able to recognize it until they use a guessing strategy based on its context or they might not be able to recognize the words but make some attempts to guess the meaning by decomposing the word into parts using their prior knowledge of word analysis. Students at tertiary level will often encounter polymorphemic words or words with multiple affixes in their reading texts. Students who are less proficient in the target language will definitely have some problem in decoding words like *superplasticizer, semipermeable, electrolytically*

and numerous other words which need morphological knowledge to help them in comprehending the complex words.

Lack of language skills in decomposing morphemes might pose a problem for students who are not proficient enough to use the words effectively and correctly. Moreover, the danger of wrongly using or interpreting the meaning of words will somehow affect their subject matter knowledge. Students' ability to recognize the key prefixes, stem words and suffixes for their field of study will enable them to decode many of the technical terms they meet in their studies as scientific terms as well as regular words in that discipline often use prefixes and suffixes to provide additional information to the stem component of a term or word. For example, if a student encounters a word '*premultiplying*' in a subject of Process Dynamic in a chemical engineering program, the student will be able to guess the meaning of the word if they are able to decompose the word into *pre+multiple+y+ing*. Hence, knowing the function of the prefix *"pre-"* and the suffix *"-ing"* will somehow assist the student into making sense of the meaning of the whole word *"premultiplying"*.

In scientific complex words, there are often common ways of forming words and knowing the ways to structure words can help students to expand their vocabulary thus enabling them to decipher complex words which are commonly found in their textbooks or other reading resources. Maxwell (2004) discusses that in coining new words affixes are manipulated in ways which are effective in communicating a particular meaning especially in the ever growing field of technology where new findings often require new word formation to display precision in conveying new ideas.

Coady and Huckin (1997) agree that word knowledge is necessary for communicative competence. In this context, vocabulary knowledge includes knowledge

of pronunciation, spelling, meanings in contexts and the frequency of use and also how the word combines with other words (Qian, 1999). However, this study will only look at one of the most beneficial ways to expand students' vocabulary that is affixed words.

An important component of skilled reading includes the knowledge of morphology which is the ability to gain information about the meaning, pronunciation, and part of speech of new words from their prefixes, roots, and suffixes. Kuo and Anderson (2006) find that children speaking several languages indicate that knowledge of inflectional morphology is acquired before knowledge of derivational morphology and the morphology of compounds. This knowledge is built-up from primary education until the secondary education. The issue on students' vocabulary knowledge and their ability to successfully comprehend what they read pose a great challenge for curriculum planners, and programme developers to plan and strategize towards an effective reading programme. Thus, one of the challenges would be in morphological awareness, which refers to the ability to reflect on and manipulate morphemes and word formation rules in a language.

Schmitt and Meara (1997) suggest that morphology and associations are related in that when more members are included within a word family because of increasing control of the morphological system, there will be a greater range of potential associations, as each affixed form will tend to bring different associations. Kuo and Anderson (2006) stress that morphological awareness is an important indicator of the ability to decode morphologically complex words, thus, contributing to the expansion of reading comprehension. Morphological awareness is an important predictor of reading as students advance to higher level of education and have to deal with more complex composition of words in their content area of study. Morphological awareness includes

other aspects of competency which involve metalinguistic awareness and linguistic competence which are essential in expanding one's linguistic competence.

1.3.1 Multi-morphemic Words.

The knowledge of the morphology of a language contributes to one's linguistic competence. Multi-morphemic words other than compounds may be divided into roots and affixes. Word formation is an important source of new words which involves in creating new words the language, or by changing the meaning of an established word. New words and meanings like *Internet, router*, and *robot* emerge as technology accelerates the development of science and technology. New words are created through the processes of compounding or by adding prefixes or suffixes to existing words to derive new forms of words which also change their meaning. Word formation is the process of creating new words. With the development of science and technology, new words are formed using various word formation processes like compounding, blending, back-formation even borrowing. A student who understands what *perm* means and what *able* means will be able to put two and two together when first confronted with the word *'permeable.'* Morphemes can also be guide to help students decide whether a word is a verb, adjective, or noun and thus provide another type of clue to the word's meaning and function in the sentence. The most crucial requirement is a basic understanding of the most important or well-used prefixes, roots and suffixes in the English language. For example, Katamba (1994) asserted that knowledge of English word-structure involves knowing the grammatical class of the words produced using a

given affix. This knowledge is useful to determine the grammatical class of the words formed in order to know the contexts to deploy them.

1.3.2 Morphological Productivity

Baayen (1992) stated that morphological categories have a fixed or declining membership, while others have a growing membership where categories with fixed or declining membership are said to be unproductive; categories with growing membership are described as productive. Baayen (1992) also describes how corpora allow researchers to explore how productivity varies across registers. Corpus-derived measures also play an increasingly important role in lexical processing in the mental lexicon. A morphological category with many members is said to be more productive in the sense that it has produced many complex words that are useful to the language community. Plag (2004) adds that morphological productivity is the property of a morphological process that give rise to new formations on a systematic basis.

In further explanation of morphological productivity, Plag also mentions that some rules or affixes are more often used to create new words compared to other types of affixes, thus indicate that morphological rules apply where patterns of complex words are formed. As mentioned earlier, many words in a language are morphologically related but not all have the same degree of productivity. Lowie (1998) asserts that recent research has shown that newly acquired words are better retained if they are initially inferred through linguistic cues rather than through context. Thus, focusing on the morphological structure of words may assist students in increasing their awareness of morphologically complex words which is an important strategy of inferring and acquiring words.

Morphological cues can be essential to vocabulary acquisition where learners with good word analysis skills will be more successful in their language skills.

Linguistic description of morphological processes details out the rules that govern the morphological processes. In other words, some morphological rules apply more easily to produce new forms than others cannot be expressed in such rules. Data from specialized corpus developed can be computed to measure the 'degrees of productivity' and to compare the productivity of different morphological processes. This information is important to determine the morphological productivity of any morphological category for course developers to put particular emphasis on. Studies on morphological productivity show that a quantitative approach to productivity requires a different qualitative view as well (Rainer, 1987; Corbin, 1987; Baayen,1991,1992; Baayen and Lieber, 1991; Baayen and Renouf, 1996). Thus, within the framework, it is possible to analyse the results, which is not possible by introspection.

1.4 Overview of Corpus Linguistics

Corpus linguistic, in a broad definition, is the systematic study of human language via analyses of corpora that consist of authentic examples of language use (Hunston, 2006; Barodal, 2002). It is a powerful research methodology which can offer new insights into features of language under study. A corpus is a linguistic database, which can consist of either spoken or written language (Barodal, 2002). Computer aided corpora-based research approaches made it possible to calculate relative frequency of words, to compare word frequency in different registers, detect the collocates of words and numerous other applications which seemed impossible decades ago. As Leech (1992) puts it, the technology gives us the ability to analyse the contents of such corpora in a