DEPTH OF VOCABULARY KNOWLEDGE:
ASSESSING ITS ROLE IN ADULTS’ READING
COMPREHENSION IN ENGLISH AS A SECOND LANGUAGE

by

DAVID D. QIAN

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Department of Curriculum, Teaching and Learning
Ontario Institute for Studies in Education of the
University of Toronto

© Copyright by David D. Qian 1998
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

0-612-33914-9
DEPTH OF VOCABULARY KNOWLEDGE: ASSESSING ITS ROLE IN ADULTS’ READING COMPREHENSION IN ENGLISH AS A SECOND LANGUAGE

David D. Qian
Doctor of Philosophy, 1998

Department of Curriculum, Teaching and Learning
Ontario Institute for Studies of Education of the
University of Toronto

ABSTRACT

This thesis research explored the relationships among vocabulary size, depth of vocabulary knowledge, and reading comprehension in English as a second language (ESL). Specifically, using multivariate and content analyses, the research assessed the role of depth of vocabulary knowledge in 74 adult Chinese and Korean speakers’ comprehension of general academic texts in English. The main study examined to what extent depth of vocabulary knowledge adds to the prediction of reading comprehension scores over and above the prediction afforded by vocabulary size, employing as instruments a portfolio of vocabulary knowledge tests, a reading comprehension test, and a background questionnaire. A follow-up study then investigated strategies for processing the meaning of unknown words used by a subsample of these ESL learners with different depths of vocabulary knowledge. For this purpose, individual interviews and a survey questionnaire on reading strategies were used as the main means of data collection.
Among other findings, the research produced empirical evidence that: (a) scores on vocabulary size, depth of vocabulary knowledge, and reading comprehension were positively, and closely, related; (b) depth of vocabulary knowledge made a unique contribution to the prediction of reading comprehension scores, over and above the prediction afforded by vocabulary size; (c) depth of vocabulary knowledge played a fundamental role in these ESL learners’ reading comprehension processes; (d) there was a positive relationship between the learners’ depth of vocabulary knowledge and their lexical inferencing ability; and (e) in processing the meaning of unknown words, all learners looked for cues to meaning, but those with greater depth of vocabulary knowledge appeared to focus more on word meanings, whereas learners with less depth of vocabulary knowledge tended to focus more on word forms. The results of this research point to the importance and necessity of improving the depth of learners’ vocabulary knowledge in their ESL learning.
ACKNOWLEDGMENTS

Many people contributed to my doctoral research. All these years, Dr. Birgit Harley, my thesis supervisor, has been extremely helpful. She was always ready to help when I had any questions. I benefitted greatly from her expertise in second language acquisition, particularly in second language vocabulary research. For all this, I extend my greatest gratitude.

I was fortunate that Dr. Alister Cumming and Dr. Esther Geva served on my thesis committee. Their insightful comments on earlier versions of the thesis significantly improved the quality of this work. Dr. Mari Wesche, External Examiner of my thesis, made some very perceptive comments regarding my research. Dr. Jim Cummins, Department Chair of the Thesis Examination Committee, also made constructive suggestions. To them, I am thankful.

I would like to thank all the volunteer participants in my research. The present thesis would not be possible without their enthusiastic participation. At the same time, I would also like to acknowledge the critical support from Marjatta Holt and her colleagues, as well as from Wenxia Wang, who opened their classrooms to me. Without this access, I would never have been able to recruit so many ESL learners within such a short period. I am much obliged to Dr. John Read, who permitted me to use his unpublished version of the Word Associates Test, which served as a major instrument in my research.

I received support from a number of individuals as a show of friendship. A special thank-you goes to Dr. Fanmin Guo of the Educational Testing Service and Dr. Lin Wang of ACT for their professional advice on measurement and statistics. I am also appreciative of Dr. Marilyn Dickson's help in organizing the native speaker testing, and to Tom Griggs, Dr.
Helen Harper, Dr. Katherine Tiede, and Susan Washington for their assistance in rating and developing research instruments.

As always, my family's support has been instrumental in my success. I will never forget all the ardent encouragement from my parents, who believe in the meaning of a Ph.D. degree. I greatly appreciate the understanding from my wife Lili, who bore with me during those numerous weekends when I sat in front of the computer analyzing the data or writing up the thesis. My son Alan has been my loyal supporter and a driving force in the past few years. He urged me on by winning four top academic awards from his school. Talking about my family I always feel gratified.

Before winding up, I would like to extend a special gratitude to Dr. Reinhard R. K. Hartmann, who offered me my first research position as an Honorary Research Fellow at the University of Exeter, U.K. It was there I developed my knowledge base and research interest in applied linguistics. I can never thank you enough, Reinhard!

Last but not least, I am grateful to the Canadian government for awarding me four scholarships in support of my doctoral study: the Social Sciences and Humanities Research Council of Canada Doctoral Fellowship (1995-96), and the Ontario Graduate Scholarships (1992-93, 1994-95, 1996-97) from the Ministry of Education and Training of Ontario. Thank you, Canada!
CONTENTS

Abstract / ii

Acknowledgements / iv

Table of Contents / vi

List of Tables / xii

List of Figures / xiv

List of Variables in the Main Study / xv

Chapter One: Introduction / 1

Chapter Two: Background / 6

2.1 Overview / 6

2.2 Lexical Knowledge and Reading: First Language Perspectives / 7
  2.2.1 The instrumentalist hypothesis / 8
  2.2.2 The aptitude hypothesis / 8
  2.2.3 The knowledge hypothesis / 9
  2.2.4 The access hypothesis / 10
  2.2.5 Summary / 12

2.3 Lexical Knowledge and Reading: Second Language Perspectives / 13
  2.3.1 Vocabulary size and L2 reading comprehension / 13
    2.3.1.1 Threshold vocabulary for reading comprehension / 14
    2.3.1.2 Tests measuring vocabulary size / 15
    2.3.1.3 Vocabulary size as a predictor of reading success / 17
    2.3.1.4 Summary / 18
  2.3.2 Issues concerning depth of vocabulary knowledge / 18
    2.3.2.1 Criteria for definition / 18
    2.3.2.2 Empirical studies on depth of vocabulary knowledge / 22
    2.3.2.3 A working definition of depth of vocabulary knowledge / 23
    2.3.2.4 Summary / 25
  2.3.3 Factors in the vocabulary-reading comprehension chain / 25
    2.3.3.1 Nation’s vocabulary-reading comprehension triangle / 26
    2.3.3.2 A model incorporating the factor of depth of vocabulary knowledge / 27
    2.3.3.3 Summary / 30
### 2.4 Hypotheses / 31
- 2.4.1 Hypothesis One / 31
- 2.4.2 Hypothesis Two / 33

### 2.5 Lexical Inferencing / 35
- 2.5.1 Defining lexical inferencing / 35
- 2.5.2 L2 reading models / 36
- 2.5.3 Proficiency levels and strategies for lexical inferencing / 39
- 2.5.4 Knowledge sources used in lexical inferencing / 41
- 2.5.5 Summary / 44

### Chapter Three: Method / 46

#### 3.1 Overview / 46

#### 3.2 The Main Study / 46
- 3.2.1 Sample / 46
  - 3.2.1.1 Criteria / 46
  - 3.2.1.2 Procedures for recruitment / 49
  - 3.2.1.3 Profile of the Korean group / 50
  - 3.2.1.4 Profile of the Chinese group / 52
- 3.2.2 Instruments / 53
  - 3.2.2.1 Background Questionnaire / 53
  - 3.2.2.2 Vocabulary size test / 53
  - 3.2.2.3 Reading comprehension test / 55
  - 3.2.2.4 Depth-of-vocabulary-knowledge test / 55
  - 3.2.2.5 Morphological Knowledge Test / 58
- 3.2.3 Procedures for data collection / 60
- 3.2.4 Data analyses / 61
  - 3.2.4.1 Phase 1 / 62
  - 3.2.4.2 Phase 2 / 63
  - 3.2.4.3 Phase 3 / 64
  - 3.2.4.4 Phase 4 / 66
  - 3.2.4.5 Phase 5 / 66

#### 3.3 Obtaining Native Speaker Norms for the Instruments / 67
- 3.3.1 Background of the native English speakers / 67
- 3.3.2 Testing procedures / 67
- 3.3.3 Results / 68

#### 3.4 Pilot Study / 70
- 3.4.1 Sample / 70
- 3.4.2 Testing procedures / 71
- 3.4.3 Analyses and results of the pilot study / 71
3.5 The Follow-Up Study / 74
3.5.1 Participants / 74
3.5.2 Material and instruments / 75
   3.5.2.1 Experimental sentences and text / 75
   3.5.2.2 Extended Vocabulary Knowledge Scale / 75
   3.5.2.3 Interview Questions / 77
   3.5.2.4 Reading Strategies Questionnaire / 78
3.5.3 Interview Procedures / 79
3.5.4 Analyses of the Interview data / 80

Chapter Four: The Main Study: Analyses and Results / 81

4.1 Overview / 81

4.2 Phase 1 of the Analysis / 81
   4.2.1 Means and ranges for the full sample / 81
   4.2.2 Comparison of group scores / 81

4.3 Phase 2 of the Analysis / 83
   4.3.1 Reliability of RC, VS, DVK, and MK / 83
   4.3.2 Correlations / 84
   4.3.3 Procedures and results of multiple regression / 85

4.4 Phase 3 of the Analysis / 87
   4.4.1 Correlations / 88
   4.4.2 Procedures and results of multiple regression / 88

4.5 Phase 4 of the Analysis / 90
   4.5.1 Correlations / 91
   4.5.2 Procedures and results of multiple regression / 91

4.6 Phase 5 of the Analysis / 92
   4.6.1 Correlations / 93
   4.6.2 Procedures and results of multiple regression / 94

Chapter Five: The Main Study: Interpretation of Findings / 95

5.1 Overview / 95

5.2 How the Various Tests Correlated / 95
   5.2.1 Intercorrelations among RC, VS, DVK, and MK / 95
   5.2.2 Intercorrelations among RC, VS, DVKmg, DVKcol, MKaffix, and MKpart / 96
   5.2.3 Intercorrelations among RC, VS, DVKold, and DVKnew / 97
5.3 Lexical Knowledge Tests as Predictors of Reading Comprehension / 97
  5.3.1 VS, DVK, and MK as predictors of RC / 98
  5.3.2 VS, DVKmng, DVKcol, MKaffix, and MKpart as predictors of RC / 99
  5.3.3 VS, DVKold, and DVKnew as predictors of RC / 100
5.4 Comparing the Results of the Korean Group and the Chinese Group / 100
5.5 Summary / 100
  5.5.1 Addressing Hypothesis One / 100
  5.5.2 Addressing Hypothesis Two / 102

Chapter Six: The Follow-Up Study: Analyses and Results / 105
  6.1 Overview / 105
  6.2 Profile of the Sample / 106
  6.3 Selection of the Test Words for the Individual Interviews / 108
  6.4 Analysis of the Interview Data / 109
    6.4.1 Framework for analysis / 109
    6.4.2 Definition of valid attempt and rate of success in lexical inferencing / 112
    6.4.3 Interrater reliability / 115
    6.4.4 Sources of knowledge used in lexical inferencing / 117
  6.5 Use of Depth of Word Knowledge in Lexical Inferencing: Some examples / 118
  6.6 Comparing Learners with Different Depths of Vocabulary Knowledge / 122
    6.6.1 Rates of success in lexical inferencing by the HPs, IPs, and LPs / 123
    6.6.2 Knowledge sources used in lexical inferencing by the HPs and LPs / 123
  6.7 Analyses of the Data from the Questionnaire on Reading Strategies / 126
    6.7.1 How did learners approach unknown words? / 127
    6.7.2 What knowledge sources did learners use in lexical inferencing while reading? / 130

Chapter Seven: The Follow-Up Study: Interpretation of Findings / 133
  7.1 Overview / 133
  7.2 Interpreting the Results of the Interview Study / 133
    7.2.1 Success rates / 133
7.2.2 Location of clues / 134
7.2.3 Use of main knowledge sources / 136
7.2.4 Use of sources representing depth of vocabulary knowledge / 136
7.2.5 Use of interlingual knowledge in lexical inferencing / 138

7.3 Relationship between Participants’ Scores on the DVK and RC / 139

7.4 Learners’ Self-Reported Behaviours in Dealing with Unknown Words While Reading / 140

7.5 How Do the Survey Results Relate to the Findings from the Interview Study? / 140

7.6 Answering Research Question Three / 143

Chapter Eight: Implications and Conclusions / 147

8.1 Overview / 147

8.2 Summary of Major Findings from the Present Research / 147
  8.2.1 Correlation between vocabulary knowledge and reading comprehension / 147
  8.2.2 Predictive powers of vocabulary knowledge tests / 148
  8.2.3 Lexical inferencing / 149
    8.2.3.1 The rate of success and the Matthew effect in reading / 149
    8.2.3.2 Lexical inferencing strategies / 150
  8.2.4 Learners’ perceptions and actual approaches to lexical inferencing / 151

8.3 Implications for Future Research / 151
  8.3.1 Relationships among sub-components of depth of vocabulary knowledge / 151
  8.3.2 Vocabulary tests as predictors of reading comprehension levels / 152
  8.3.3 Vocabulary tests as predictors of other language skill levels / 153
  8.3.4 The threshold for lexical inferencing / 154

8.4 Implications for Second Language Education / 155
  8.4.1 Vocabulary knowledge tests for placement / 155
  8.4.2 Importance of depth of vocabulary knowledge in reading comprehension / 156
  8.4.3 Appropriate use of lexical inferencing / 157
  8.4.4 Skills in dictionary use / 158

8.5 Limitations of the Present Research / 158
  8.5.1 Population representation / 159
  8.5.2 Test reliability and accuracy / 159
8.5.3 Selection of the head words for the new part of the DVK / 160
8.5.4 The form of qualitative data collection / 160

8.6 Concluding Remarks / 161

References / 164

Appendix A: Letter of Informed Consent / 175
Appendix B: Background Questionnaire / 176
Appendix C: Depth-of-Vocabulary-Knowledge Test (Items 31-40) / 177
Appendix D: Morphological Knowledge Test / 178
Appendix E: Questionnaire on Reading Strategies / 179
Appendix F: Episodes from Individual Interviews / 180
LIST OF TABLES

3-1: Age ranges of the Korean group and the Chinese group / 51
3-2: Education levels of the Korean and Chinese ESL learners / 51
3-3: The Korean and Chinese ESL learners’ duration of stay in Canada / 51
3-4: Variables for the statistical analyses in the main study / 63
3-5: Native speaker scores on VS, DVK, MK, and RC / 68
3-6: Native speakers’ score means, standard deviations and ranges / 69
3-7: The pilot study: Learner means, standard deviations and score ranges / 72
3-8: Pilot study with ESL learners: K-R reliability coefficients / 73
3-9: Pilot study with ESL learners: Intercorrelations among the RC, VS, DVK, and MK / 73
4-1: Means, standard deviations, and score ranges on the RC, VS, DVK, and MK for the full sample / 82
4-2: Means, standard deviations and ranges of the Korean group’s test scores and the Chinese group’s test scores / 82
4-3: Test reliability coefficients in the main study and the pilot study based on K-R 21 / 84
4-4: Pearson correlation coefficients among RC, VS, DVK, and MK for the full sample / 85
4-5: Regression results of the full sample (N=74) with VS, DVK, and MK as the independent variables / 86
4-6: Pearson correlation coefficients among RC, VS, DVKmng, DVKcol, MKaffix, and MKpart for the full sample / 88
4-7: Regression results of the full sample using the forced entry procedures with VS, DVKmng, and DVKcol as the selected independent variables / 89
4-8: Pearson correlation coefficients among RC, VS, DVKold, and DVKnew for the full sample / 91
4-9: Regression results of the full sample using the forced entry procedures with VS, DVKold, and DVKnew as the independent variables / 92

4-10: Pearson correlation coefficients among RC, VS, DVK, and MK for the Korean group and Chinese group / 93

4-11: Regression results for the Korean group with VS, DVK, and MK as the independent variables / 94

4-12: Regression results for the Chinese group with VS, DVK, and MK as the independent variables / 94

6-1: Backgrounds of the participants in the follow-up study / 107

6-2: Participants' individual scores on the RC, VS, DVK, MK, & EVKS / 107

6-3: List of test words used in the individual interviews / 116

6-4: Use of knowledge sources by all participants / 117

6-5: Use of knowledge sources by the HP group (Participants #1-#4) / 124

6-6: Use of knowledge sources by the LP group (Participants #9-#12) / 125

6-7: Frequency of learners' self-reported behaviours in dealing with unknown words while reading / 128

6-8: Ranking of frequency of learners' self-reported behaviours in dealing with unknown words while reading / 130

6-9: Learners' self-reported strategies for inferencing unknown words in reading / 131

6-10: Ranking of learners' perceived strategies for inferring the meaning of unknown words in reading / 132

7-1: Questionnaire options and their corresponding knowledge sources / 141

7-2: Lexical meaning inferencing: Comparing what the learners perceived with what they actually did / 142
LIST OF FIGURES

Figure 2-1: A model hypothesizing the relationships among various factors in the chain of vocabulary knowledge-reading comprehension / 28

Figure 2-2. The hypothesized role of vocabulary knowledge and world knowledge in reading comprehension / 31
LIST OF VARIABLES IN THE MAIN STUDY

RC  Score on the Reading Comprehension Test (RC)

VS  Score on the vocabulary size test (VS), originally known as the Vocabulary Levels Test

DVK  Score on the depth-of-vocabulary-knowledge test (DVK), originally known as the Word Associates Test

DVKmng  Score on the meaning part of the DVK

DVKcol  Score on the collocation part of the DVK

DVKold  Score on the general vocabulary knowledge part of the DVK (Items 1-30)

DVKnew  Score on the RC-text-related part of the DVK (Items 31-40)

MK  Score on the Morphological Knowledge Test (MK)

MKaffix  Score on the affixation part of the MK

MKpart  Score on the part-of-speech part of the MK

L1  Learner’s first language background
CHAPTER ONE
INTRODUCTION

The present thesis research investigated the relationship between different aspects of second language (L2) vocabulary knowledge and academic reading comprehension. Specifically, the study assesses the role of breadth and depth of vocabulary knowledge in the comprehension of general academic texts in English by a group of adult Chinese and Korean speakers in two academic ESL programs in Southern Ontario.

Research on the relationship between vocabulary knowledge and reading comprehension in L2 is just beginning to gain attention. Within this area of interest, however, the two ends of the scale are still rather unbalanced: the greater part of the literature has been on how L2 learners acquire their vocabulary through reading; while, except for a very limited number of studies (e.g., de Bot, Paribakht & Wesche, 1997; Laufer, 1989, 1992a, 1992b, 1996), few studies have attempted to determine what role vocabulary knowledge plays in L2 reading comprehension in academic contexts. Even within the small number of studies that have assessed the relationship between vocabulary knowledge and reading comprehension in L2, the majority focus on the relationship between vocabulary size, or breadth of vocabulary knowledge, and reading comprehension. Little recognition is accorded to the role other aspects of vocabulary knowledge play in reading comprehension.

A word can be known in varying degrees (Baumann & Kameenui, 1991; Graves, 1984; McKeown & Beck, 1988; Nagy & Anderson, 1984). In first language (L1) research on vocabulary knowledge, various proposals have been made as to what vocabulary knowledge consists of (e.g., Cronbach, 1942; Dale, 1965). L2 researchers have also discussed what is meant
by knowing a word (e.g., Nation, 1990; Richards, 1976). The notions that have been proposed, although different in perspective, are generally complementary. In their attempt to define aspects and components of vocabulary knowledge, a number of L1 researchers have noted the importance of depth, or quality, of vocabulary knowledge in reading comprehension. Although there appears to be a virtual lack of empirical research on this topic in L1, the principle of the importance of depth of vocabulary knowledge has been generally accepted (Beck, Perfetti, McKeown, 1982; Mezynski, 1983).

However, the same cannot be said of the situation in L2 research, where there has been little recognition of the importance of depth of vocabulary knowledge in reading comprehension, let alone empirical investigations on the topic. Although researchers (e.g., Read, 1989, 1993, 1995; Wesche & Paribakht, 1994) have been developing instruments to measure depth of vocabulary knowledge, only one known study (de Bot, Paribakht & Wesche, 1997) has, in some way, linked analyses about depth of vocabulary knowledge to reading comprehension. Many questions still remain unanswered.

In addition to the foregoing theoretical concern, there is a practical need motivating the present research. Each year witnesses a continuous influx of international students into English as a second language (ESL) schools in English-speaking countries. Many come to ESL programs to prepare themselves for university studies in English-speaking institutions. In Canada, demand for short-term intensive academic ESL programs has been strong. In response to this demand, a large number of universities, colleges, and private language schools have set up short-term programs specifically catering to these academic-oriented learners.

In Ontario, the typical cycle of this type of program lasts four to eight weeks, which
means a high turnover of learners. As a result, there is a high demand for assessment of newcomers to the programs. Traditional assessments are mostly conducted by administering a placement test composed of a battery of tests on different aspects of ESL skills and knowledge, such as listening comprehension, reading comprehension (especially of general academic texts), grammar, vocabulary, etc. Administration of these tests is normally time-consuming but is required by most of the ESL programs in Ontario as their only method of placing newcomers in appropriate classes. The practice has resulted in great pressure on teachers as well as administrators, since administering and marking the tests requires substantial time.

The pressure of placement testing on short-term ESL programs has, in fact, been an international problem. In describing this problem in the Eurocentres, a consortium of private language schools in Europe, Meara and Jones (1988) note:

The main problem with tests of this sort is that they take a long time to administer and mark. In a situation where time is at a premium because classes cannot be started until the placement procedure is completed this is obviously a serious shortcoming (p. 80).

To overcome this shortcoming, researchers in ESL testing and assessment have been exploring alternative ways of learner assessment. Most notably, Meara and his colleagues (see Meara, 1996; Meara & Buxton, 1987; Meara & Jones, 1987, 1988) have developed a simple form of placement test, the Eurocentres Vocabulary Test, for the Eurocentres schools. According to Meara and Jones (1988), the test, which normally takes about 15 minutes to complete, has shown promise as an indicator of learners' ESL proficiency. Based on signal detection theory, the Eurocentres Vocabulary Test requires only "yes" or "no" answers from the test-taker. This computerized test was designed to measure the vocabulary size of the test-taker instead of several aspects of the learner's English, but it still serves the purpose of being a reliable
placement test. Referring to Anderson and Freebody (1981), Meara and Jones (1988) argue that the reason for the test's effectiveness is "that vocabulary knowledge is heavily implicated in all practical language skills, and that in general, speakers with a large vocabulary perform better on a wide range of linguistic indicators than speakers with a more limited vocabulary" (p. 80).

Even though its scope is limited to the size of vocabulary, or breadth of vocabulary knowledge, the assessment research carried out by Meara and his colleagues has been notable among very few documented attempts to describe and justify the role of vocabulary knowledge in relation to other aspects of language skills and knowledge. The general lack of such research is especially evident with respect to the relationship between vocabulary knowledge and academic reading comprehension. However, as was discussed earlier in this chapter, since there is more to vocabulary knowledge than breadth, vocabulary size tests may not be the only feasible solution to the placement problem. One question, for example, is whether depth-of-vocabulary-knowledge tests are also a feasible solution? If so, what type of depth tests would be most effective, especially in indicating learners' levels of academic reading comprehension? Questions such as these give rise to a strong need for further exploration of the relationship between vocabulary knowledge and academic reading comprehension since academic ESL programs in English-speaking countries continue to be a popular choice for many international students.

The purpose of the present research, accomplished through a main study composed of a portfolio of written tests and a follow-up study consisting of semi-structured individual interviews and a questionnaire on reading comprehension strategies, was to assess the relationships among variables involved in the L2 reading process from the angle of vocabulary
knowledge in order to answer the following three research questions:

1. How do scores on vocabulary size, depth of vocabulary knowledge, and reading comprehension correlate with each other?

2. To what extent does depth of vocabulary knowledge add to the prediction of reading comprehension scores over and above the prediction afforded by vocabulary size?

3. How do ESL learners make use of their depth of vocabulary knowledge in the process of comprehending a general academic text?

The thesis is composed of eight chapters. Chapter 2 discusses the research background and reviews the relevant literature on the role of vocabulary knowledge in reading comprehension. In this chapter, influential L1 theories related to vocabulary knowledge and reading comprehension are examined. Recent findings from empirical studies concerning this theme in the L2 context, with a focus on assessment issues related to depth of vocabulary knowledge, are discussed. The notion of depth of vocabulary knowledge is analyzed. This provides the conceptual framework for the present research. Chapter 3 presents an overview of the design and methods of the present research. The chapter also details how the quantitative and qualitative data were collected and how two trials were conducted to test the instruments before the main study got under way. Chapter 4 describes the procedures for the quantitative data analyses in the main study and results of the analyses. Chapter 5 discusses the findings reported in Chapter 4. In Chapter 6, the procedures for analyzing the qualitative data and the results from the follow-up study are presented. Chapter 7 discusses and interprets the findings reported in Chapter 6. In drawing conclusions from the research, Chapter 8 considers some implications from the research for future research and educational practice.
CHAPTER TWO

BACKGROUND

2.1 Overview

Research on the relationship between L2 vocabulary knowledge and reading comprehension in a second language is a topic currently drawing the attention of researchers, though only after a long dearth of research, owing largely to the influence of the syntactically-oriented trends of linguistic theory (Richards, 1976). Audio-lingual teaching methods, which surfaced in the 1940s and dominated the L2 scene for more than two decades, favoured limited purposeful vocabulary learning, assuming that “good language habits, and exposure to the language itself, would eventually lead to an increased vocabulary” (Coady, 1993, p. 5). The next dominant theory, the communicative approach, came to the L2 scene in the 1970s and is still alive and well today. The communicative approach does not see the revival of vocabulary instruction in the classroom as a major concern, contending that L2 acquisition is analogous to first language (L1) acquisition and that vocabulary will be naturally acquired in the process (Coady, 1993).

It is only in recent years that vocabulary has reclaimed its pre-1940s status as an important focus in the L2 classroom. In the past few years, researchers in L2 education have begun to direct their attention to various issues related to L2 vocabulary. More research papers on vocabulary are emerging (Qian, 1993). However, most recent vocabulary studies focus on vocabulary learning. There have been relatively few studies exploring the role of vocabulary knowledge in the measurement of communicative competence.

This chapter examines the literature pertaining to the relationship between vocabulary knowledge and reading comprehension with particular attention to performance measurement in
order to provide a theoretical framework for the present research. The chapter first examines relevant theoretical positions taken by L1 researchers, and then considers their influence on L2 research in the area of vocabulary and reading comprehension. Arising from the available research literature on L2 vocabulary, a number of issues relevant to the present research are discussed in detail. First, the concept of two components of vocabulary knowledge, vocabulary breadth and vocabulary depth, are considered and defined. Second, the role of each component and the relationship between them is explored in the context of reading comprehension. Third, some existing measures of vocabulary knowledge are reviewed. This review serves as an assessment leading to the selection of data-collection instruments for the present research, especially for the main study. Finally, available studies on lexical inferencing in L2 are reviewed. This review provides a conceptual basis for the follow-up study, which is the second component of the present research.

2.2 Lexical Knowledge and Reading: First Language Perspectives

Studies concerned with the role of vocabulary knowledge in L1 reading comprehension abound (e.g., Anderson & Freebody, 1981, 1983; Beck, Perfetti, & McKeown, 1982; Kameenui, Carnine, & Freschi, 1982; Mezynski, 1983; Stahl, 1983, 1985, 1986); the most influential of these has been Anderson and Freebody's (1981) analysis of the relationship between knowledge of word meanings and level of reading comprehension. Summarizing earlier research, Anderson and Freebody identified three distinct positions on the role of vocabulary knowledge in reading comprehension, namely, the instrumentalist hypothesis, the aptitude hypothesis, and the knowledge hypothesis.
2.2.1 The instrumentalist hypothesis

The instrumentalist hypothesis sees vocabulary knowledge as a major prerequisite to reading (Anderson & Freebody, 1981). According to the instrumentalist view, individuals who perform better on a vocabulary test probably understand more of the words in texts they read than individuals who score lower do. The central idea of this hypothesis is straightforward: knowing the words enables reading comprehension. In other words, this position claims that vocabulary knowledge is a direct factor in the causal chain resulting in reading comprehension. The educational implications of the instrumentalist hypothesis are apparent: in order to improve students' reading ability, vocabulary teaching should be regarded as a priority in the curriculum. The larger a student's vocabulary, the better he or she will understand a text.

2.2.2 The aptitude hypothesis

The aptitude hypothesis (Hunt, 1978) claims that achievement on a vocabulary test is indicative of an individual's verbal aptitude: A person scoring high on a vocabulary test should be mentally more agile. The heart of the aptitude hypothesis is that the direct and key factor which results in an individual's better reading comprehension is the person's mental capacities rather than his or her large vocabulary. Other things being equal, an individual with superior mental agility should be able to acquire a larger vocabulary than his or her peers. Likewise, this same person should also be better at written discourse comprehension as a result of superior mental capacities. In this model, a large vocabulary is not considered making a direct contribution to better reading comprehension: A person with a larger vocabulary is better at comprehension because of his or her mental agility, and the large vocabulary this person has is just a reflection

The pedagogical implication of the aptitude hypothesis is rather discouraging: because individuals' abilities in reading are directly based on their verbal aptitude, reward for vocabulary training will be very limited. Although it is possible to improve students' vocabulary knowledge through training (Mezynski, 1983), such training alone will not likely result in a significant change in ranking in terms of students' reading achievements if all the students undergo the same amount of training.

2.2.3 The knowledge hypothesis

The third position, the knowledge hypothesis, sees a person's scores on vocabulary tests as an indicator of the person's knowledge of the world. Consequently, a person with a higher score should have deeper and broader knowledge of the world than a person with a lower score. The central idea of this position is that it is this background knowledge that is crucial for reading comprehension. In this model, knowing a word's meaning is merely a sign that the person is likely to possess the background knowledge needed to comprehend a written discourse. Although a strong case has now been made by Anderson (1978) that background knowledge is crucially important for reading comprehension, there is little evidence in support of the argument that scores derived from vocabulary tests reflect such background knowledge.

The knowledge hypothesis conforms to schema and constructivist theories (Kintsch & van Dijk, 1978; Rumelhart, 1977, 1981), which describe the process of reading comprehension as an interaction of conceptually-driven, or top-down, and data-driven, or bottom-up, activities. In these models, conceptually-generated knowledge of the world interacts actively with
information from the written text obtained through the eye. Depending on the context, comprehension can be activated either by world knowledge or by written symbols (Rumelhart, 1981). The pedagogical implication of the knowledge hypothesis is that vocabulary should be taught in context and in respect to the acquisition of specific domains of knowledge.

Specifically, as Mezynski (1983) puts it:

New vocabulary should be taught in the context of learning new subject matter. In this way, word meanings can be related to one another, and where possible, to information already possessed by the learner. According to the knowledge position, if students are taught groups of words that are semantically unrelated, the students may learn definitions but fail to learn where the word fits in with their store of related knowledge (p. 255).

The above three hypotheses on the role of vocabulary knowledge in reading comprehension are derived from different theoretical perspectives. None of them can claim to be the only correct position, and "no serious scholar in reading or related fields rigidly adheres to any of these positions" (Anderson & Freebody, 1981, p. 89) because they in fact do not contradict one another. However, it is difficult to identify which of the three is most tenable since "there are neither theoretical tools nor hard data to justify a conclusion to date" (Anderson & Freebody, 1981, p. 82).

2.2.4 The access hypothesis

Even though the instrumentalist view seems to be the most unequivocal of the three positions because of its straightforwardness, studies trying to prove this theory have produced conflicting results. While results of a number of studies (Chall, 1958; Wittrock, Marks, & Doctorow, 1975) are in favour of this position, others were unable to prove its soundness (Jenkins, Pany, & Schreck, 1978; Pany & Jenkins, 1977; Tuinman & Brady, 1974). Tuinman and Brady (1974)
rejected the instrumentalist hypothesis when an experiment involving sixteen 45-minute periods of vocabulary instruction failed to improve fourth, fifth and sixth grade students’ comprehension of three passages containing a large proportion of difficult words that they had been directly taught. Jenkins, Pany, and Schreck (1978) and Pany and Jenkins (1977) also failed to establish that vocabulary instruction could help increase their students’ story comprehension after experimenting with a number of different teaching methods to no avail.

Having reviewed eight earlier experimental studies on this topic, Mezynski (1983) concluded that it is not the instrumentalist hypothesis *per se* that is to blame. What should be noted is that while vocabulary knowledge can be a major prerequisite for comprehension, a link between the two variables, i.e., vocabulary knowledge and reading comprehension, is missing. This link Mezynski identified as effective practice. She found that studies with results in favour of the instrumentalist hypothesis contained carefully designed activities to practise the target words, whereas those with findings that failed to support the instrumentalist hypothesis put less or no emphasis on practice. Mezynski (1983) noted that through vocabulary instruction "students can often improve their performance on vocabulary tests yet fail to improve on reading comprehension tests" (pp. 275-276), which suggests that "automaticity of access may be especially important for successful comprehension" (p. 276). The importance of rapid access to word meanings has been substantiated and stressed in several studies (LaBerge & Samuels, 1974; Lesgold & Perfetti, 1978; Perfetti & Lesgold, 1977). Mezynski (1983) summarizes them in the following way:

To the extent that reading processes share a limited resource pool, skill in reading requires that some component processes occur automatically (or with low resource demands). If all words in a text are familiar to the reader, processing can focus on extracting the overall meaning, and other "higher order" activities. The
more interruptions that occur because of unknown or incomplete word meanings, the greater this will interfere with text processing (p.262).

On this theoretical basis, Mezynski established a fourth hypothesis, the access view. This hypothesis regards verbal aptitude "as an amalgam of potentially trainable subskills" (Mezynski, 1983, p. 254): "In reading comprehension, these might be skills involved in accessing word meanings and using those meanings efficiently in text processing" (p. 254). The essential claim of this position is the need for automaticity of access to vocabulary knowledge as a prerequisite for reading comprehension. Pedagogically, the access view implies that sufficiency of practice is crucial in vocabulary acquisition, and that unless taught words truly become part of the students' lexical repertoire these words will not be really useful in reading comprehension (Mezynski, 1983).

2.2.5 Summary

The four foregoing positions take different perspectives on the issue of how vocabulary knowledge relates to reading comprehension. The instrumentalist hypothesis, which regards good vocabulary knowledge as the primary factor in successful reading comprehension, "suggests that what is important is the number of words taught" (Mezynski, 1983, p. 255). The knowledge hypothesis, based on a schema-theoretic and constructivist perspective, sees vocabulary as indicative of world knowledge, in which case knowledge of individual word meanings is no longer the primary cause of successful comprehension; rather it is the underlying knowledge, normally referred to as background knowledge or world knowledge, that leads to comprehension. The aptitude hypothesis considers vocabulary knowledge and reading comprehension to be two unrelated outcomes of mental aptitude. The access hypothesis recognizes an important
relationship between vocabulary and reading comprehension, given that the vocabulary is easily accessible through effective training. Apparently, there is strength in the four hypotheses as a combined whole. However, one must be wary of accepting them as a package without screening. For instance, the excessive emphasis of the instrumentalist hypothesis on vocabulary knowledge as a direct factor in the causal chain in reading comprehension could prematurely reduce investigation of the possible effect of other factors as a result.

2.3 Lexical Knowledge and Reading: Second Language Perspectives

Compared to the substantial L1 research on the role of vocabulary knowledge in reading comprehension, L2 research on this topic is only just emerging, and the research literature is still amazingly scant. However, within the small body of literature, a few studies seem to be of relevance to the present research.

In the discussions to follow, it is important to distinguish two dimensions of vocabulary knowledge, namely, breadth and depth. In the present research, breadth of vocabulary knowledge is defined as vocabulary size, or the number of words for which a learner has at least some minimum knowledge of meaning. Depth, or quality, of vocabulary knowledge, is provisionally defined as one's level of understanding of various important aspects of a word. A more detailed working definition of depth of vocabulary knowledge appears later in this chapter (see Section 2.3.2.3).

2.3.1 Vocabulary size and L2 reading comprehension

Two important issues arising from research in this line of inquiry are: (a) What is the threshold
level of vocabulary size for minimum proficiency in academic reading comprehension? and (b) does vocabulary size serve as an effective predictor of academic reading comprehension?

2.3.1.1 Threshold vocabulary for reading comprehension. There has been some research in the past decade on the relationship between vocabulary size and L2 reading comprehension. Liu and Nation (1985) investigated this issue with 59 ESL teachers attending a diploma course in New Zealand. The great majority of these teachers were ESL learners themselves. These teachers were asked to infer the meaning of vocabulary in context under two conditions: One passage had a maximum of 96% vocabulary coverage for the participants (the other 4% words in the passage were nonsense words), and the other passage had a maximum of 90% lexical coverage for the participants (there were about 10% nonsense words in the passage). Liu and Nation found that the density of unknown words in a text affected the success rate of guessing from context: the fewer the unknown words in a passage, the higher the success rate of guessing. This finding suggests that vocabulary size affects one's ability to guess from context, and that lexical guessing from a text of 95% lexical coverage should yield more satisfactory results than performing a similar task with a text of 90% lexical coverage.

Laufer's (1989, 1992a, 1996) research with first year university students, whose native language background was Hebrew or Arabic, demonstrated that 95% lexical coverage should be the minimum required level for L2 learners to reach an adequate comprehension level (a reading comprehension score of 56% and above) in academic reading. Laufer (1992b, 1996) found that a vocabulary of 3,000 word families, which should normally provide 90% to 95% lexical coverage of a text, was the normal cutoff for L2 learners being able or unable to use their
knowledge of the world to compensate for their vocabulary size to help comprehension.

According to Nation (1993), when each different word form is counted as a different word, 5,000 words should provide lexical coverage of over 95% of academic texts in English, and when inflected forms and closely related derivatives are considered as one word type, these 5,000 word forms will translate into around 3,000 word families. Since Laufer adopted Nation’s concept of “word” in her studies, it can therefore be concluded that 3,000 word families, or about a 95% lexical coverage, should be the threshold vocabulary for adequately coping with academic texts in English.

2.3.1.2 Tests measuring vocabulary size. To date, there have been two types of widely known tests that measure vocabulary size, or breadth of vocabulary knowledge, in ESL. One is Nation’s (1983) Vocabulary Levels Test, and the other is The Eurocentres Vocabulary Tests developed by Meara and his colleagues (Meara & Buxton, 1987; Meara & Jones, 1988). Both types of test were developed to serve the needs of ESL practitioners and researchers for a reliable but simple measure of vocabulary knowledge for learner assessment, and both have achieved some success in serving this purpose. Nation’s (1983) test is a paper-and-pencil test with testing items evenly distributed at five difficulty levels (2,000 words, 3,000 words, 5,000 words, university vocabulary, and 10,000 words). At each level, there are six items. For each item, the test-taker is required to match three meanings with three of the six words listed (see Section 3.2.2.2 for more detailed description of the test). Nation’s test has been accepted by a number of L2 researchers as a reliable measure of vocabulary size (e.g., Laufer, 1992a, 1996; Yu, 1996). The test is easily accessible as it is appended to two publications (Nation, 1983, 1990).
The Eurocentres Vocabulary Tests are based on signal detection theory in psychology (Anderson & Freebody, 1983; Zimmerman, Broder, Shaughnessy & Underwood, 1977). The format of the Eurocentres tests is simpler than the Vocabulary Levels Test in that it only requires "Yes" or "No" from the test-taker. Each test contains two types of words: 60 real English words and 40 pseudo-words that appear in English lexical patterns. The tests are available as computerized tests as well as paper-and-pencil tests. According to Meara (1996) and Meara and Jones (1988), the Eurocentres tests have demonstrated themselves as reliable measures of vocabulary size. Other L2 researchers (e.g., Harley, Howard, & Roberge, 1996; Lessard-Clouston, 1996) have also used the format of these tests in their studies to develop vocabulary measures for their specific purposes. More details about the tests and the rationale behind the format are discussed in a number of articles (Meara, 1996; Meara & Buxton, 1987; Meara & Jones, 1987, 1988; Meara, Lightbown, & Halter, 1994). Despite the promising signs of the Eurocentres tests, however, some problems have emerged. For example, Meara (1996) found that the tests worked better with certain L1 groups than others. For groups whose L1 was closely cognate with English, the results were particularly problematic. It was also reported that the tests did not produce satisfactory results either from low proficiency L2 learners (Meara, 1996) or from advanced L2 learners (Wesche, Paribakht, & Ready, 1996).

Although both the Vocabulary Levels Test and the Eurocentres tests serve the function of measuring vocabulary size, there is a difference between the two types of test in that, because of their different formats and requirements, the former requires a clear focus on meaning while the latter only requires some reflective recognition of the word form. In other words, Nation’s Vocabulary Levels Test measures more depth of lexical knowledge than the Eurocentres tests.
2.3.1.3 Vocabulary size as a predictor of reading success. Recently, a number of studies (e.g., Koda, 1989; Laufer, 1992a, 1996) have used scores on vocabulary size to predict levels of academic reading comprehension. Laufer (1996) found "good and significant correlations" (p. 55) between the above vocabulary size tests and reading comprehension tests in her studies. In one study with 92 first-year university students whose native language was either Hebrew or Arabic (Laufer, 1992a), the correlation between the scores on the Vocabulary Levels Test (Nation, 1983) and scores on reading comprehension was .50, and that between the scores on a Eurocentres Vocabulary Test (Meara & Jones, 1989) and scores on reading comprehension was .75. In another study involving 80 first-year university students of similar L1 backgrounds (Laufer, 1996), a correlation of .71 was reported between students' scores on reading comprehension and those on the Vocabulary Levels Test.

Koda's (1989) study of 24 college students who were learning Japanese as a foreign language found equally strong correlations between a self-made vocabulary test and two reading tests, one being a cloze test and the other paragraph comprehension. Koda reported a correlation of .69 between the learners' scores on the vocabulary test and the cloze test, and a correlation of .74 between their scores on the vocabulary test and the paragraph comprehension test.

Research by Coady, Magott, Hubbard, Graney, and Mokhttari (1993) with 79 students studying English in a university academic preparation program found that two experimental groups, which had received special training in high frequency vocabulary, achieved better ESL reading comprehension at the end of the experiments than did a control group which had not received such treatment. The study was carried out to verify the proposition that "there is a positive and significant relationship between knowledge of high-frequency vocabulary and
reading proficiency" (p. 217). Based on the results of their study, Coady and his colleagues argued that special training in the 2,000 most frequent English vocabulary items could improve learners' reading proficiency. However, they were cautious about reaching a conclusion on the relationship between vocabulary and reading proficiency, acknowledging a need for further investigation in this area.

2.3.1.4 Summary. It is apparent that, within the topic under discussion, the theoretical foundation on which L2 vocabulary research is developing has largely been that of L1 research. L2 research on the relationship between vocabulary size and reading comprehension is ongoing. Nevertheless, existing research in L2 has somehow failed to give adequate attention to another aspect which may be equally important, depth of vocabulary knowledge, even though Anderson and Freebody (1981) and Mezynski (1983) specifically noted the importance of this aspect while introducing the four hypotheses on the vocabulary-reading comprehension connection.

2.3.2 Issues concerning depth of vocabulary knowledge

2.3.2.1 Criteria for definition. In L1 research, the importance of depth of vocabulary knowledge has been widely recognized. Research by Beck, Perfetti, and McKeown (1982) indicates that lexical automaticity without some depth of vocabulary knowledge will not necessarily help understand texts. Mezynski (1983) also notes:

Word meanings can be "known" to varying degrees. Depending on the task, a person could perform adequately with relatively imprecise knowledge. In other situations, a much finer notion of the word's meaning might be required (p. 265).

Indeed, although sometimes L2 learners only need partial knowledge of a word in
comprehension, more lexical knowledge is obviously desirable in many situations. How much knowledge is deemed sufficient, then, and what is it? It is relatively easy to reel off a list of aspects: pronunciation, spelling, multiple meanings, connotations, inflections and derivations, syntactic properties, appropriate uses, collocations, semantic associations, and idioms containing the target words. The list can go on and on. It is not surprising that, more often than not, even an advanced L2 user’s knowledge of a given word can be superficial and incomplete.

Over the years, lexical researchers have developed various criteria for measuring word knowledge. Cronbach (1942) provided five categories:

- generalization: being able to define the word;
- breadth of meaning: recalling the different meanings of the word;
- precision of meaning: applying the word correctly to all possible situations;
- availability: being able to use the word productively;
- application: selecting an appropriate use of the word.

Here, the focus appears to be on meaning and use. Cronbach’s categories can be basically divided into two main categories: knowledge of word meaning (generalization, breadth of meaning, and precision of meaning) and levels of accessibility to this knowledge (availability and application). The obvious weakness in the criteria is the lack of a place for other aspects of lexical knowledge, such as spelling, pronunciation, and morpho-syntactic properties.

Richards (1976) offered eight assumptions on what is involved in knowing a word. These assumptions of lexical knowledge are much more inclusive than the framework defined by Cronbach, as Richards not only incorporated morphological and syntactic properties into the concept, but also considered such aspects as word frequency and register characteristics.
However, pronunciation, spelling, and collocations seemed to be the obvious missing aspects in the framework. His eight assumptions emphasize that to know a word, one should be aware of:

- its relative frequency in the language;
- its register characteristics, which may include social, temporal and geographic variations, and field and mode of discourse;
- the syntactic behaviour associated with the word;
- its underlying form and the derivations that can be made from this form;
- the network of associations between that word and other words in the language, which may include such associative links as antonymy, synonymy, and subordinate, coordinate, and superordinate classifications;
- its semantic features and connotations; and
- the different meanings associated with the word (pp. 78-84).

In the form of a table, Nation (1990, p. 31) proposed a set of 16 questions to reflect four dimensions of a learner’s vocabulary knowledge. These questions, largely reflecting Richards’ (1976) foregoing framework, are classified as follows:

- form: spoken form and written form
- position: grammatical patterns and collocations
- function: frequency and appropriateness
- meaning: concept and association

The strength of Nation’s (1990) classification is that it has not only retained the primary aspects of Richards’ (1976) definition, but also added such important aspects as spoken form, written form, and collocations, which were missing elements from Richards’ otherwise
comprehensive framework. Meanwhile, the plain language Nation used to compose the 16 questions under the above four categories makes the questions easily understandable not only to L2 researchers but also to ESL teachers and learners.

In L1 research, it has been well accepted in principle that a word can be known at different levels (Baumann & Kameenui, 1991; Graves, 1984; McKeown & Beck, 1988; Nagy & Anderson, 1984). However, researchers differ in opinion as to how to define these levels and what are workable ways to differentiate them. Various approaches have been suggested for capturing the features capable of distinguishing one level from another. Dale (1965, p. 898) established a four-level continuum to assess knowledge of a word in research with L1 users who were fifth grade students in the U.S.:

Stage 1: "I never saw it before."
Stage 2: "I've heard of it, but I don't know what it means."
Stage 3: "I recognize it in context - it has something to do with ..."
Stage 4: "I know it."

Beck, McCaslin, and McKeown (1980) recommended scaling the quality of lexical knowledge in terms of three levels of lexical access: unknown, acquainted, and established word knowledge. Stahl (1985, 1986) proposed an intuitive three-level scale of the processing of lexical knowledge in an ascending order of difficulty: association, comprehension, and generative processing. Kameenui, Dixon, and Carnine (1987) introduced a three-level continuum of lexical knowledge in a descending order of difficulty: full concept, partial concept, and verbal association. Although no conclusions have been reached as to which is the best approach, all these approaches imply that depth of lexical knowledge is not a solid mass whose qualities
cannot be differentiated. In other words, it is feasible to characterize different levels of lexical knowledge. The methods mentioned above suggest possible ways of approaching the problem and, in fact, are complementary and have a great deal in common.

2.3.2.2 Empirical studies on depth of vocabulary knowledge. To date, there are no empirical results available on how depth of vocabulary knowledge functions in its own right in L2 reading comprehension, although some related areas have recently been touched upon: Laufer (1992b) has looked at how world knowledge compensates for limited vocabulary in reading comprehension, and a number of other studies (Read, 1988, 1989, 1993, 1994, 1995; Paribakht & Wesche, 1993; Wesche & Paribakht, 1993, 1996) have explored the issue of measuring the depth of vocabulary knowledge in an L2.

In answer to the need for assessment procedures, Read (1989) attempted to set up an interview protocol for the purpose of assessing the depth of vocabulary knowledge of adult ESL learners in New Zealand. The interview questions (Read, 1989, p. 13) covered a wide range of aspects such as:

- the ability to pronounce the word;
- explanation of what the word means;
- identification of the field(s) of study to which the word belongs;
- the naming of other words with which the given word collocates or is associated;
- knowledge of other forms of the word.

Read's (1989) study provided many valuable insights into designing and conducting interviews of this nature. However, due to the lack of previous research in the area, his study
was basically exploratory and unable to establish any workable measures or procedures. However, Read did not stop there. He continued researching along this line (Read, 1993, 1994, 1995) and has in recent years developed an innovative test on depth of vocabulary knowledge, known as the Word Associates Test, which was used as an important instrument in the present research (see 3.2.2.4 for details).

More recently, Wesche and Paribakht (1996) have devised a Vocabulary Knowledge Scale (VKS), and used it in a pilot study on the vocabulary development of students in a comprehension-based intermediate ESL program at the University of Ottawa (Paribakht & Wesche, 1993). The VKS includes a five-level elicitation scale for self-reporting and demonstrating vocabulary knowledge, and a corresponding five-level set of scoring categories (see Wesche & Paribakht, 1996, p. 30, Figures 1 & 2, for more details of the VKS and its scoring categories). The ratings of the VKS "range from complete unfamiliarity, through recognition of the word and some idea of its meaning, to the ability to use the word with grammatical and semantic accuracy in a sentence" (Wesche & Paribakht, 1996, p. 29). According to Paribakht and Wesche (1993), the VKS is capable of measuring progress in the developing knowledge of particular words and of showing intra-group change as well as inter-group differences in gains of content vocabulary resulting from a brief instructional period. However, the scale in its present form does not purport to be useful for estimating general lexical knowledge (Paribakht & Wesche, 1993), nor has it been employed to determine the relationship between vocabulary knowledge and reading comprehension, although presumably it could be.

2.3.2.3 A working definition of depth of vocabulary knowledge. The foregoing definitions
of what is involved in knowing a word can be categorized into three types. The first is basically component-focused (e.g., Richards, 1976), the second is largely based on the learner's familiarity with a given word (e.g., Dale, 1965; Paribakht & Wesche, 1993; Wesche & Paribakht, 1996), and the third is a combination of the above two (e.g., Cronbach, 1942; Nation, 1990; Read, 1989). Whichever one may support, it is evident that there is a consensus among these researchers that "knowing a word involves more than being able to recall the meaning (or L1 translation) of a presented word form" (Nation, 1994, p. 121). In order to provide a feasible framework for the present thesis research, a working definition of depth of vocabulary knowledge is proposed below. Based on the literature reviewed in Section 2.3.2, on the definitions of relevant terms by Hartmann and Stork (1972) and Richards, Platt and Platt (1992), and particularly on the definitions proposed by Richards (1976), Nation (1990) and Read (1989) concerning what is involved in knowing a word, the following are considered main aspects of the depth of knowledge of a word relevant to reading comprehension:

- Pronunciation and spelling: This involves knowing how different forms of the word are pronounced and spelled;
- Morphological properties: This involves knowing the word's stem, its capability of inflection, derivation, and other word formation devices, and its possible parts of speech;
- Syntactic properties: this involves knowing the word's possible positions and its syntagmatic relations, including collocational relations, with other words in a sentence;
- Meaning: This not only involves identification of the denotative meaning of a word
in context, but also, where applicable, knowledge of connotations, as well as polysemy, antonymy, synonymy and other paradigmatic relations the word may have;

- Register, or discourse features: These include possible adherence to a stylistic, social or regional variety, and the field, mode and manner of discourse concerning the application of the word;

- Frequency of the word in the language, or whether this word is a commonly used word or a rarely used word only appearing in some specialized texts.

2.3.2.4 Summary. Although the importance of depth of vocabulary knowledge has been stressed by vocabulary researchers and some progress has been made, work still needs to be done in exploring the connection between vocabulary knowledge and reading comprehension in L2. As Read (1988) notes:

> [W]e need to concentrate more on depth of knowledge: how well are smaller sets of key vocabulary items known? The problem is that we currently lack procedures for measuring the depth of vocabulary knowledge of second language learners. This is an area in which research is just beginning. (pp. 16-17)

In this section, earlier definitions of what is involved in knowing a word have been listed, recent empirical studies on assessing depth of vocabulary knowledge in L2 have been reviewed, and a working definition of depth of vocabulary knowledge, from the perspective of word properties, has been proposed. The design of the present thesis research generally follows this definition as a guideline in the development of methodology and procedures.

2.3.3 Factors in the vocabulary-reading comprehension chain

The instrumentalist, aptitude, access, and knowledge hypotheses all have important implications
for L2 curriculum, as they do for L1 pedagogy, although refinement of each and improvement of the framework as a connected whole are still desirable.

2.3.3.1 Nation's vocabulary-reading comprehension triangle. Nation (1993), who examined the four L1 vocabulary-reading hypotheses in relation to L2, regarded verbal aptitude as instructionally irrelevant but identified vocabulary size, language skill, and world knowledge as three main factors involved in L2 reading comprehension. Nation proposed a triangular diagram (see Nation, 1993, p. 117 for details) to describe the relationship between these three factors. According to Nation (1993), the relationships among these factors are dynamic: They constantly change depending on the relative strength of each factor. Skill in language use, which includes reading comprehension, is dependent upon vocabulary size: "In order to be successful in academic studies, it is necessary to be familiar not only with the high frequency words of English but also with the general academic vocabulary that is common to many academic disciplines" (p. 120); "knowledge of the world depends on skill in language use" (p. 120); and "vocabulary growth is affected by knowledge of the world" (p. 121).

Nation's triangle deserves some criticism. The question of how vocabulary growth will be affected by knowledge of the world is beyond the scope of the present discussion, but several other points are worth noting. With respect to the effect of vocabulary size on reading comprehension, Laufer (1992a) cautions that, even though the results of her studies (Laufer, 1989, 1992a) have shown high correlations between vocabulary size and reading comprehension, this nevertheless does not imply a causal relationship between the two factors. The level of a learner's reading ability is not necessarily a direct result of his or her vocabulary size, even
though it is feasible to predict a reader's comprehension level by measuring his or her vocabulary size. This is a conclusion reached by earlier studies on L1 as well (Davis, 1968, 1972; Thorndike, 1973; Spearritt, 1972), which recognize vocabulary, among all factors involved in reading comprehension, as having the most important and strongest predictive value.

In spite of agreeing with Nagy that caution should be exercised about adopting the instrumentalist hypothesis as "it can lead to a reductionist view of reading, i.e., if you know the vocabulary then that is about all you need to be able to read" (Nation 1993, pp. 115-116), Nation still seems to have neglected this very warning when he posits that "learners' skill in use depends on the size of their recognition vocabulary" (Nation 1993, p. 117). It is apparent that Nation's position here is akin to a reductionist view of reading.

What seems also to have been neglected in Nation's discussion is the important dimension of depth, or quality, of vocabulary knowledge. Unlike L1 vocabulary researchers such as Anderson and Freebody (1981) and, particularly Mezynski (1983), who all emphasized, to varying degrees, the importance of the quality of readers' vocabulary knowledge as opposed to simply concentrating on the size of their vocabularies, Nation, in stating his position on the relationship between vocabulary size, skill in use, and world knowledge, totally ignored the dimension of depth of vocabulary knowledge, which may be equally responsible for affecting reading comprehension. Without recognition of this dimension of vocabulary knowledge, Nation's position is unconvincing.

2.3.3.2 A model incorporating the factor of depth of vocabulary knowledge. Taking into account the potential importance of depth of vocabulary knowledge in reading comprehension,
the following model (Figure 2-1) is proposed. This model, founded on the four vocabulary knowledge-reading comprehension hypotheses (namely, the instrumentalist, knowledge, aptitude and access hypotheses), identifies the following as important learner factors in the vocabulary knowledge-reading comprehension chain: breadth and depth of vocabulary knowledge, automaticity of access to vocabulary knowledge, other linguistic knowledge, world knowledge, reading comprehension, and cognitive and affective variables.

Figure 2-1. A model hypothesizing the relationships among various factors in the chain of vocabulary knowledge-reading comprehension.
The model emphasizes the following aspects. First, it defines vocabulary knowledge as a combination of *vocabulary size* and *depth of vocabulary knowledge*. These two dimensions are interactive and interdependent; in exploring the role of vocabulary knowledge in reading comprehension, these two dimensions deserve equal attention. In addition, other linguistic knowledge, most prominently that of sentence grammar, also plays an indispensable role in the comprehension process.

Second, the model recognizes automaticity of access to vocabulary knowledge as an important factor. However, the kind of vocabulary knowledge automatically accessed should have depth as well as breadth. If automaticity involves only arriving at simple and discrete meanings of lexical items without any supporting depth of vocabulary knowledge, this automaticity would be theoretically questionable.

Third, the importance of world knowledge, or background knowledge, cannot be overemphasized: "The term background knowledge has evolved into an educational buzzword" (Bernhardt 1991, p. 95). While recognizing that the reader's knowledge of the world plays an important role in the comprehension process, Bernhardt (1991) also points out that this knowledge can be divided into at least three categories: (a) local-level knowledge, or idiosyncratic knowledge of individual communities, (b) domain-specific knowledge, and (c) culture-specific knowledge. It would be naive to assume "knowledge" or lack thereof simply on the basis of ethnic heritage. World knowledge often interacts with and reinforces the quality of vocabulary knowledge.

Fourth, cognitive and affective variables influence the comprehension process. Cognitive variables which may bear on the process include, among others, linguistic intelligence (Gardner,
1983), memory and analytic ability, and relevant affective variables include motivation, attitudes, and emotions.

Fifth, the model recognizes a reciprocal causal relationship in the chain: Vocabulary knowledge can be expanded as a result of reading comprehension (Stanovich, 1986). For L2 learners, this link is especially significant since reading may be used as an important avenue for vocabulary development in both breadth and depth.

The model hypothesizes interactive relationships among the above factors. In the process of reading comprehension, these factors inform and influence one another: A change in the strength of one factor will likely cause change in other factors. By the same token, any weakness in one factor may be compensated by relative strength in other factors (Stanovich, 1980, 1984). The primary purpose for proposing the model has been to identify important learner factors that may influence the reading comprehension process. The present research, however, focuses on only a part of this model (see Figure 2-2): the role of depth of vocabulary knowledge in reading comprehension, although the investigation also explores, to a lesser extent, the roles that vocabulary size and world knowledge play in the comprehension process.

2.3.3.3 Summary. This section has (a) critically examined part of Nation's (1993) vocabulary-reading comprehension triangle, (b) proposed a model identifying factors and relationships involved in the process of reading comprehension based on earlier L1 vocabulary-reading comprehension hypotheses, and based on the proposed model, (c) specified the factors for investigation in the present research.

Although still lacking empirical evidence at this stage, the following position is argued
with the theoretical support of the earlier discussion on depth of vocabulary knowledge. The breadth and depth of vocabulary knowledge are two interconnected dimensions, the development of which will be interdependent to some extent. With the exception of adults particularly knowledgeable in a certain domain, due to experience or specialized training, it would be rare for an L2 reader to have very superior and in-depth knowledge of vocabulary in a given language when his or her overall size of vocabulary in that language is very limited. By the same token, a large vocabulary would probably not take a reader very far if his or her knowledge of this vocabulary is shallow and superficial. Therefore, while the issue of vocabulary size matters a great deal in exploring the relationship between vocabulary knowledge and reading comprehension, the important status of depth of vocabulary knowledge should also be accorded equal recognition.

Figure 2-2. The hypothesized role of vocabulary knowledge and world knowledge in reading comprehension investigated in the present research.
2.4 Hypotheses

The following two hypotheses are derived from the above discussion of the relevant literature.

2.4.1 Hypothesis One

Hypothesis 1 is stated as follows:

In ESL, both vocabulary size and depth of vocabulary knowledge are important components in the vocabulary-reading comprehension chain, and correlations of these two components with reading comprehension, and with each other, will all be at a minimum level of $r = .50$ ($p < .05$).

This hypothesis addresses Research Question 1 (see Chapter 1) concerning the correlational relationship among scores on tests of vocabulary size, depth of vocabulary knowledge, and reading comprehension. In the behavioral sciences, a correlation $r$ of .50 is regarded as indicating a "moderate positive relationship" (Hamilton, 1990, p. 481) between any two variables considered, though this relation is contingent on sample size and the distribution of variables.

In the field of L2 vocabulary-reading research, recent empirical studies (Koda, 1989; Laufer, 1992a, 1996) have reported high correlations ($r \geq .50$) between students' scores on vocabulary size tests and reading comprehension tests. Koda's (1989) study, in which a self-made vocabulary test was used, reported a correlation of .74 ($p < .05$) between scores on the vocabulary test and a paragraph comprehension test. In Laufer's (1992a) study, which employed the Vocabulary Levels Test (Nation, 1983), and the Eurocentres Vocabulary Test (Meara & Jones, 1989), the correlation between the scores on Nation's test and reading comprehension
scores was .50 ($p < .05$) and that between the scores on the Eurocentres' test and the reading comprehension scores was .75 ($p < .05$). In a more recent study (Laufer, 1996), a correlation of .71 ($p < .05$) was found between the scores on Nation's (1983) Vocabulary Levels Test and reading comprehension scores.

Meanwhile, it has been noted that quality of vocabulary knowledge is also important in achieving reading comprehension (Anderson & Freebody, 1981; Mezynski, 1983), although at present empirical evidence is lacking to substantiate this argument. Earlier discussion (Section 2.3.3) called for a recognition of interconnectedness and interdependence in considering the relationship between breadth and depth of vocabulary knowledge. In other words, depth of vocabulary knowledge should be considered as having a role as prominent as breadth of vocabulary knowledge, or vocabulary size, in reading comprehension. It was based on the above arguments that Hypothesis 1 was established.

### 2.4.2 Hypothesis Two

Hypothesis 2 is stated as follows:

For ESL learners whose vocabulary size is beyond the 3,000-word threshold for reading comprehension, scores on depth of vocabulary knowledge will make a unique, and distinctive, contribution to the prediction of reading comprehension scores, over and above the prediction afforded by vocabulary size scores.

This hypothesis arises from Research Question 2 (see Chapter 1), which aims to find out what predictive value depth of vocabulary knowledge offers over and above the prediction afforded by breadth of vocabulary knowledge, or vocabulary size. The threshold hypothesis in
reading comprehension (Clarke, 1979, 1980; Cummins, 1979; Laufer, 1989, 1992a, 1996; Nation, 1990) postulates that, in terms of vocabulary size, there is a threshold level below which the reader will be handicapped by a lack of comprehension and above which the reader will be able to apply his or her reading strategies to help comprehension and achieve better results. Findings from Laufer's (1989, 1992a, 1996) studies, in which general academic texts in English were used to measure comprehension, support this hypothesis. Laufer pointed out that a threshold of 95% lexical coverage of a text is needed for comprehension and is critical in distinguishing a relatively proficient reader (with a minimum of 56% text comprehension in Laufer's studies) from a non-reader. This 95% lexical coverage translates into around 3,000 word families1, or 5,000 individual words (Laufer, 1996). In light of this 95% coverage finding, Hypothesis 2 focuses on ESL learners whose vocabulary sizes are 3,000 and over.

According to Laufer (1989, 1992a, 1996), Meara and Buxton (1987), Meara and Jones (1987, p. 6; 1988), and Nation (1993), and some earlier researchers studying L1 vocabulary knowledge (e.g., Thorndike, 1973), vocabulary size can serve as a good indicator of L2 learners' reading ability. Nevertheless, although empirical results are still lacking, the importance of a second dimension of vocabulary knowledge, i.e., depth of vocabulary knowledge, vis-à-vis reading comprehension has been repeatedly stressed by L1 lexical researchers (e.g., Anderson & Freebody, 1981; Beck, Perfetti, & McKeown, 1982; Mezynski, 1983). This lends support to the argument introduced in Section 2.3.3 that the two aspects of vocabulary knowledge, i.e.,

---

1 Nation (1990) introduced a formula for conversion between word forms and word families. He states that the coefficient for the conversion is 1.6. In her investigation, which concluded that 3,000 word families formed the threshold vocabulary for comprehending academic texts, Laufer (1992) adopted this formula for converting the number into equivalent lexical items, that is, 48,000. To be in line with Laufer's (1992a) investigation and Nation's (1990, 1993) studies when referring to vocabulary size, the figures used in the present research will mean word families, instead of individual word forms.
vocabulary size and depth of vocabulary knowledge, are interconnected and that their
development is, to some extent, interdependent. It was based on the above ground that
Hypothesis 2 was proposed.

2.5 Lexical Inferencing

In recent decades, there has been some research interest in the area of lexical inferencing in L2 reading (e.g., Bensoussan & Laufer, 1984; Haastrup, 1987, 1991; Morrison, 1996). Regarding the third major research question of the present thesis research, which looks at strategies of informed lexical guessing in reading comprehension that are used by ESL learners, a review of relevant studies will be presented in the following sub-sections. Definitions of the term inferencing will first be examined, and then knowledge sources that L2 learners rely on in reading comprehension, especially in making lexical inferences, will be considered.

2.5.1 Defining lexical inferencing

Definitions of what is meant by inferencing have been provided by Carton (1971), Bialystok (1978, 1981, 1983), Færch, Haastrup and Phillipson (1984), and Haastrup (1987, 1991). Among them, efforts by Carton and Haastrup have been most notable. Carton (1971) proposed that language learning is an intellectually challenging process of problem solving, to which learners try to apply, wherever possible, their experience and knowledge. Carton suggested that:

[inferencing] is intended to refer to a process of identifying unfamiliar stimuli...
In inferencing, attributes and contexts that are familiar are utilized in recognizing what is not familiar (p.45).

A more recent, and precise, definition was offered by Haastrup (1987):
The process of inferencing involves making informed guesses as to the meaning of an utterance in the light of all the available linguistic cues in combination with the learner's general knowledge of the world, her awareness of the situation and her relevant linguistic knowledge (p. 197).

Hastrup (1991) assumed that inferencing at the text level and at the word level bear a close relationship, and therefore, that lexical inferencing, like text inferencing, can be considered a comprehension process. According to Hastrup's (1987) definition above, lexical inferencing can be identified with making informed guesses at the word level, thus informed guessing (as opposed to wild guessing) as a term is interchangeable with inferencing. Based on Hastrup's definition, the present research uses lexical guessing as equivalent to lexical inferencing.

2.5.2 L2 reading models

Research on strategies for lexical guessing has been influenced by theoretical models of L2 reading. In the 1960s and 1970s, the concept of bottom-up processing was dominant (e.g., Gough, 1972) and L2 reading was largely regarded as a decoding process (Barnett, 1989), in which there is little room for lexical guessing. The reader is supposed to reconstruct the writer's intended meaning by recognizing letter combinations and words as meaningful units (Rivers, 1981; Plaister, 1968; Yorio, 1971).

Later on, Smith (1971) and Goodman (1967, 1968, 1981) proposed top-down processing models, which eclipsed the bottom-up approach. Describing reading comprehension as a psychological guessing game, Goodman (1967, 1968) defined it as "an interaction between reader and written language, through which the reader attempts to reconstruct a message from the writer" (Goodman, 1968, p. 15). Smith (1971) emphasized the importance of meaning and prediction in reading, and posited that reading is not only purposeful and selective, but the reader
is also expected to bring certain prior knowledge to the text, and that it is the interaction of this prior and higher-level knowledge, the expectation of comprehending, and the purpose in reading that lead the reader to anticipate the text content. Both Goodman's and Smith's views of reading have had a tremendous impact on L2 reading theory. In fact, during the 1970s and 1980s, their models served as "the overwhelming conceptual framework within second language reading" (Bernhardt, 1991, p. 22).

Proponents of top-down approaches argue that readers use their knowledge of syntax and semantics to reduce their dependence on the print and phonics of the text. This approach generally specifies four processes in reading, as Barnett (1989) summarizes:

First, readers make predictions about the grammatical structure in a text, using their knowledge of the language and supplying semantic concepts to get meaning from the structure. Then, they sample the print to confirm their predictions. They neither see nor need to see every letter or word. The more highly developed the readers' sense of syntax and meaning, the more selective the readers can be in sampling. After sampling, they confirm their guesses or, alternatively, correct themselves if what they see does not make sense or if the graphic input predicted is not there (pp. 19-20).

An important impact of the top-down approach on ESL reading instruction has been that it has caused a shift of emphasis in vocabulary teaching from isolated word learning in favour of learning words in context, based on the arguments that learning words in isolation is not productive often due to their polysemic nature and that the meaning of many words can be inferred from contextual clues as long as learners have developed necessary reading skills (Morrison, 1996; Twaddell, 1973). However, this position is not supported by results from a number of vocabulary retention studies, which found either that contextualized vocabulary learning is not superior to decontextualized vocabulary learning (Bensoussan & Laufer, 1984; Gershman, 1970; Lado, Baldin, & Lobo, 1967; Mondria & Wit-de Boer, 1991; Morgan &
Bailey, 1943) or that, under certain circumstances, learners may achieve a better retention of vocabulary learned without context than with a context (Prince, 1996; Qian, 1996; Sibert, 1930).

The advent of interactive models of reading (e.g., Rumelhart, 1977; Stanovich, 1980) ushered in a new era in L2 reading research. While interactive models recognize that higher-level processing influences lower-level processing, they also acknowledge the importance of the printed text for comprehension. However, unlike bottom-up models, interactive models are basically reader-driven, or knowledge-driven. The reader is regarded as an active processor and all of the reader's prior knowledge and experience play a major role in the comprehension process. This expected role of the reader, complemented by the recognition of the importance of the printed text, makes the interactive approach superior to the pure bottom-up or top-down approach.

A typical interactive model of reading (Rumelhart, 1977) assumes that, in the comprehension process, visual information is first scanned from the printed words, and important features of this information are then sent to a message centre, or pattern synthesizer, where various sources of the reader's prior knowledge, linguistic and otherwise, come together to interpret the information received, and "the reading process is the product of the simultaneous joint application of all the knowledge sources" (Rumelhart, 1977, p. 588). Linguistic knowledge sources used by this pattern synthesizer include orthography, lexis, syntax, and semantics. Unlike pure bottom-up or top-down models, which typically emphasize linear and unidirectional processing, interactive models generally feature simultaneous, parallel and multi-level processing, and "a convergence of top-down and bottom-up hypotheses strengthens both" (Rumelhart, 1977, p. 598). Stanovich (1980, 1984) improved the Rumelhart (1977) model by adding to the model
a compensatory aspect, which assumes that "a deficit in any knowledge source results in a heavier reliance on other knowledge sources, regardless of their level in the processing hierarchy" (p. 63), although such over-reliance on one particular knowledge source does not necessarily improve the outcome. Other interactive models (e.g., Lesgold & Perfetti, 1978, 1981; Perfetti & Lesgold, 1977; van Dijk & Kintsch, 1983) vary from Rumelhart's (1977) original model in some aspects, but the interaction of knowledge sources and processing levels still remains the common ground for these models.

In interactive models, identification of a word meaning often involves processing at multiple levels, which typically include lower-level processes of visual information on letters, letter clusters, and the word shape, and some higher-level tasks involving such knowledge sources as syntax, meaning, or knowledge of the world. In lexical inferencing, because of the parallel and multi-level nature of the interactive approach, the visual lexical form is seen to be just as important as the context involved because, even if a higher-level analysis can influence or partially determine a comprehension output, confirmation from the printed symbols is still needed to support the higher-level hypothesis in order to produce the best guesses. Thus, the interactive approach has, to some extent, resurrected the important status of printed words in the reading comprehension process. This resurrection is perhaps partially responsible for the renewed claim of the important status of vocabulary instruction in L2 classrooms.

2.5.3 Proficiency levels and strategies for lexical inferencing

L2 researchers posit that lexical guessing is no easy task for L2 learners, especially when their proficiency level is not up to par (dos Santos & Sanpedro Ramos, 1993; Haastrup, 1990, 1991;
Haynes, 1993; Morrison, 1996). Haastrup (1990, 1991) suggested that there is a threshold level for meaningful guessing, that proficiency level in the language in question is often critical for making successful lexical inferences, and that high proficiency L2 learners normally make more successful guesses than low proficiency L2 learners do. According to dos Santos and Sanpedro Ramos (1993), who looked into the difference between good and poor readers in a study of 140 ESL students who were attending a Brazilian university, poor readers tend to over-use top-down processing strategies, and results of their guessing are often less than satisfactory. This finding is in line with the results obtained by Morrison (1996), who, based on a study of 20 first-year university students of French as a second language (FSL) in Canada, reported that low proficiency (LP) learners tended to over-use world knowledge as an information source, that high proficiency (HP) learners were more capable of using a variety of linguistic knowledge sources than LP learners, and that, at a success rate of 74%, the HP learners were more successful in guessing the meaning of unknown words than were the LP learners, who only achieved a success rate of 34%.

In her study of 63 adult ESL learners from four different linguistic backgrounds, Haynes (1993) found that lack of vocabulary knowledge was a major factor that hindered learners from performing lexical guessing tasks satisfactorily, and that attending to word shape was a popular strategy among ESL learners, but mismatches often happened, especially with learners at lower proficiency levels. Cziko (1980), based on the results of his study on oral reading errors with two proficiency-level groups of FSL students and one group of native French speakers, reported that low proficiency learners tended to focus more on surface graphemic shape and less on contextual information indicating syntactic, semantic or discourse constraints.
However, findings from L1 research (Stanovich, 1990; Stanovich, Cunningham, & Feeman, 1984) argue otherwise: Unless the difficulty level of the material in use is beyond their reach, less skilled readers use as much context in comprehension as skilled readers; and skilled readers rely as much on graphemic cues as less skilled readers do, but in doing so, good readers expend less capacity than less skilled readers, thus still having the capacity to attend to other types of cues simultaneously.

2.5.4 Knowledge sources used in lexical inferencing

Using think-aloud protocols among pairs of students, Haastrup (1987, 1990, 1991) studied the lexical inferencing procedures of high school students who were Danish-speaking learners of English. In developing a taxonomy of knowledge sources for her data analysis based on Carton's (1971) framework of knowledge sources for inferencing, Haastrup (1991) identified three main sources of knowledge L2 learners may use in lexical inferencing: contextual, intralingual and interlingual.

In her taxonomy, Haastrup divided contextual knowledge into two subcategories: knowledge of co-text and knowledge of the world. Co-text includes four subcategories: (1) one or two words from the immediate co-text of the test word, (2) three or more words from the surrounding sentence, (3) any specific part of the co-text beyond the sentence containing the test word, and (4) more global use of the context. Knowledge of the world includes factual knowledge, attitudes and beliefs. The second source of knowledge, intralingual knowledge, is divided into two categories: the test word itself, and the syntax of the sentence containing the test word. Under the category of test word, six subcategories were established: (1)
phonology/orthography, (2) morphology, (3) lexis, (4) word class, (5) collocations, and (6) semantics. Under the category of the syntax of the sentence containing the test word, four subcategories were identified: (1) definite articles, (2) adjectives, (3) prepositions, and (4) number. The third source of knowledge, interlingual knowledge, contains two categories, which are L1 and Ln. Ln refers to all other languages except the informants’ L1 and the target language of the experiment (see Haastrup, 1991, pp. 92-100 for more details).

Among the many findings from her study, Haastrup (1991) found that, based on the total number of 803 valid attempts of lexical inferencing made by her informants, contextual knowledge was the most heavily used source of knowledge by her informants, use of intralingual knowledge ranked second, and interlingual knowledge was relatively infrequently used. At the subcategory level, the most heavily used information sources were co-text, test word came as the second heavily used source, and knowledge of the world was the third heavily used source. The least used knowledge sources were Ln and syntax of the sentence containing the test word.

Morrison’s (1996) study partially replicated Haastrup’s (1991) study, using pair think-aloud processes to infer lexical meanings. At the level of main knowledge sources and based on the total number of 40 valid attempts at lexical inferencing that were made, Morrison found that her results generally corroborated Haastrup’s (1991) with respect to frequencies of use of knowledge sources. At the subcategory level, while co-text was still the most heavily activated information source in the lexical inferencing process -- a finding conforming to that of Haastrup (1991) -- Morrison noted that knowledge of the world was the second most frequently used source, and knowledge of the test word was a distant-third source. These results partially match Haastrup’s (1991) findings.
Haastrop's taxonomy of knowledge sources is comprehensive but its three-level structure is complicated in organization. Although its concept of "co-text knowledge" is useful in indicating locations of clues that informants used in inferencing, its many subcategories of "intralingual knowledge" make the framework too rigid for use in other situations where, for example, different types of text may be used for experiments. Also, it is conceptually debatable whether co-text is a distinct knowledge source.

De Bot, Paribakht, and Wesche (1997) identified a set of eight knowledge sources used in inferring meanings of unknown words, based on evidence from their introspective verbal protocols of 10 ESL learners. They interpreted this finding to support an L2 lexical processing model adapted from Levelt's (1989, 1993) model for L1 speech production. The eight knowledge sources are:

- sentence level grammar;
- word morphology;
- punctuation;
- world knowledge;
- discourse and text;
- homonymy;
- word associations; and
- cognates.

Although organized in a different way, these knowledge sources generally correspond to categories of Haastrop's (1991) taxonomy. The most noticeable difference between the two frameworks is that co-text knowledge, which is an important category in Haastrop's taxonomy,
is not incorporated as a distinct category in de Bot, Paribakht, and Wesche's (1997) criteria, although a number of categories, such as sentence-level grammar, and discourse and text, do imply the use of contextual cues. The criteria proposed by de Bot, Paribakht, and Wesche are obviously an improvement over Haastrup's taxonomy in that, (a) organizationally, their criteria are much simpler than Haastrup's, and therefore easier to use, (b) conceptually, the misconception that contextual knowledge is a knowledge source has been corrected, and (c) they avoid the over-refined, and therefore rigid, subcategories of intralingual knowledge by adopting more inclusive categories such as "word morphology" and "word associations", thus making the taxonomy more flexible to use.

2.5.5 Summary

Section 2.5 examined the definition of inferencing made by other researchers and defined lexical inferencing for the present research. The section also reviewed three distinctive L2 reading approaches, that is, the bottom-up approach, the top-down approach, and the interactive approach, and their relationships to lexical inferencing. Finally, relevant findings from a number of studies on lexical inferencing were presented, and different patterns of using lexical inferencing strategies by learners at different proficiency levels were noted. In addition, some knowledge sources employed in lexical inferencing by L2 learners were summarized based on two studies (de Bot, Paribakht, & Wesche, 1997; Haastrup, 1991). These studies, especially their taxonomies for data analysis, bear on the design of the qualitative data analysis for the present research, which was conducted to answer the third major research question posed in Chapter 1:
How do ESL learners make use of their depth of vocabulary knowledge in the process of comprehending a general academic text?
CHAPTER THREE

METHOD

3.1 Overview

This chapter describes the overall design of the present thesis research, which is composed of two related studies, a main study, and a follow-up study. Details are provided on the sample, research instruments, materials, procedures of data collection, and procedures of data analysis. In addition, the procedures and results of an initial instrument piloting as well as a pilot study, both conducted to prepare for the main study, are described.

The main study was designed to test the two hypotheses stated in Chapter 2, which in turn addressed the first two major research questions introduced in Chapter 1. The follow-up study, with interviews as its mainstay, was designed to address the third research question, which was also introduced in Chapter 1.

3.2 The Main Study

3.2.1 Sample

This section describes the general criteria for choosing the sample for the main study and procedures for selecting participants. Following this, background profiles of two subsamples, namely, a Korean L1 group and a Chinese L1 group, are presented.

3.2.1.1 Criteria. Several criteria were used to identify participants for this research. First, in line with Hypothesis 2, this study was aimed at an ESL population whose L2 vocabulary size was at the 3,000-word level or better. In other words, the participants should all have reached
Second, to minimize the influence of L1 on their ESL performance, especially the factor of guessing in testing, the L1 background of the participants needed to be a language that is not cognate with English, thereby avoiding effects of linguistic relationships of students' mother tongue with English. For this reason, and because I am a speaker of Chinese myself, Chinese ESL learners were first sought as a target group. Chinese, with its ideographs, is not cognate with English relative, for example, to French or German, either orthographically or phonetically. Therefore, in comprehending an English word or text, these aspects of knowledge of their mother tongue will not help Chinese readers guess the meaning of unknown English words.

The third criterion focused on the education level of the participants. Ideally, the participants were expected to have similar L1 education levels in order to minimize the influence of world knowledge and knowledge of academic subjects on their performance in the testing. However, it proved impossible to find a sufficiently large number of participants at precisely the same education level who also met the other criteria, namely, the lexical threshold level for L2 reading and particularly the L1 background. It was therefore decided that the education levels of the participants should range from the completion of high school at the lower end, to the completion of university undergraduate studies at the higher end. At these education levels, learners should be able to handle general academic texts, which composed an important part of the research instruments, selected and developed in line with the three research questions and two hypotheses introduced in Chapters 1 and 2.

In Ontario, a number of universities run intensive ESL programs in their continuing education departments, two of which indicated their support for this research project by
providing access to their ESL students. Of these two departments, one offers a regular 6-8 week multi-level, academic-focused intensive ESL program all year round. The program is designed for those intending to study in English-speaking institutions. According to its brochure, the goal of the program is to develop learners’ academic language skills. The program emphasizes academic reading, essay writing and grammar, as well as listening, note-taking, and talking. The minimum age requirement for entering this program is 18. Learners are recruited from different parts of the world. These learners are placed in classes at 16 different ESL proficiency levels ranging from low intermediate to advanced, based on the results of a placement test administered at the beginning of each program. There is usually a large number of Chinese speakers in the program. The ESL program offered by the other Ontario university was specially set up to prepare some 20 learners from China. These learners, at ESL proficiency levels ranging from high beginners to high intermediate, planned to start university studies after a month’s intensive training in English academic listening, speaking, reading, and writing.

It was assumed that the majority of Chinese-speaking students in these two ESL programs would meet the criteria for participant selection. Therefore it was decided that participants in this research should be recruited from these two programs. During the data collection process it was found that, although there were a large number of Chinese speakers in the ESL programs where participants were being sought, the number who signed the consent forms was not large enough to guarantee the scope of data analyses I had envisioned. On the other hand, there was a large number of Korean learners in the multi-level program described above, and there was a strong interest from this group of learners to participate in the study. Therefore, a Korean group was also included in the study. As it turned out, the number of
participants of Korean L1 background exceeded that of Chinese learners.

Although Korean and Chinese are linguistically unrelated, the original linguistic criterion for the selection of Chinese learners was also applicable to the Korean learners. Linguistically, Korean and English belong to typologically different language systems. Orthographically and phonetically, English and Korean are very different and unrelated to each other. Therefore, it could be assumed that their L1 would play a minimal role in the Korean learners’ guessing the meaning of unfamiliar words during testing.

Data for the research were eventually generated from a sample of 74 ESL learners1 with either a Chinese or a Korean L1 background. All the participants were taking intensive ESL programs in the two continuing education departments mentioned earlier. Of these learners, many were preparing themselves for further studies in a Canadian or American university where English would be the language of instruction, while others were attending the programs to improve their English for other purposes.

3.2.1.2 Procedures for recruitment. These learners participated in the present study on a voluntary basis. About one to two weeks before the testing, letters describing the purpose of the research and the procedures for the data collection were distributed to all Chinese and Korean learners in the two programs, mostly through their ESL teachers. I also met with some of the learners to explain what was expected of participants. Those who were willing to participate all signed a consent form attached to the letter (see Appendix A). After a participant list was

---

1 Altogether 80 learners were tested, but the scores of 6 participants were discarded because they failed to meet the 3,000-word threshold level. Therefore, the numbers discussed elsewhere in this thesis do not include these 6 learners.
established based on the information collected from the consent forms, these volunteers were contacted individually, mostly via telephone, to set up the time for testing.

3.2.1.3 Profile of the Korean group. The number of participants in the Korean group totaled 41, 5 male and 36 female. They were all attending the multi-level intensive academic ESL program described earlier: (a) 15 of them were preparing for entry into an undergraduate program in a Canadian or U.S. university, (b) 6 intended to pursue graduate education in Canada at the conclusion of the ESL training, (c) 10 were attending the program for the purpose of improving their English, and they intended to go back to their home country to continue their university education there at the conclusion of the ESL program, and (d) 10 intended to look for a job after taking the program.

All the Korean participants had started to learn English at the beginning of junior high school, which was at about 13 years of age. The average age of this group was 24 years, with a range from 18 to 42 years. As revealed in Table 3-1, the majority (90%) were between the ages of 18 and 27. With respect to educational level, all members of this group had completed senior high school, and 16 had completed university undergraduate education. None of them had started any graduate education. These participants were in a variety of specialized fields, which included over 30 disciplines in the humanities, social sciences, science, and engineering.

The length of their stay in Canada varied slightly. Thirty-five learners had been in Canada for no more than six months. All of the 41 learners had been in Canada for less than a year. A summary of the background information about the Korean group appears in Tables 3-1, 3-2, and 3-3.
### Table 3-1. Age ranges of the Korean group (n=41) and the Chinese group (n=33)

<table>
<thead>
<tr>
<th>Age Range</th>
<th>Number of Learners</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Korean</td>
<td>Chinese</td>
</tr>
<tr>
<td>18-22</td>
<td>20</td>
<td>8</td>
</tr>
<tr>
<td>23-27</td>
<td>17</td>
<td>12</td>
</tr>
<tr>
<td>28-32</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>33-37</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>38-42</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

### Table 3-2. Education levels of the Korean and Chinese ESL learners

<table>
<thead>
<tr>
<th>Educational Level</th>
<th>Number of Learners</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Korean</td>
<td>Chinese</td>
</tr>
<tr>
<td>High school graduates</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>University first year</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>University second year</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>University third year</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>University fourth year</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>University graduates</td>
<td>16</td>
<td>6</td>
</tr>
<tr>
<td>College graduates</td>
<td>0</td>
<td>18</td>
</tr>
</tbody>
</table>

### Table 3-3. The Korean and Chinese ESL learners’ duration of stay in Canada

<table>
<thead>
<tr>
<th>Duration of Stay in Canada</th>
<th>Number of Learners</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Korean</td>
<td>Chinese</td>
</tr>
<tr>
<td>1-3 months</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td>4-6 months</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>7-12 months</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>13-18 months</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>
3.2.1.4 Profile of the Chinese group. The number of participants in the Chinese group totaled 33, 5 male and 28 female. They were all attending the intensive ESL programs in the two continuing education departments described earlier: (a) 22 were preparing for entry into an undergraduate program in a Canadian or U.S. university, (b) 5 intended to pursue graduate education in Canada at the conclusion of the ESL training, (c) 4 planned to attend a community college or a vocational school in Canada, and (d) 2 intended to look for a job in Toronto at the conclusion of the program, believing that completing the program would give them an edge in securing employment.

These participants had started to learn English at various ages. Three started to learn English before the age of 10. Twenty received their first English lessons at the beginning of junior high school. The remaining 10 started learning English when they began their university education at about 18-19 years old. The average age of this group was 26.9 years, with 40 being the oldest and 19 the youngest. The education level of this group was between the completion of senior high school and the completion of university undergraduate education. Like the Korean group, none of the Chinese participants had been exposed to graduate education. Their academic fields included the humanities, social sciences, science, and engineering. Altogether, 18 disciplines were represented.

The length of their stay in Canada also varied. The majority (28 learners) had been in Canada for no more than 6 months. Thirty-one had been in Canada for less than a year. Tables 3-1, 3-2, and 3-3 also provide a summary of the background profile of the Chinese group.
3.2.2 Instruments

3.2.2.1 Background Questionnaire. This questionnaire was designed to obtain background information on participants' age, education level, academic field, L1, length of learning English, length of stay in Canada, and plans after the completion of the ESL program (see Appendix B).

3.2.2.2 Vocabulary size test. Originally called the Vocabulary Levels Test, this vocabulary size test (VS) was developed by Nation (1983, 1990) to measure the size of the test-taker's vocabulary. The test has been accepted by a number of L2 researchers as an appropriate measure of vocabulary size (cf. Laufer, 1992a, 1996; Yu, 1996). Yu (1996), in his study of Chinese and Japanese university students' use of English motion verbs, used the VS to determine the initial vocabulary size of his participants. Before putting the test to formal use in his main study, Yu piloted two levels of the test (the 2,000- and 3,000-word levels) with a sample of 47 adult ESL students, together with 32 items selected from two established English proficiency tests (Test of English as a Foreign Language and Michigan Placement Test). Yu reported a very high correlation (.99) between the VS and the TOEFL/Michigan test items, thus confirming the concurrent validity of the VS for his study.

The VS (see Nation, 1983, pp. 19-24, for the complete test) is composed of five parts representative of five different vocabulary size levels in English, namely, the 2,000 word level, the 3,000 word level, the 5,000 word level, the university word level, and the 10,000 word level. At each vocabulary size level, there are six test items, each comprising six words and three definitions. The test-taker is required to match the three definitions with three of the six words provided by writing the corresponding number of the words beside the definitions. An example
taken from the instruction part of the VS is given below:

1 business
2 clock 6 part of a house
3 horse 3 animal with four legs
4 pencil 4 something used for writing
5 shoe
6 wall

Each level contains 18 correct choices. The words at each level of the test were selected "so that they would be representative of all the words at that level" (Nation, 1983, p. 14). Because of the way the test was constructed, the chance of guessing correctly is low, and testees' scores can be regarded as "a close approximation to the proportion of words in the test that they know" (Nation, 1990, p.262).

The 2,000 and 3,000 word levels of the VS include high frequency words in English; the 5,000 word level is a boundary level between the high frequency level and low frequency level; and the 10,000 word level is composed of low frequency words. The university word level contains specialized vocabulary needed for academic studies. The vocabulary items selected for the university word level represent words frequently appearing in university textbooks.

In scoring, each word correctly chosen is awarded 1 point. The maximum possible score is 90 for the same number of words. In the present research, all the five levels of the VS were used. The maximum possible score for the test in this research was therefore 90.

In interpreting test scores, Nation (1983) states that a score of 12 or fewer out of 18 points at a vocabulary size level is an indication that this level has not been mastered. The present study required the starting vocabulary size level for the data to be at 3,000 because the second hypothesis of this study was aimed at a sample population with a minimum vocabulary size of 3,000, which is considered the threshold level for ESL reading comprehension (Laufer,
Therefore, to be eligible for inclusion in this study, a learner had to score 13 points or more at the 3,000-word level of the VS. Based on the experience of the pilot study (see Section 3.4), the maximum time allowed for completing the VS in the main study was 20 minutes.

3.2.2.3 Reading comprehension test. This test was a standardized multiple-choice reading comprehension test (hereafter RC), which was taken from a version of the Test of English as a Foreign Language (TOEFL) (Educational Testing Service, 1987, pp. 93-100). It was a section of an official TOEFL test. The original test was composed of six reading passages and 30 comprehension questions. Because of the anticipated time constraints of administering the tests, I perceived a need to shorten the test. As a result, two passages were randomly removed along with 10 comprehension questions. This left 20 multiple-choice comprehension questions to measure reading comprehension levels. In scoring, each correct answer to a comprehension question was awarded 1 point. As there was a total of 20 questions, the maximum possible score was 20.

As an established standardized language test, all of the official TOEFL tests have been carefully pretested for validity and reliability before being put into actual use. In this study, therefore, the RC was not a main focus of concern in instrument piloting. The maximum time allowed to complete the RC in the main study was 25 minutes, based on the experience of administering the test in the pilot study (see Section 3.4).

3.2.2.4 Depth-of-vocabulary-knowledge test. Originally called the Word Associates Test, this
depth-of-vocabulary-knowledge test (hereafter DVK) was developed by Read (1989, 1993, 1994, 1995) to assess depth of vocabulary knowledge in English. The most recently revised version of the DVK (Read, 1994, 1995) is composed of 40 items designed to measure two aspects of the depth of vocabulary knowledge: meaning and collocation, or the paradigmatic and syntagmatic relationships of words. The aspects that the DVK measures match some important components of the operational definition of depth of word knowledge that has been proposed in the present research (see Chapter 2).

Each of the 40 DVK items consists of one stimulus word, which is an adjective, and two boxes, each containing four words. Among the four words in the left box, one to three words can be synonymous to one aspect of, or the whole, meaning of the stimulus word, while among the four words in the right box there can be one to three words that collocate with the stimulus word. Each item always has four correct choices. However, these choices are not evenly spread. There are three possible situations: (1) the left and right boxes both contain two correct answers; (2) the left box contains one correct choice, while the right box contains three correct answers; (3) the left box contains three correct answers, while the right box contains only one correct choice. This arrangement effectively reduces the chances of guessing. An example taken from the instruction part of the test is provided below.

\[ \text{sudden} \]

\[
\begin{array}{|c|c|c|}
\hline
\text{beautiful} & \text{quick} & \text{surprising} & \text{thirsty} \\
\hline
\end{array}
\quad
\begin{array}{|c|c|c|}
\hline
\text{change} & \text{doctor} & \text{noise} & \text{school} \\
\hline
\end{array}
\]

The DVK was originally designed to measure the depth of vocabulary knowledge of adult ESL students in New Zealand; it went through a process of repeated piloting and refining, including a large-scale field testing (N=103, Read, 1994, 1995). In one recent trial (N=94), a
high correlation ($r = .82$) was found between the DVK and a comparison vocabulary matching test. The Rasch reliability coefficient of the DVK in this trial was .93, and that of the matching test was .90 (Read, 1995). In an earlier trial ($N = 38$), a correlation coefficient of .85 ($p < .01$) was reported (Read, 1994) between the DVK and the matching test.

In the present study, the content of the DVK was partially modified for the following reason. Although not reflected in the major research questions, it was of interest to find out whether performance on words selected from passages in the reading comprehension test would be a better indicator of the readers’ comprehension level than performance on words that did not appear in the reading comprehension test (E. Geva, personal communication, January 1995). To address the question, it was planned to add 10 new items to the existing DVK, using adjectives selected from the RC as stimulus words. However, time constraints on the data collection did not allow lengthy tests to be administered. Consequently, 10 items in the original DVK had to be removed in order to make room for ten new stimulus words to be tested.

As investigator in this research and as an ESL teacher with fifteen years’ experience, I worked with two native English-speaking ESL teachers in modifying the DVK. First, criteria were established for removing some stimulus words, and therefore items, from the original DVK. An item would be discarded if two out of the three teachers judged that the difficulty level of the whole section was too low, or that there was an ambiguous choice in the item. As a result, nine items were discarded either because they were too easy or contained some questionable choices. These nine items were then replaced by nine new items with their stimulus words selected from reading passages in the RC. Coincidentally, one of the remaining 31 original items, with the stimulus word "tight", also qualified as a replacement item because the word
"tight" happened to appear in a passage in the RC. Therefore, this item was retained but counted as an RC-text-related item.

Two criteria were followed in selecting the nine new target words from the RC: (1) they must be adjectives; and (2) these adjectives must all have multiple meanings and/or a relatively wide range of uses. The final appearance of the version of the DVK developed was as follows: Items 1-30 and 32 were original DVK items that were re-numbered (Item 32 is headed by "tight"), and Items 31, and 33-40 were new. Thus, Items 1-30 were considered to be the DVK's non-RC-related part, or the general part, while Items 31-40 (see Appendix C) were the RC-related part. The 10 adjectives selected as the DVK's stimulus words accounted for 3.6% of the total number of words appearing in the texts of the RC\(^2\). The maximum time allowed to complete the revised DVK was 30 minutes.

In scoring, each word correctly chosen was worth 1 point. A wrong choice would get 0. Therefore the maximum possible score was 40 x 4 = 160 points for the 40 test items. That is, an individual's total score would equal the total number of correct choices he or she made.

3.2.2.5 Morphological Knowledge Test. This was a test designed for the purpose of the present research (see Appendix D). The Morphological Knowledge Test (hereafter MK) was developed to measure morphological aspects of vocabulary knowledge. The MK has 10 stimulus words shared by the two parts of the test. One part, which contains 10 items, tests the knowledge of lexical affixes and stems. All the 10 stimulus words of the affix subtest contain a prefix or a

\(^2\) This number only includes the first occurrence of words in the texts. Repeated occurrences of the same word, or words assuming different forms but from the same word family, for example, *produce* and *produced*, were counted as one word.
suffix. The test requires the test-taker to define, in English or the person’s L1, the 10 stimulus words, including the affixes. The other part measures knowledge of parts of speech. In this part-of-speech subtest, the stimulus words appear in two forms, the stem and the stem plus an affix. This part requires the test-taker to identify the correct parts of speech of the two forms of a given word. The 10 stimulus words of the MK were all selected from the reading passages in the RC. The instructions and an example of the test items follow:

**Instructions:** Explain the meaning of the following 10 words in English, or translate them into your first language. If the underlined part of the word has changed the part of speech of the base word, please indicate this by using appropriate terms provided below.

Possible terms: **noun, verb, adjective, adverb, preposition**

Example:

<table>
<thead>
<tr>
<th>Word</th>
<th>Meaning or translation of the word</th>
<th>The change in the part of speech caused by the underlined part</th>
</tr>
</thead>
<tbody>
<tr>
<td>homeless</td>
<td>without home</td>
<td>n. → adj.</td>
</tr>
</tbody>
</table>

In scoring, each item was worth 1 point. A wrong answer would get 0. In the case of the affix subtest, a half point was awarded to answers which bore a close meaning to the stimulus word but failed to accurately define the word. The maximum possible score on the affix subtest was 10 for 10 correct definitions, and in the part-of-speech subtest, 20 for 20 correct identifications of part of speech.

Since the affix subtest required some productive work to define the 10 stimulus words, it was marked by two raters in order to establish interrater reliability. When differences arose due to different rater opinions, a third rater was called upon, in which case the score of the controversial item was decided according to the majority opinion. I took responsibility as the
first rater. The other two raters were also experienced ESL teachers in established, and well-reputed, adult ESL programs in Toronto. Out of a total of 760 items, the first two raters had different opinions on 37 items, or about 5% difference. In other words, the interrater agreement was about 95%. The 37 discrepancies were resolved with the input of the third rater, a native English speaker. Based on the experience of the pilot study (see Section 3.4), the maximum time allowed for completing this test was 15 minutes.

3.2.3 Procedures for data collection

Data collection for the main study was in the form of paper-and-pencil testing. All the four tests, the VS, DVK, MK, and RC, were administered in single sessions. A letter of information, together with a consent form, was given out to all the potential test-takers well in advance of the testing time. Only those who agreed to participate by signing the consent form were invited to the testing. Because the test-takers were all volunteers who had busy and different schedules, arrangements for the testing had to be made individually through telephone calls. As a result, 35 testing sessions were held with groups of different sizes, the largest being 13, and the smallest 1. The four tests were administered in five different orders to these groups in order to minimize the possible influence of learning effects since 10 of the stimulus words in the DVK and all the stimulus words in the MK also appeared in the RC. These orders were:

1. Questionnaire-VS-DVK-RC-MK;
2. Questionnaire-VS-RC-DVK-MK;
3. Questionnaire-VS-MK-RC-DVK;
4. Questionnaire-VS-MK-DVK-RC; and
5. Questionnaire-VS-RC-MK-DVK.

Attention was paid to strictly controlling the time in each session. I took sole responsibility as the administrator of the testing, in order to ensure that all the testing sessions were conducted within the same time frame, and that the participants would receive the same type of instructions and explanations. The scores were reported in writing to the test-takers by mail, together with individual recommendations for improvement from their existing vocabulary knowledge levels.

3.2.4 Data analyses

The data analysis was designed as a five-phase process, with different purposes and a different combination of variables in each phase. The data collected through the four tests, namely, the RC, VS, DVK, and MK, were all of an interval nature. Because the primary purpose of this study was to find out: (a) the intercorrelations among the learners' scores on the reading comprehension, vocabulary size, and depth of vocabulary knowledge, and (b) to what extent scores on depth-of-vocabulary-knowledge tests contribute to predicting reading comprehension scores in addition to the prediction already accomplished by scores on the vocabulary size test, Pearson product-moment correlations and multiple regression were chosen as the dominant techniques for the statistical analyses. Means, score ranges, and standard deviations were computed when necessary. In addition, oneway analysis of variance (ANOVA) was also performed in comparing the means of the four tests between the Korean group and Chinese group, in order to determine whether the data from the two groups were comparable.

The SPSS for UNIX, release 5.0, was used for the statistical analyses. SPSS was chosen
as the program for the statistical analyses in this study mainly for two reasons. First, SPSS was specifically designed for statistical analyses in the social sciences. Second, SPSS has a special function to detect serious multicollinearity\(^3\) that may influence the variance of the estimated regression coefficients (Berry & Feldman, 1985). Its use would therefore protect the integrity of the regression model in the process of multiple regression analysis, which was the central technique of data analysis in the main study. The variables that were used in the five phases of the data analysis are listed in Table 3-4. All the phases are described below.

3.2.4.1 Phase 1. As noted previously (see Section 3.2.1), the original design of the study involved only one L1 population, namely, Chinese ESL learners. However, due to the difficulty in obtaining a sufficient number of participants from the target population, and also because of the availability and willingness of Korean learners to participate, 41 Korean learners were included in the study. The background of these two groups was generally comparable apart from their L1. In order to determine whether the data generated from the two L1 groups were also comparable and whether the two groups could therefore be treated as one sample in the later phases of analysis, the means, ranges, and standard deviations of the scores of the two groups were computed, and oneway ANOVAs were conducted on four variables, RC, VS, DVK, and MK, to determine whether differences existed between the group means. If statistical differences between the group means were not found, the data analyses could then proceed to Phase 2.

---

\(^3\) According to Norušis (1990), in the SPSS process of multiple regression, all independent variables to be included in a regression equation are automatically tested for tolerance. They first have to pass both tolerance and minimum tolerance tests in order to gain entry into the equation. If a variable fails to meet the tolerance criterion, a warning will be issued and the variable will be rejected.
Table 3-4. Variables for the statistical analyses in the main study

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>MPS$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>RC</td>
<td>Score on the Reading Comprehension Test (RC)</td>
<td>20</td>
</tr>
<tr>
<td>VS</td>
<td>Score on the vocabulary size test (VS), originally known as the Vocabulary Levels Test</td>
<td>90</td>
</tr>
<tr>
<td>DVK</td>
<td>Score on the depth-of-vocabulary-knowledge test (DVK), originally known as the Word Associates Test</td>
<td>160</td>
</tr>
<tr>
<td>DVKmg</td>
<td>Score on the meaning part of the DVK</td>
<td>76</td>
</tr>
<tr>
<td>DVKcol</td>
<td>Score on the collocation part of the DVK</td>
<td>84</td>
</tr>
<tr>
<td>DVKold</td>
<td>Score on the general vocabulary knowledge part of the DVK (Items 1-30)</td>
<td>120</td>
</tr>
<tr>
<td>DVKnew</td>
<td>Score on the RC-text-related part of the DVK (Items 31-40)</td>
<td>40</td>
</tr>
<tr>
<td>MK</td>
<td>Score on the Morphological Knowledge Test (MK)</td>
<td>30</td>
</tr>
<tr>
<td>MKaffix</td>
<td>Score on the affixation part of the MK</td>
<td>10</td>
</tr>
<tr>
<td>MKpart</td>
<td>Score on the part-of-speech part of the MK</td>
<td>20</td>
</tr>
<tr>
<td>L1</td>
<td>Learner’s first language background</td>
<td>0 or 1</td>
</tr>
</tbody>
</table>

a. MPS = maximum possible score

3.2.4.2 Phase 2. Phases 2 to 4 were designed on the assumption that results from Phase 1 would indicate no statistical difference between the group means on the RC, DVK, VS, and MK. The two L1 groups would then be treated as one sample in the analyses to be conducted in Phases 2 through 4. The primary reason to combine the two groups was a purely technical one in that reliable statistical results, especially those of multiple regression, could only be obtained when a sufficiently large number of data points was available.

In this phase, the one-tailed product-moment correlation was computed for scores from the RC, VS, DVK, and MK to determine the intercorrelations among the four tests which
represent, respectively, reading comprehension (RC), vocabulary size (VS), and depth of vocabulary knowledge (DVK and MK).

In the regression analyses in this phase, RC was used as the dependent variable and VS, DVK, and MK as the independent variables. The vocabulary knowledge-reading comprehension model proposed in Chapter 2 specifies two major dimensions of vocabulary knowledge, i.e., breadth and depth. According to the working definition of vocabulary knowledge introduced in the same chapter, the following are all main components of the depth dimension: morphological properties, lexical meaning, and lexical collocation. In the present research, the morphological component was represented by the MK, and the meaning and collocation components were covered by the DVK. In conformity with the above conceptual framework, the main purpose for this phase was to determine the extent to which scores on the DVK and MK can add to the prediction of RC scores over and above the prediction afforded by VS scores.

More specifically, this phase of the analyses was designed to: (a) check the reliability of RC, VS, DVK, and MK based on the scores of the whole sample, as compared with the reliability coefficients obtained in the pilot study (see Section 3.4); (b) determine the intercorrelations among RC, VS, DVK, and MK; and (c) determine to what extent DVK and MK will improve the prediction of RC on top of the prediction already provided by VS.

3.2.4.3 Phase 3. The third phase also involved full-sample analyses. In this phase, RC was still the dependent variable and VS remained an independent variable. However, the other two independent variables used in Phase 2, DVK and MK, were split into four, namely, DVKmg, DVKcol, MKaffix, and MKpart. In other words, the DVK and MK were now treated as the
following four separate tests:

1. The lexical meaning part, or the left boxes, of the DVK (represented by DVKmng), with a maximum possible score of 76;
2. The word collocation part, or the right boxes, of the DVK (represented by DVKcol), with a maximum possible score of 84;
3. The affixation part of the MK (represented by MKaffix), with a maximum possible score of 10; and
4. The part-of-speech part of the MK (represented by MKpart), with a maximum possible score of 20.

There were two reasons for dividing the two tests into four subtests. First, the definition of depth of vocabulary knowledge proposed in Chapter 2 specifies that the main components of this dimension of vocabulary knowledge include: (a) lexical-morphological properties, (b) lexical meaning, and (c) lexical collocation (for details, see Section 2.3.2.3). The MK covered two major parts of (a), namely, affixes and parts of speech, and the DVK represented (b) and (c).

Second, prompted by Verhallen and Schoonen’s (1993) claim of a significant paradigmatic-syntagmatic distinction in the development of an L2 vocabulary in child learners, Read (1995) recently proposed that the DVK was well-suited for doing research into this topic, since this test covers both paradigmatic and syntagmatic components and, organizationally, the two components are clearly distinguished in the left and right parts of test items. This structure made it feasible to build the paradigmatic-syntagmatic comparison into the present design in tandem with the main purposes of the present study. This third phase of analysis was designed to achieve the following purposes:
1. To determine the intercorrelations among RC, VS, DVKmng, DVKcol, MKaffix, and MKpart.

2. To examine individually the magnitude of contribution of each independent variable to the prediction of RC, so as to determine which dimensions and components of vocabulary knowledge play a more important role in predicting reading comprehension.

3.2.4.4 Phase 4. The fourth phase of the analysis also used the full sample. The variables included in the analyses were RC as the dependent variable and VS, DVKold, and DVKnew as the independent variables. As mentioned in Section 3.2.2.3, there has been some research interest in whether scores on a depth of vocabulary knowledge test with target words selected from a related reading comprehension test will provide better prediction of reading comprehension levels than scores on a similar test whose target words do not appear in the related reading comprehension test. This phase of the analysis was designed to shed some light on this question, and the modified DVK was considered to be a suitable research instrument for this comparison.

Variable DVKold represented scores on DVK Items 1-30, which do not contain head words appearing in the RC passages, and are thus deemed a general DVK test. DVKnew stood for scores on DVK Items 31-40, which are headed by 10 words selected from the passages of the RC, and thus considered to be RC-related.

3.2.4.5 Phase 5. The fifth phase was composed of some split-sample analyses. In this phase, the sample was divided into two groups according to the learners' L1 background, namely, Korean
and Chinese, and split-file analyses were carried out. The purpose of this phase was to find out whether similar patterns were produced by each L1 group when their scores on the RC, VS, DVK, and MK, were analyzed separately. The variables involved in this phase were RC as the dependent variable, and VS, DVK, MK, and L1 as the independent variables.

3.3 Obtaining Native Speaker Norms for the Instruments

Before a pilot study was conducted, a group of native English speakers were invited to pre-test the major instruments to be used in the main study, in order to make sure that all the test items were of good quality and the choices provided, especially in the DVK, conformed to Canadian usage. The instruments tested included the RC, VS, DVK, MK, and the Background Questionnaire. Attention was focused on two tests, namely, the DVK and MK, since the MK was a new test and the DVK was a modified version of the original test, which in itself was a recent invention.

3.3.1 Background of the native English speakers

This exercise involved eight educated Canadian native-speakers of English. The participants, who took the tests on a voluntary basis, were all mature adults over 30 years of age in six different professions, which included law, law enforcement, computer engineering, electronics technology, social work, and education. The education levels of the participants are provided in Table 3-5.

3.3.2 Testing procedures

The testing took place on five different occasions within one week. All the four tests were
administered to the participants in single sessions. Before the testing proceeded, the purpose of the testing was clearly explained to the participants such that they were expected to provide some input on the format and contents of the tests in addition to writing the tests. Immediately after the testing, a brief discussion always took place concerning the quality and appropriateness of the tests.

3.3.3 Results

The native speaker scores on the tests and some descriptive statistics, including means, score ranges, and standard deviations, are provided in Tables 3-5 and 3-6. The primary purpose of the testing was to seek input from the participants to ensure that all the test items were acceptable to native Canadian English speakers. This purpose was fulfilled satisfactorily.

Table 3-5. Native speaker scores on the VS, DVK, MK, and RC (n=8)

<table>
<thead>
<tr>
<th>Par No.</th>
<th>Educ Level</th>
<th>VS MPS: 90</th>
<th>DVK = DVKold + DVKnew MPS: 160 = 120 + 40</th>
<th>MK = MKaffix + MKpart MPS: 30 = 10 + 20</th>
<th>RC MPS: 20</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PhD</td>
<td>154=116+38</td>
<td>29=10+19</td>
<td></td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>BSc</td>
<td>153=113+40</td>
<td>30=10+20</td>
<td></td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>MEd</td>
<td>153=114+39</td>
<td>29= 9+20</td>
<td></td>
<td>19</td>
</tr>
<tr>
<td>4</td>
<td>MEd</td>
<td>151=113+38</td>
<td>30=10+20</td>
<td></td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td>Cold</td>
<td>139=103+36</td>
<td>30=10+20</td>
<td></td>
<td>20</td>
</tr>
<tr>
<td>6</td>
<td>MSc</td>
<td>151=112+39</td>
<td>30=10+20</td>
<td></td>
<td>19</td>
</tr>
<tr>
<td>7</td>
<td>Col</td>
<td>138=103+35</td>
<td>20=10+10</td>
<td></td>
<td>19</td>
</tr>
<tr>
<td>8</td>
<td>HS*</td>
<td>150=112+38</td>
<td>16= 8+ 8</td>
<td></td>
<td>17</td>
</tr>
</tbody>
</table>

a. Par = participant  b. Educ = education  c. MPS = maximum possible score  d. Col = community college  e. HS = high school
Table 3-6. Native speakers’ score means, standard deviations, and ranges (n=8)

<table>
<thead>
<tr>
<th>Test</th>
<th>M</th>
<th>SD</th>
<th>Score Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>RC</td>
<td>19.25</td>
<td>1.04</td>
<td>17 - 20</td>
</tr>
<tr>
<td>VS</td>
<td>89.25</td>
<td>1.17</td>
<td>87 - 90</td>
</tr>
<tr>
<td>DVK</td>
<td>148.62</td>
<td>6.39</td>
<td>138 - 154</td>
</tr>
<tr>
<td>MK</td>
<td>26.75</td>
<td>5.52</td>
<td>16 - 30</td>
</tr>
</tbody>
</table>

Some helpful comments were also made on all of the four tests. For example, the participants generally agreed that the four tests were in good shape, and would be able to achieve what was expected to measure. But a few participants noted that the instructions for the MK were not clear enough. One participant even missed the part-of-speech subtest at first. Also, two participants argued that a few items in the VS were culturally biased: for instance, that the target words *angel* and *sermon* in the VS would not likely be known to those who did not have a Christian background. A few sections in the DVK were considered to be "tricky" because there appeared to be more than four possible choices in each of these sections. One test-taker also complained that the texts in the RC were boring, which may negatively affect the learners' performance on the testing.

It can be observed from Table 3-6 that the score range of the MK was fairly widespread. This was due to the low score of one person, a high school graduate who had very limited knowledge of parts of speech. The score range would have been improved to 20-30 had this person been removed from the list. Based on the input, modifications were made to the MK and DVK. The VS and RC were, however, kept intact since the problem pointed out in the VS was not considered serious enough to affect test-takers’ scores considerably, and the "boredom" of
the RC would likely be applied to the learners across the board without discrimination.

3.4 Pilot Study

Following the testing with the eight native English speakers, a pilot study was conducted with ESL learners in order to test the reliability of the RC, VS, DVK, and MK, which were to be employed as the major instruments for the main study.

3.4.1 Sample

The sample for the pilot study was composed of 12\(^4\) young adult ESL learners, including 1 male and 11 female learners. Among them, 8 had a Korean L1 background, and 4 Chinese. When the pilot study took place, they were all attending an intensive academic ESL program in one of the universities mentioned earlier (see Section 3.2). They came to the ESL program in order to start university in an English-speaking environment, or to fare better in their university studies in their home countries. The range of their length of stay in Canada was from 1 month to 18 months, with an average of 4 months. Their ages ranged from 18 to 27 years, averaging 22.8. Their lengths of learning English ranged from 4 years to 12 years, averaging 8 years. Their education levels ranged from the completion of high school to the completion of undergraduate studies in their home countries. They were at different ESL levels from low intermediate to advanced, and therefore in different proficiency-level classes.

\(^4\) Altogether 16 learners were included for the pilot testing. However, scores of 4 learners (two Chinese and two Korean) on the VS failed to meet the 3,000-word threshold level for reading comprehension. These 4 learners' scores were, therefore, discarded.
3.4.2 Testing procedures

At the beginning of the testing session, the participants were asked to complete the Background Questionnaire. After that, the RC, VS, DVK, and MK were administered to them separately in a number of mixed orders. All four tests were completed in single sessions. Although several occasions were arranged to accommodate these volunteer participants’ different schedules, the actual maximum time spent on each test was strictly controlled and kept identical on all occasions in order to ensure the reliability of the results.

3.4.3 Analyses and results of the pilot study

When the scores on the four tests were obtained, I carried out certain statistical analyses:

1. Computation of means, ranges and standard deviations of the scores on the four tests to obtain information on the distribution of the scores;

2. Pearson product-moment correlations among scores on the four tests with a view to determining the intercorrelations among scores on the RC, VS, DVK, and MK; and

3. Computation of reliability coefficients of the four tests using Kuder-Richardson Formula 21 (K-R 21), in order to ensure that the reliability coefficients of the four tests were all at acceptable levels.

Table 3-7 summarizes the means, ranges and standard deviations of the scores on the RC, VS, DVK, and MK in the pilot study. The reason for conducting the pilot study was to ensure that the instruments to be used in the main study had acceptable levels of reliability and to find

---

5 Multiple regression procedures were not performed in the pilot study because the sample size was too small for the purpose.
out how the intercorrelations among the scores on the RC, VS, DVK, and MK would look. However, it should be cautioned that because of the small size of available learners, the statistical results listed in Tables 3-8 and 3-9 could only provide some preliminary profiles rather than serve as meaningful indicators.

**Table 3-7.** The pilot study (N = 12): Learner means, standard deviations, and score ranges

<table>
<thead>
<tr>
<th>Test</th>
<th>MPS&lt;sup&gt;a&lt;/sup&gt;</th>
<th>M</th>
<th>SD</th>
<th>Score Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>RC</td>
<td>20</td>
<td>16.33</td>
<td>2.74</td>
<td>11 - 20</td>
</tr>
<tr>
<td>VS</td>
<td>90</td>
<td>66.58</td>
<td>15.29</td>
<td>38 - 81</td>
</tr>
<tr>
<td>DVK</td>
<td>160</td>
<td>122.50</td>
<td>18.21</td>
<td>87 - 142</td>
</tr>
<tr>
<td>MK</td>
<td>30</td>
<td>25.08</td>
<td>5.90</td>
<td>12.5 - 30</td>
</tr>
</tbody>
</table>

<sup>a. MPS = maximum possible score.</sup>

As the method of rational equivalence to check internal consistency of the tests (Alderson, Clapham, & Wall, 1995; Borg, 1981; Harrison, 1983; Richards, Platt, & Platt, 1992), K-R 21 was computed for the scores with a view to obtaining reliability coefficients for the RC, VS, DVK, and MK. Pearson product-moment correlations were calculated for intercorrelations among the scores on the RC, VS, DVK, and MK. Table 3-8 lists the reliability coefficients for the scores on the four tests, and Table 3-9 provides the product-moment correlations among the RC, VS, DVK, and MK scores.

The reliability coefficients appearing in Table 3-8 looked adequate for my purpose. The \( r \) values of three of the four tests, namely, the VS, DVK, and MK, were very high. The \( r \) of the RC, which had been expected to yield a higher value than the other three tests, was however
relatively low at .63. This level, however, should still be acceptable as both the sample size (12) and the number of items (20) were small. According to Alderson, Clapham, and Wall (1995) and Borg (1981), K-R 21 normally produces lower reliability coefficients than would be yielded by other methods. In other words, K-R 21 generally provides the minimum conservative estimate of the reliability of a test. However, in hindsight, the shortening of the RC from 30 items to 20 probably negatively affected its reliability.

As Table 3-9 shows, the intercorrelation of all the four tests in the pilot study was high, all over .80 (p < .05). The highest r was for the DVK with VS at .94, and the lowest r was for the MK with RC at .83, which was still high. These coefficients supported Hypothesis 1.

Table 3-8. Pilot study with ESL learners (N = 12): K-R reliability coefficients

<table>
<thead>
<tr>
<th>Test</th>
<th>K-R Reliability Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>RC</td>
<td>.63</td>
</tr>
<tr>
<td>VS</td>
<td>.94</td>
</tr>
<tr>
<td>DVK</td>
<td>.92</td>
</tr>
<tr>
<td>MK</td>
<td>.91</td>
</tr>
</tbody>
</table>

Table 3-9. Pilot study with ESL learners (N = 12): Intercorrelations among the RC, VS, DVK, and MK

<table>
<thead>
<tr>
<th>Test</th>
<th>RC</th>
<th>VS</th>
<th>DVK</th>
</tr>
</thead>
<tbody>
<tr>
<td>VS</td>
<td>.84*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVK</td>
<td>.89*</td>
<td>.94*</td>
<td></td>
</tr>
<tr>
<td>MK</td>
<td>.83*</td>
<td>.87*</td>
<td>.88*</td>
</tr>
</tbody>
</table>

* Significant at .05.
In summary, the results of the pilot study were generally satisfactory. The values of the coefficients obtained through Pearson correlation and K-R 21 revealed an acceptable level of reliability among scores on the RC, VS, DVK, and MK. Therefore, no adjustments were made to the four tests. It was then decided to include the scores of these 12 learners in the data of the main study since there were no changes made to the RC, VS, DVK, and MK after the pilot study, because participants for the main study would be recruited from the same source where the 12 learners for the pilot study were from, and because these learners all met the recruitment criteria for both the main and follow-up studies.

3.5 The Follow-Up Study

Research Question 3 called for a qualitative investigation, following the data collection for the main study.

3.5.1 Participants

From the sample for the main study, a subsample of 12 volunteers, which was the expected size, took part in this follow-up study. However, in selecting these participants, it was not feasible to use methods of random selection because by the time the data collection for the follow-up study was ready to begin, a large number of the participants in the main study had already left Toronto. The reason for this was that the process of data collection for the main study, which involved 35 individual testing sessions all administered by myself, took some time to complete, and most of the participants in the main study were staying at their temporary residences in Toronto for only a limited period of time to attend ESL courses.
Under the circumstances, I made telephone calls in an attempt to reach the remaining participants, using information available on the main-sample list. After numerous attempts, I contacted about 25 and 12 agreed to participate in the follow-up study. Of the 12 participants, 10 were Korean learners and two were Chinese. A profile of these participants appears in Chapter 5, in conjunction with the analyses and results of the follow-up study.

3.5.2 Material and instruments

3.5.2.1 Experimental sentences and text. The follow-up investigation took the form of individual interviews based on four experimental sentences chosen from the RC passages in the main study, and a short experimental text entitled "The Greenhouse", adapted from *Tests in English for Overseas Students* (Kassem, 1982, p. 175). The experimental sentences were chosen on the assumption that they would contain words unknown to some interviewees.

In the experimental text, there were 10 highlighted words I had pre-selected on the same assumption that they would be new to some interviewees. These words are: *greenhouse, feature, indispensable, edible, irrespective, permanent, functional, conduct, devices,* and *free-standing.* These ten words were used as stimulus words to elicit responses from the learners in order to determine: (a) whether they understood the correct meaning of these words, and (b) if these words were unknown to them, what knowledge sources they used in trying to guess the meaning of these words in the comprehension process.

3.5.2.2 Extended Vocabulary Knowledge Scale. This was a version of the Vocabulary Knowledge Scale (Paribakht & Wesche, 1993, p. 15; Wesche & Paribakht, 1993, 1996) upwardly
extended to increase its potential to tap greater depth of vocabulary knowledge by providing room for the learner to indicate a second meaning of a target word and to make a sentence using the target word in its second meaning. The Extended Vocabulary Knowledge Scale (hereafter EVKS) was used to measure the depth of the participants' knowledge of the 10 pre-selected words in the experimental text, in order to provide a profile of the interviewees, in addition to the information provided by their scores on the RC, VS, DVK, and MK. The participants were asked to complete the EVKS in writing. The part added to the original VKS was as follows:

VI. I know more than one meaning of the word. It can mean (synonyms or translations)

1 ______________________________________________________________________ 2 ______________________________________________________________________

VII. The following sentences show all the meanings I can think of for the word. (If you do this section, please also do Section VI.)

1 ______________________________________________________________________

2 ______________________________________________________________________

An 8-level scoring system adapted from Read (1994, p. 7) was used (see below). This system was originally a five-level scale developed by Wesche and Paribakht (1993) and was later modified and extended to eight levels by Read. In the present study, the wording of Levels 6-8 was modified slightly to make the extended part of the scale focus more explicitly on second word meanings in order to match the purpose of the EVKS.

Level 1: The word is not familiar.

Level 2: The word is familiar but the meaning is not known.

Level 3: One meaning of the word is partly known.
Level 4: One meaning of the word is known.
Level 5: One meaning of the word is known and its use is correctly exemplified.
Level 6: A second meaning of the word is partly known.
Level 7: A second meaning of the word is known.
Level 8: A second meaning of the word is known and its use is correctly exemplified.

In scoring, each level stood for one point. The maximum possible score was therefore eight for each stimulus word, and the maximum possible score for the 10 stimulus words was 80. The learners’ responses were scored by two raters, both experienced ESL teachers. When different opinions arose on a response, the average of the two scores was accepted. The interrater agreement was 94.2%.

3.5.2.3 Interview questions. A set of questions was used in the semi-structured interviews to understand how the participants made use of their vocabulary knowledge in comprehending the text and what sources of knowledge they used in the comprehension process. These questions, which are listed below, were used in conjunction with the experimental sentences and text. The participants were asked some or all of the following questions depending on the stimulus words and the availability of time:

1) How do you explain this sentence in your own words?
2) What is the meaning of ____ (a stimulus word) in this sentence?
3) Why did you think this was the meaning of the word here?

---

6 The actual maximum possible score may be lower since not every stimulus word has more than one meaning.
4) Did you consider other meanings before deciding on this one?

5) You have indicated you did not know this word. So how did you work out the meaning of this word?

6) What helped you understand this sentence?

7) Is there anything else that helped you understand this sentence?

More specific questions were developed, based on these generic questions, after the stimulus words were finalized at the beginning of each interview.

3.5.2.4 Reading Strategies Questionnaire. This was a questionnaire about the learners’ ESL reading strategies when they encountered an unfamiliar word. The format of the questionnaire and its first question were adapted from a questionnaire used by Harley and Hart (in preparation) in a study of secondary school students learning French as a second language. There were two general questions in the adapted questionnaire:

1. How often do you do each of the following when you meet an unfamiliar word in reading an English text?

2. What kind of information do you use when trying to guess the meaning of an unfamiliar word in an English text?

The respondents were provided with four choices of time frequency to indicate how often they did each thing in the situation described in the above questions. The frequency choices were: often, sometimes, rarely, and never. For the first question, eight ways of dealing with a new word were described. For the second question, six strategies were described for making use of various levels and aspects of one’s knowledge to deal with an unknown word in
comprehending a text (see Appendix E).

3.5.3 Interview procedures
At the beginning of each interview, I, who conducted all the individual interviews, asked the participant to underline all the unknown or unfamiliar words in the four experimental sentences. Then, the interviewee was given a few minutes to work out the meaning of each sentence. After the interviewee indicated that the task was completed, I would ask the interviewee to explain the meaning of the sentences and some words in the sentences. The words the interviewee underlined were usually the focus of the questions, which I asked based on the generic questions described in Section 3.5.2.3.

When the sentence task was completed, I would ask the interviewee to write the EVKS (see Section 3.5.2.2) to demonstrate level of knowledge of the 10 stimulus words, which would thereafter appear in the experimental text. When the EVKS was completed, the interviewee was presented with the experimental text, and was first asked to underline any unknown or unfamiliar words. After that, the interviewee would be asked to describe the main idea of the text, and then asked questions about the test words. Words underlined by the interviewee qualified as test words. When some of the 10 highlighted words in the experimental text were underlined by the interviewee, these words normally became the focus of questions. Occasionally, I would add a limited number of words to the pool of test words when I noticed that the interviewee did not present correctly the general idea of the experimental text, probably due to misunderstanding some words, or when the interviewee failed to explain correctly in the EVKS a word that the person later did not underline. Due to time constraints, I randomly omitted a small number of
words that might have been used as test words. At the end of the interview, the participant would be asked to complete the Questionnaire on Reading Strategies.

Each interview lasted about 45 minutes. The interviews were carried out either in English or in Chinese. Because I am not proficient in Korean, Korean participants had to converse in English, or occasionally in Chinese when they preferred, especially when they were at a loss for an English word. All the interviews were audio-taped and later transcribed. The Chinese transcripts were all later translated into English.

3.5.4 Analyses of the interview data

The analyses of the interview data were primarily descriptive-interpretive in nature, aided by some quantitative analyses. In order to obtain insights into the role that vocabulary knowledge plays in the reading comprehension process, the interview data were analyzed qualitatively using content analyses, based on an analytic framework grounded on the working definition of depth of vocabulary knowledge (see Chapter 2), as well as on some features of Haastrup’s (1991) and de Bot, Paribakht and Wesche’s (1997) taxonomies of knowledge sources for lexical inferencing. The analysis (see Chapter 6 for details) focused on the following questions.

1. What knowledge sources did the learners use in inferring the meaning of unknown words in the experimental sentences and text?

2. What types of vocabulary knowledge did the learners use in the lexical inferencing process?

---

7 Some of the Korean participants told me that they took Mandarin courses while studying in school, and that Mandarin courses were popular in Korea.
CHAPTER FOUR

THE MAIN STUDY: ANALYSES AND RESULTS

4.1 Overview

This chapter describes the statistical analyses for the main study and the results produced through the five-phase analysis. These five phases, which involved different combinations of the variables, were designed for various specific purposes. Description of the variables used in the five phases was provided in Table 3-4.

4.2 Phase 1 of the Analysis

The analyses carried out in the first phase were preliminary steps. The purpose of this phase of the analysis was to ensure that the data elicited from the two L1 groups were comparable and therefore that the two groups could be treated as one sample in the later analyses.

4.2.1 Means and ranges for the full sample

The full sample’s means, standard deviations, and ranges of the scores on the RC, VS, DVK, and MK were computed in order to provide a general profile of the full sample. Table 4-1 presents the results of these computations.

4.2.2 Comparison of group scores

Following the full-sample analysis, the means, standard deviations, and ranges of the scores on the RC, VS, DVK, and MK obtained by the Korean group and Chinese group were computed separately. The results are provided in Table 4-2.
Table 4-1. Means, standard deviations, and score ranges on the RC, VS, DVK, and MK for the full sample (N=74)

<table>
<thead>
<tr>
<th>Test</th>
<th>MPS²</th>
<th>M</th>
<th>SD</th>
<th>Score range</th>
</tr>
</thead>
<tbody>
<tr>
<td>RC</td>
<td>20</td>
<td>15.89</td>
<td>2.62</td>
<td>9 - 20</td>
</tr>
<tr>
<td>VS</td>
<td>90</td>
<td>62.77</td>
<td>13.02</td>
<td>34 - 86</td>
</tr>
<tr>
<td>DVK</td>
<td>160</td>
<td>119.51</td>
<td>17.75</td>
<td>71 - 148</td>
</tr>
<tr>
<td>MK</td>
<td>30</td>
<td>24.03</td>
<td>5.23</td>
<td>9 - 30</td>
</tr>
</tbody>
</table>

a. MPS = Maximum possible score.

Table 4-2. Means, standard deviations and ranges of the Korean group’s test scores (n=41) and the Chinese group’s test scores (n=33)

<table>
<thead>
<tr>
<th>Test</th>
<th>M Kor</th>
<th>SD Kor</th>
<th>M Chin</th>
<th>SD Chin</th>
<th>Korean</th>
<th>Chinese</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>RC</td>
<td>16.07</td>
<td>2.26</td>
<td>3.02</td>
<td>11 - 20</td>
<td>9 - 20</td>
<td>.44</td>
<td>n.s.*</td>
<td></td>
</tr>
<tr>
<td>VS</td>
<td>65.02</td>
<td>11.72</td>
<td>14.16</td>
<td>46 - 86</td>
<td>34.0 - 84</td>
<td>2.82</td>
<td>n.s.</td>
<td></td>
</tr>
<tr>
<td>DVK</td>
<td>120.54</td>
<td>15.02</td>
<td>20.82</td>
<td>71 - 143</td>
<td>74.0 - 148</td>
<td>.30</td>
<td>n.s.</td>
<td></td>
</tr>
<tr>
<td>MK</td>
<td>25.02</td>
<td>4.77</td>
<td>5.23</td>
<td>9 - 30</td>
<td>9.5 - 30</td>
<td>3.46</td>
<td>n.s.</td>
<td></td>
</tr>
</tbody>
</table>

a. n.s. = Not significant at .05.

It can be observed from Table 4-2 that, although the values of the corresponding parameters of the two groups are not identical, they appear to have similar patterns in general. In confirmation of this observation, oneway ANOVAs found no significant differences between the Chinese and Korean group score means on the RC, VS, DVK, and MK. Based on this finding, it was therefore decided that the two L1 groups could be merged for subsequent analyses in Phase 2 through Phase 4 in order to maintain a substantial sample size, which would enhance the power of the multiple regression analyses.
4.3 Phase 2 of the Analysis

There were three purposes for this phase of the analysis: (a) to obtain the reliability coefficients of the RC, VS, DVK, and MK based on the scores of the whole sample, as compared with the reliability coefficients obtained in the pilot study; (b) to determine the intercorrelations among RC, VS, DVK, and MK; and (c) to determine whether, and to what extent, DVK and MK add to the prediction of RC over and above the prediction already afforded by VS. K-R 21 was used to calculate the reliability coefficients of the four tests. Pearson correlations were computed for the intercorrelations among the same four tests. To fulfill purpose (c), the forced-entry and stepwise procedures of multiple regression were performed.

4.3.1 Reliability of the RC, VS, DVK, and MK

The reliability coefficients of the RC, VS, DVK, and MK were first calculated using the method of rational equivalence based on K-R 21 (see Table 4-3). Also listed in the table for comparison are the reliability coefficients of the same four tests obtained in the pilot study. The standard error of measurement of the scores on the RC, VS, DVK, and MK in the main study were 1.76, 4.12, 5.33 and 2.03, respectively.

Table 4-3 shows that, although the reliability coefficients of the four tests for the main study still ranked in the same order as they did in the pilot study, the values of the four coefficients are slightly lower than their counterparts in the pilot study, with the \( r \) value for the RC, which had been the lowest among the four coefficients in the pilot study, further decreased by .08 to .55. In order to determine whether the \( r \) value for the RC could be improved if the number of items in the RC increased, an estimation of reliability was conducted using the
Spearman-Brown Formula (Linn, 1989, pp. 116). The results of this calculation forecast a reliability of .71 if the RC were to be lengthened to 40 items, and .79 if the RC were increased to 60 items. This indicated that the reliability of RC scores would be improved if the test was lengthened (i.e., as it had originally been devised for the TOEFL). Although an $r$ of .55 was less than satisfactory, I continued the data analysis as planned since this was a post hoc analysis and the results of the Spearman-Brown Formula analysis were satisfactory (see also Section 8.5.2).

**Table 4-3.** Test reliability coefficients in the main study (N=74) and the pilot study (n=12) based on K-R 21

<table>
<thead>
<tr>
<th>Test</th>
<th>Reliability $r$</th>
<th>Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Main Study</td>
<td>Pilot Study</td>
</tr>
<tr>
<td>RC</td>
<td>.55</td>
<td>.63</td>
</tr>
<tr>
<td>VS</td>
<td>.90</td>
<td>.94</td>
</tr>
<tr>
<td>DVK</td>
<td>.91</td>
<td>.92</td>
</tr>
<tr>
<td>MK</td>
<td>.85</td>
<td>.91</td>
</tr>
</tbody>
</table>

4.3.2 Correlations

Table 4-4 reports the results of the one-tailed product-moment correlation analyses for this phase. It is evident that all the correlation coefficients were higher than .60, well over the minimum $r$ value of .50 predicted in Hypothesis 1. However, since the $r$ value for DVK with VS slightly exceeded .80, it indicated collinearity between these two predictor variables. Because of this collinearity, the regression results obtained from these data should be interpreted with caution.
Table 4-4. Pearson correlation coefficients among RC, VS, DVK, and MK for the full sample (N=74)

<table>
<thead>
<tr>
<th>Variable</th>
<th>RC</th>
<th>VS</th>
<th>DVK</th>
</tr>
</thead>
<tbody>
<tr>
<td>VS</td>
<td>.78*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVK</td>
<td>.82*</td>
<td>.82*</td>
<td></td>
</tr>
<tr>
<td>MK</td>
<td>.64*</td>
<td>.69*</td>
<td>.63*</td>
</tr>
</tbody>
</table>

* significant at .05.

4.3.3 Procedures and results of multiple regression

Section 3.2.4.2 provided the theoretical basis for multiple regression procedures in this phase. Briefly, the main purpose for this phase was to determine the extent to which scores on the two vocabulary depth tests, namely, the DVK and MK, can add to the prediction of the RC scores over and above the prediction afforded by the VS scores. For this purpose, forced entry procedures were used in order to manually determine the order of variable entry into the regression equation so that the order would conform to the operational framework specified in Section 3.2.4.2. The analyses focused on examining the changes in the magnitude of the $R^2$, or the variance in the dependent variable that was accounted for by the independent variables (Smith & Glass, 1987). This was achieved by observing the size of the portion of $R^2$ that was added by DVK, as well as by DVK and MK jointly, to cover the dimension of depth of vocabulary knowledge when VS was already in the equation. The stepwise procedure was thus used to remove from the equation any independent variables that did not qualify as predictors of the dependent variable RC. These procedures and $R^2$ changes are summarized in Table 4-5 in addition to the text description below.
1. Using the forced entry procedure, VS was entered into the equation at the first step. At this point, the $R^2$ was $.60$ ($F = 110.15, p < .05$).

2. The entry of DVK at the second step changed the size of the $R^2$ to $.71$, showing an increase of $.11$, or $11\%$ of the explained variance in RC ($F$ change $= 26.26, p < .05$).

3. When MK was added to the equation at the third step, the $R^2$ increased by only $0.01$ to $0.72$ ($F$ change $= 2.27$, not significant at $p = .05$).

4. A stepwise procedure was then conducted in order to remove any independent variables that did not meet the minimum criteria$^1$ for a good predictor. The conclusion of this step saw MK removed, and the $R^2$ was brought back to $.71$, the same level as before MK was added to the equation.

Table 4-5. Regression results of the full sample (N=74) with VS, DVK, and MK as the independent variables

<table>
<thead>
<tr>
<th>Step</th>
<th>Procedure</th>
<th>Variable</th>
<th>Status</th>
<th>$R^2$</th>
<th>$R^2$ change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>forced entry</td>
<td>VS</td>
<td>in</td>
<td>$.60^*$</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>forced entry</td>
<td>DVK</td>
<td>in</td>
<td>$.71^*$</td>
<td>$.11^*$</td>
</tr>
<tr>
<td>3</td>
<td>forced entry</td>
<td>MK</td>
<td>in</td>
<td>$.72^*$</td>
<td>.01</td>
</tr>
<tr>
<td>4</td>
<td>stepwise</td>
<td>MK</td>
<td>out</td>
<td>$.71^*$</td>
<td>-.01</td>
</tr>
</tbody>
</table>

$^*$ Significant at $.05$.

In summary, among the three predictor variables, namely, VS, DVK, and MK, DVK added a unique portion ($11\%$) of explained variance in RC on top of the $60\%$ variance accounted for

$^1$ The default values of the SPSS for the minimum criteria of stepwise procedures are: $F$-to-enter $= 3.84$; probability of $F$-to-enter $= .05$; $F$-to-remove $= 2.71$; and probability of $F$-to-remove $= .10$. All the analyses in the present study used these default values.
by VS. In contrast, MK's contribution of 1% additional explained variance in RC was negligible.

As an additional procedure, a fresh model was built to determine (a) the predictive power of DVK in forecasting RC scores when DVK was entered into the equation first, and (b) the size of prediction of RC added by VS on top of the prediction already afforded by DVK. The results showed that the $R^2$ was .68 ($F = 152.17, p < .05$) when DVK alone was in the equation. After VS was added to the equation, the $R^2$ increased by 0.03 to 0.71 ($F$ change $= 8.03, p < .05$). That is, VS added 3% explained variance in RC on top of the prediction already provided by DVK.

4.4 Phase 3 of the Analysis

This phase, also involving full-sample analyses, was designed for exploratory analyses. In this phase, RC was still the dependent variable, and VS remained an independent variable representing the size of vocabulary knowledge. However, the two other independent variables used in Phase 2, DVK and MK, were split into four variables, namely, DVKmng, DVKcol, MKaffix, and MKpart. In other words, the DVK and MK were now divided into four separate tests.

The objectives of this phase were: (a) to determine the intercorrelations among RC, VS, DVKmng, DVKcol, Mkaffix, and Mkpart; and (b) to examine the magnitude of contribution each independent variable makes to the prediction of RC, in order to determine which tests, and therefore components, of vocabulary knowledge are more useful in predicting reading comprehension scores. For objective (a), Pearson product-moment correlations were employed, and for objective (b), the forced entry and stepwise procedures of multiple regression were performed.
4.4.1 Correlations

Table 4-6 reports the one-tailed product-moment correlations. The table reveals that while the values of most of the correlation coefficients were higher than .50 ($p < .05$), the $r$ was .47 ($p<.05$) for Mkpart with DVKmng, and .48 ($p < .05$) for Mkpart with DVKcol. These values were slightly below .50, the minimum level predicted in Hypothesis 1.

Table 4-6. Pearson correlation coefficients among RC, VS, DVKmng, DVKcol, Mkaffix, and Mkpart for the full sample (N=74)

<table>
<thead>
<tr>
<th>Variable</th>
<th>RC</th>
<th>VS</th>
<th>DVKmng</th>
<th>DVKcol</th>
<th>Mkaffix</th>
</tr>
</thead>
<tbody>
<tr>
<td>VS</td>
<td>.78*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVKmng</td>
<td>.79*</td>
<td>.78*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVKcol</td>
<td>.78*</td>
<td>.75*</td>
<td>.80*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mkaffix</td>
<td>.61*</td>
<td>.75*</td>
<td>.66*</td>
<td>.57*</td>
<td></td>
</tr>
<tr>
<td>Mkpart</td>
<td>.55*</td>
<td>.54*</td>
<td>.47*</td>
<td>.48*</td>
<td>.55*</td>
</tr>
</tbody>
</table>

* Significant at .05. RC = Score on the Reading Comprehension Test; VS = Score on the vocabulary size test; DVKmng = Score on the meaning part of the DVK; DVKcol = Score on the collocation part of the DVK; Mkaffix = Score on the affix part of the MK; Mkpart = Score on the part-of-speech part of the MK.

4.4.2 Procedures and results of multiple regression

To fulfil objective (b) described in Section 4.4, the stepwise and forced entry procedures of multiple regression were performed. The reason that the stepwise procedure was used in this phase was to select predictors of RC from among DVKmng, DVKcol, Mkaffix, Mkpart, and VS. The forced entry procedures were used in order to manually determine the order of variable entry into the regression equation so that some purposeful comparisons could be made of each selected predictor variable (Smith & Glass, 1987): After predictor variables were selected through the stepwise procedure, each qualified variable was compared with the rest of the
selected variables individually. Two steps were carried out at each stage of comparison. At the first step, one variable was entered. At the second step, another variable was entered. The $R^2$ change after the entry of the second variable indicates the additional portion of explained variance in RC contributed by the second variable. This procedure was repeated until all the variables selected by the stepwise procedure were compared.

Table 4-7. Regression results of the full sample (N=74) using the forced entry procedures with VS, DVKmng, and DVKcol as the selected independent variables

<table>
<thead>
<tr>
<th>Stage</th>
<th>Step</th>
<th>Variable</th>
<th>Status</th>
<th>$R^2$</th>
<th>$R^2$ change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>DVKmng</td>
<td>in</td>
<td>.63*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>VS</td>
<td>in</td>
<td>.69*</td>
<td>.07*</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>VS</td>
<td>in</td>
<td>.60*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>DVKmng</td>
<td>in</td>
<td>.69*</td>
<td>.09*</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>DVKmng</td>
<td>in</td>
<td>.63*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>DVKcol</td>
<td>in</td>
<td>.68*</td>
<td>.06*</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>DVKcol</td>
<td>in</td>
<td>.60*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>DVKmng</td>
<td>in</td>
<td>.68*</td>
<td>.08*</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>DVKcol</td>
<td>in</td>
<td>.60*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>VS</td>
<td>in</td>
<td>.69*</td>
<td>.09*</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>VS</td>
<td>in</td>
<td>.60*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>DVKcol</td>
<td>in</td>
<td>.68*</td>
<td>.08*</td>
</tr>
</tbody>
</table>

* Significant at .05 level. VS = Score on the vocabulary size test; DVKmng = Score on the meaning part of the DVK; DVKcol = Score on the collocation part of the DVK.

The stepwise procedure performed at the beginning of the process allowed three independent variables, namely, DVKmng, VS, and DVKcol, into the equation, which means that the two other independent variables, Mkaffix and Mkpart, did not meet the minimum criteria of $F$-values and $p$-values set up by the SPSS and therefore did not qualify as contributors to
predicting RC. Mkaffix and Mkpart were thus discarded in the subsequent analyses. Following
the stepwise procedure, the forced entry analysis was conducted in comparing the predictive
values of DVKmng, VS, and DVKcol. The procedures and $R^2$ changes are reported in Table 4-7.

In summary, among the five predictor variables involved, Mkaffix and Mkpart failed to
qualify as predictors of RC. The remaining three variables, DVKmng, VS and DVKcol, all
qualified as distinctive RC predictors with similar powers.

4.5 Phase 4 of the Analysis

Like Phase 3, this phase of the analysis was exploratory in nature. The variables involved in
Phase 4 were RC, VS, DVKold, and DVKnew. DVKold represented Items 1-30 of the DVK,
or the general vocabulary knowledge part of the DVK. DVKnew stood for Items 31-40 of the
DVK, or the part on RC text-related vocabulary knowledge in the DVK. The objectives of this
phase were to determine (a) the respective power of VS, DVKold, and DVKnew as RC
predictors; and (b) how the four variables correlated with one another. In other words, this phase
was designed to find out whether scores representing knowledge of words appearing in the RC
texts could serve as a better predictor of RC scores than scores on general vocabulary
knowledge, or test words that did not appear in the RC texts.

To meet these two objectives, the forced entry and stepwise procedures of multiple
regression were employed as the primary methods for the data analyses, with RC as the
dependent variable and VS, DVKold, and DVKnew as the independent variables. Product-
moment correlations were also performed to examine the intercorrelations among the four
variables.
4.5.1 Correlations

Table 4-8 lists the results of the one-tailed correlational analysis in Phase 4. It can be observed that all the correlations for the variable pairs among RC, VS, DVKold, and DVKnew were over .70 ($p < .05$).

Table 4-8. Pearson correlation coefficients among RC, VS, DVKold, and DVKnew for the full sample (N=74)

<table>
<thead>
<tr>
<th>Variable</th>
<th>RC</th>
<th>VS</th>
<th>DVKold</th>
</tr>
</thead>
<tbody>
<tr>
<td>VS</td>
<td>.78*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVKold</td>
<td>.80*</td>
<td>.79*</td>
<td></td>
</tr>
<tr>
<td>DVKnew</td>
<td>.73*</td>
<td>.72*</td>
<td>.72*</td>
</tr>
</tbody>
</table>

* Significant at .05. VS = Score on the vocabulary size test; DVKold = Score on the general vocabulary knowledge part of the DVK (Items 1-30); DVKnew = Score on the RC-text-related part of the DVK (Items 31-40)

4.5.2 Procedures and results of multiple regression

In order to determine the predictive values of VS, DVKold, and DVKnew in forecasting RC, the stepwise and forced entry procedures were performed. As in Phase 3, all the predictor variables selected by the stepwise procedure in this phase were compared individually with one another in a two-step equation building process. The analyses focused on examining the changes in the $R^2$.

A stepwise procedure conducted at the beginning of the process selected all the three independent variables, namely, DVKold, VS, and DVKnew, as qualified predictors of RC. This means that the three variables met the minimum criteria for being RC predictors. Comparisons of VS, DVKold, and DVKnew via the forced-entry procedures and their $R^2$ changes are summarized in Table 4-9.
Table 4-9. Regression results of the full sample (N=74) using the forced entry procedures with VS, DVKold, and DVKnew as the independent variables

<table>
<thead>
<tr>
<th>Stage</th>
<th>Step</th>
<th>Variable</th>
<th>Status</th>
<th>R²</th>
<th>R² change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>DVKold</td>
<td>in</td>
<td>.63*</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>DVKnew</td>
<td>in</td>
<td>.68*</td>
<td>.05*</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>DVKnew</td>
<td>in</td>
<td>.53*</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>DVKold</td>
<td>in</td>
<td>.68*</td>
<td>.15*</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>VS</td>
<td>in</td>
<td>.60*</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>DVKold</td>
<td>in</td>
<td>.69*</td>
<td>.09*</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>DVKold</td>
<td>in</td>
<td>.63*</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>VS</td>
<td>in</td>
<td>.69*</td>
<td>.06*</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>VS</td>
<td>in</td>
<td>.60*</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>DVKnew</td>
<td>in</td>
<td>.66*</td>
<td>.06*</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>DVKnew</td>
<td>in</td>
<td>.53*</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>VS</td>
<td>in</td>
<td>.66*</td>
<td>.13*</td>
</tr>
</tbody>
</table>

* Significant at .05. VS = Score on the vocabulary size test; DVKold = Score on the general vocabulary knowledge part of the DVK (Items 1-30); DVKnew = Score on the RC-text-related part of the DVK (Items 31-40)

As shown in the table, the relative strength of DVKold in contrast to that of DVKnew was particularly noticeable: The former added 15% explained variance in RC over and above the 53% variance already explained by DVKnew, whereas in a reversed-order comparison, DVKnew added only 5% explained variance in RC on top of the 63% variance that was already accounted for by DVKold. Meanwhile, DVKold and VS appeared to be similarly powerful as RC predictors, and each made a unique contribution to the prediction.

4.6 Phase 5 of the Analysis

In the fifth phase of the analysis, the sample was divided into two groups according to the learners’ L1 background, namely, Korean and Chinese, and split-file analyses were carried out.
Identical measures were applied to the two L1 groups. The purpose of this phase of the analyses was to find out whether similar results were obtained by each L1 group when their scores were analyzed separately. In order to fulfil this purpose, product-moment correlations, and the forced entry and stepwise procedures of multiple regression were conducted with RC as the dependent variable and L1, VS, DVK, and MK as the independent variables.

4.6.1 Correlations

Tables 4-11 reports the one-tailed Pearson correlations for the Korean and the Chinese groups, which were obtained from the split-sample analyses.

### Table 4-10. Pearson correlation coefficients among RC, VS, DVK, and MK for the Korean group (n=41) and Chinese group (n=33)

<table>
<thead>
<tr>
<th>Var</th>
<th>RC Korean</th>
<th>Chinese</th>
<th>VS Korean</th>
<th>Chinese</th>
<th>DVK Korean</th>
<th>Chinese</th>
</tr>
</thead>
<tbody>
<tr>
<td>VS</td>
<td>.67*</td>
<td>.87*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVK</td>
<td>.72*</td>
<td>.89*</td>
<td>.73*</td>
<td>.89*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MK</td>
<td>.57*</td>
<td>.70*</td>
<td>.66*</td>
<td>.68*</td>
<td>.61*</td>
<td>.66*</td>
</tr>
</tbody>
</table>

* Significant at .05.

In Table 4-10, all the correlations were higher than .50 ($p < .05$), the minimum hypothesized level. However, the table shows that the correlations among RC, VS, DVK, and MK for the Chinese group were higher than their counterparts for the Korean group. This may be because the Korean group had smaller score variances than the Chinese group, since the standard deviations of the means of the four tests for the Korean group were all smaller than those for the Chinese group.
4.6.2 Procedures and results of multiple regression

The purpose of performing the multiple regression procedures in this phase was to find out whether similar statistics were obtained by the two L1 groups when the data of each of the two groups were analyzed separately. The focus of the analyses was on the changes in $R^2$, or the variance in RC that was accounted for by the independent variables, which were, in this case, VS, DVK, and MK. These procedures and $R^2$ changes are summarized in Tables 4-11 and 4-12, which show that the patterns of the regression results from the two L1 groups were generally similar, although the actual values of the figures obtained were not identical.

Table 4-11. Regression results for the Korean group (n=41) with VS, DVK, and MK as the independent variables

<table>
<thead>
<tr>
<th>Step</th>
<th>Procedure</th>
<th>Variable</th>
<th>Status</th>
<th>$R^2$</th>
<th>$R^2$ change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>forced entry</td>
<td>VS</td>
<td>in</td>
<td>.45*</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>forced entry</td>
<td>DVK</td>
<td>in</td>
<td>.57*</td>
<td>.12*</td>
</tr>
<tr>
<td>3</td>
<td>forced entry</td>
<td>MK</td>
<td>in</td>
<td>.57*</td>
<td>.01</td>
</tr>
<tr>
<td>4</td>
<td>stepwise</td>
<td>MK</td>
<td>out</td>
<td>.58*</td>
<td>-.01</td>
</tr>
</tbody>
</table>

* Significant at .05.

Table 4-12. Regression results for the Chinese group (n=33) with VS, DVK, and MK as the independent variables

<table>
<thead>
<tr>
<th>Step</th>
<th>Procedure</th>
<th>Variable</th>
<th>Status</th>
<th>$R^2$</th>
<th>$R^2$ change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>forced entry</td>
<td>VS</td>
<td>in</td>
<td>.75*</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>forced entry</td>
<td>DVK</td>
<td>in</td>
<td>.82*</td>
<td>.07*</td>
</tr>
<tr>
<td>3</td>
<td>forced entry</td>
<td>MK</td>
<td>in</td>
<td>.83*</td>
<td>.01</td>
</tr>
<tr>
<td>4</td>
<td>stepwise</td>
<td>MK</td>
<td>out</td>
<td>.82*</td>
<td>-.01</td>
</tr>
</tbody>
</table>

* Significant at .05.
CHAPTER FIVE

THE MAIN STUDY: INTERPRETATION OF FINDINGS

5.1 Overview

This chapter summarizes the findings obtained from the data analyses for the main study and answers Research Questions 1 and 2 by testing the two hypotheses stated in Chapter 2. Research Question 3 is dealt with separately in Chapters 6 and 7 through the follow-up study.

5.2 How the Various Tests Correlated

The purpose of this section is to test Hypothesis 1, which reads: In ESL, both vocabulary size and depth of vocabulary knowledge are important components in the vocabulary knowledge-reading comprehension chain, and correlations of these two components with reading comprehension, and with each other, will all be at a minimum level of \( r = .50 \) \((p < .05)\).

The results of product-moment correlation analyses obtained in Phases 2, 3 and 4 of the data analyses in the main study address this hypothesis. Each of these three phases examined the intercorrelations among a different combination of variables.

5.2.1 Intercorrelations among RC, VS, DVK, and MK

According to the results of the Pearson correlation analyses performed in the second phase of the data analyses (see Table 4-4), all the Pearson correlation coefficients obtained for the variable pairs of RC, VS, DVK, and MK were higher than .60 \((p < .05)\). In other words, high and positive intercorrelations existed among the four tests. The correlation between DVK and RC was among the highest, which indicates that, for this given sample and among the four tests
under discussion, the scores on the DVK, which represented the meaning and collocation components of depth of vocabulary knowledge, were very highly correlated with learners' reading comprehension scores, and therefore should be most closely associated with the test-takers' general academic reading comprehension levels. On the other hand, the high correlation between the scores on the DVK and VS suggests that learners' scores on the depth and breadth dimensions of the vocabulary knowledge are closely, and positively, associated. Therefore, development of the two dimensions appears interdependent. This confirms the hypothesized relationship between depth and breadth of vocabulary knowledge proposed in Section 2.3.3.2.

5.2.2 Intercorrelations among RC, VS, DVKmng, DVKcol, MKaffix, and MKpart

According to the results of the Pearson correlation analyses performed in the third phase of the data analyses (see Table 4-6), most of the correlations obtained for the variable pairs of RC, VS, DVKmng, DVKcol, and Mkaffix were higher than .50 ($p < .05$), which was the minimum hypothesized correlation level. However, the pairing of Mkpart with two variables failed to produce satisfactory results: The correlation for Mkpart with DVKmng and for Mkpart with DVKcol were both slightly below the hypothesized minimum level of .50 ($p < .05$). That is, when the DVK and MK were split into four separate tests of word meaning, word collocation, affixation and part of speech, the scores on the part-of-speech subtest, which was one of the two sub-components of the morphological knowledge component of depth of vocabulary knowledge tested in this study, just barely failed to meet the expected minimum level of correlation with the scores on the meaning and collocation components of the depth of vocabulary knowledge.

These results indicate (a) that there were positive, and relatively strong, intercorrelations
among the scores on most of the components of vocabulary knowledge, including such aspects as vocabulary size, word meaning, word collocation, and affixation, and (b) that performance on these components of vocabulary knowledge correlated well with the learners' performance on general academic reading comprehension levels. These findings suggest that scores on most sub-components of vocabulary knowledge are closely, and positively, associated with one another, as well as with reading comprehension levels.

5.2.3 Intercorrelations among RC, VS, DVKold, and DVKnew

According to the results of the Pearson correlation analyses (see Table 4-8), all the product-moment coefficients for the variable pairs of RC, VS, DVKold, and DVKnew were higher than .70 ($p < .05$). This indicates that scores on both of the two DVK subtests, namely, the original 30-item part, and the new 10-item part with its head words selected from the RC texts, correlate highly with each other, as well as with the scores on the RC and the VS, respectively.

5.3 Lexical Knowledge Tests as Predictors of Reading Comprehension

Hypothesis 2 states: For ESL learners whose vocabulary size is beyond the 3,000-word threshold for reading comprehension, scores on depth of vocabulary knowledge will make a unique, and distinctive, contribution to the prediction of reading comprehension scores, over and above the prediction afforded by vocabulary size scores. The results from the multiple regressions that were performed in Phases 2, 3 and 4 address this hypothesis.
5.3.1 VS, DVK, and MK as predictors of RC

In Phase 2 of the data analyses, the changes in the magnitude of the R²'s of VS, DVK, and MK were observed by performing the forced entry and stepwise procedures of the multiple regression, in order to determine whether the three independent variables could predict RC, and whether DVK and MK would add to the prediction of RC over and above the prediction already provided by VS. The results (see Table 4-5) showed that both VS and DVK contributed significantly to the prediction of RC, and that DVK added a noticeable portion (11%) of explained variance in RC on top of the prediction already afforded by VS. In other words, DVK, which represented scores on two components of depth of vocabulary knowledge (lexical meaning and lexical collocation), made a unique, and distinctive, contribution to the prediction of scores on academic reading comprehension on top of the prediction provided by scores on vocabulary size.

In contrast, it was determined in the same process that MK was not capable of predicting RC because of the negligible magnitude that MK added to the explained variance in RC during the regression model building process, and also because MK was eliminated from the regression model by the stepwise procedure.

Additional regression analyses indicated that DVK was slightly more powerful than VS in predicting RC, based on their initial contributions, as the first predictor variable in the equation, to the explanation of the variance in RC, and on the proportion of the additional RC variance they accounted for as the second predictor variable entered in the equation.
5.3.2 VS, DVKmng, DVKcol, Mkaffix, and Mkpart as predictors of RC

The results of multiple regression procedures performed in Phase 3 (see Table 4-7) revealed a number of findings. First, Mkaffix, and Mkpart did not emerge as significant predictors. That is, scores on both the affixation and part-of-speech subtests, representing the lexical-morpho-syntactic component of depth of vocabulary knowledge, were not useful in predicting learners’ academic reading comprehension levels in this study. Second, in comparing DVKmng, VS, and DVKcol as predictors of RC when DVKmng and DVKcol were treated as scores on two separate depth-of-vocabulary-knowledge tests, the three variables appeared to have similar levels of predicting powers.

In other words, scores on the VS as a vocabulary size test, on the word meaning subtest of the DVK, and on the word collocation subtest of the DVK, were all somewhat unique predictors of reading comprehension levels, though they also shared common properties. At the same time, scores on the two morphological knowledge subtests of the MK, namely, the affixation and the part-of-speech subtest, failed to contribute meaningfully to the prediction of reading comprehension scores. This may be due to a ceiling effect, as suggested by the score distribution of especially the part-of-speech subtest. It seems that both subtests of the MK were too easy for a considerable number of participants who had a relatively high level of ESL proficiency, since the scores from both subtests were, to varied extents, negatively skewed (Krathwohl, 1993). For the affixation subtest, the mean was 7.32 (SD 2.11), and the median was 7.5 on a scale of 10. In the case of the part-of-speech subtest, the mean was 16.70 (SD 3.77), and the median was 18.00 on a scale of 20.
5.3.3 VS, DVKold, and DVKnew as predictors of RC

Based on the $R^2$ changes obtained through the multiple regression procedures (see Table 4-9), VS, DVKold, and DVKnew, representing respectively scores on vocabulary size, depth of vocabulary knowledge in general, and depth of vocabulary knowledge related to the reading comprehension test, were all statistically significant predictors of RC, each contributing with some unique basis to explaining the variance in RC. Of these three predictors, DVKold was found to be the most powerful predictor of RC, followed by VS, and then by DVKnew. These findings suggest (a) that even at the shortened length of 30 items, the original DVK was still a powerful predictor of RC, and (b) that depth of the knowledge of the 10 adjectives from the RC texts did not provide as much prediction of reading comprehension scores as that of general vocabulary did. However, this result should be interpreted with caution because the two DVK subtests contained different numbers of items (see also Section 8.5.3).

5.4 Comparing the Results of the Korean Group and the Chinese Group

The same pattern emerged from the regression analyses in Phase 5: In both the analyses for the Korean and Chinese groups, MK was removed by the stepwise procedure because of its very weak predicting power; for both the Korean and Chinese groups, DVK added a significant portion of explained variance in RC over and above the prediction already provided by VS.

5.5 Summary

5.5.1 Addressing Hypothesis One

The following findings address Hypothesis 1.
1. The Pearson correlations between the reading comprehension scores represented by RC, vocabulary size scores represented by VS, and depth of vocabulary knowledge scores represented by DVK and MK, were all over .60 ($p < .05$). This finding supports Hypothesis 1.

2. The Pearson correlations between the scores on the RC, the VS, and the two components of depth of vocabulary knowledge represented by variables DVKmng and DVKcol, were all over .70 ($p < .05$). This finding further confirms Hypothesis 1.

3. The scores on the affixation aspect of the morphology component correlated fairly well with the scores on the RC, the VS, the word meaning and word collocation subtests of the DVK, and the part-of-speech subtest, which was the second aspect of the morphology component in the present study. This finding also supports Hypothesis 1.

4. However, when the part-of-speech subtest was put in the spotlight, the situation became slightly different. Of the five related Pearson correlation coefficients obtained, only three $r$ values (for MKpart with RC, MKpart with VS, MKpart with MKaffix) exceeded .50 ($p < .05$). The other two (for MKpart with DVKmng and for MKpart with DVKcol) were both just under .50 ($p < .05$). This suggests that scores on one aspect of morphological knowledge, namely, part-of-speech, does not correlate highly with two other components of depth of vocabulary knowledge, i.e., meaning and collocation, although a correlation was nonetheless evident here. This finding provides only partial support for Hypothesis 1.

5. The Pearson correlations between the scores on depth of general vocabulary knowledge (represented by DVKold), on knowledge of 10 adjectives appearing in the RC (represented by DVKnew), on the RC, and on the VS were all over .70 ($p < .05$). This strongly supports
Hypothesis 1.

In summary, with the exception of the part-of-speech subtest, all the vocabulary knowledge and reading comprehension measures tested in this study correlated strongly, and positively, with one another. The values of their correlation coefficients were all over .50 ($p < .05$), which is the minimum level hypothesized in Chapter 2. Therefore, Hypothesis 1 is applicable to scores on the RC, the VS, the DVK, the MK as one aggregate test, and the affixation subtest of the MK, whether or not the DVK is regarded as one entity or separated as two subtests. However, when the part-of-speech subtest is joined in the ranks, the results of the correlational analyses provide only partial support to the hypothesis, since two variable pairs (MKpart with DVKmng and MKpart with DVKcol) were both just under the hypothesized minimum level of .50 ($p < .05$).

5.5.2 Addressing Hypothesis Two

The following findings from the regression analyses address Hypothesis 2:

1. The scores on the DVK, which represented word meaning and word collocation components of depth of vocabulary knowledge, proved to be a powerful predictor of reading comprehension scores, as scores on the DVK added a significant portion of explained variance in RC on top of the prediction provided by the scores on the vocabulary size.

2. The scores on aspects of morphological knowledge, whether as one aggregate test or two separate subtests, did not effectively predict reading comprehension scores.

The findings from the multiple regression analyses support Hypothesis 2 with qualification because, while the DVK scores examined in this study turned out to be a powerful predictor of reading comprehension scores, the scores on the MK, a test purporting to represent one
component of depth of vocabulary knowledge, failed to prove itself as a meaningful predictor of reading comprehension scores, whether treated as one entity or two separate subtests.

Nevertheless, the DVK, a test representing word meaning and word collocation components of depth of vocabulary knowledge, emerged as a useful, and relatively unique, predictor of reading comprehension scores, whether or not it was treated as one vocabulary knowledge test, or as two subtests in two different ways (as the word meaning subtest and the word collocation subtest, or as the depth of general vocabulary knowledge subtest and the subtest of depth of knowledge of RC-text-related words). It nonetheless also shared many common elements with other measures of vocabulary utilized in this study, such as the VS, which was also closely associated with the RC scores.

These findings suggest that, properly designed and developed, depth of vocabulary knowledge tests can provide unique prediction of reading comprehension scores on top of the prediction afforded by vocabulary size scores. Based on Anderson and Freebody (1981), Meara and Jones (1988) contend that vocabulary size is an important factor in a learner’s L2 performance. In fact, Anderson and Freebody’s (1981) concept of vocabulary knowledge also includes depth of vocabulary knowledge as an important dimension. Now the findings from the present study have provided empirical evidence that depth of vocabulary knowledge has an integral, and fundamental, place in the relationship between vocabulary knowledge and reading comprehension. This lays a theoretical foundation for using depth-of-vocabulary-knowledge measures for placement purposes in educational settings.

The present findings that both vocabulary size and depth of vocabulary knowledge scores are positively, and strongly, associated with reading comprehension scores have provided
empirical support for the instrumentalist hypothesis (Anderson & Freebody, 1981) with a new perspective, i.e., the role of depth of vocabulary knowledge in this relationship, although the present study did not investigate whether vocabulary knowledge is a direct factor in the causal chain resulting in reading comprehension.
CHAPTER SIX
THE FOLLOW-UP STUDY: ANALYSES AND RESULTS

6.1 Overview

As described in Chapter 3, a follow-up study was carried out after the completion of the data collection for the main study. The follow-up study was based on individual interviews focused on learners' processing of lexical meaning and on a questionnaire concerning reading strategies. The present chapter discusses the procedures for the analysis of the qualitative data and reports the findings from these analyses. These address the third major research question posed in Chapter 1: How do ESL learners make use of their depth of vocabulary knowledge in the process of comprehending a general academic text?

Specifically, this follow-up study endeavoured to find out how learners made use of their depth of vocabulary knowledge to deal with unknown or unfamiliar words in the process of comprehending the experimental sentences and text. This chapter is divided into a number of sections. First, a profile of the participants in the study is presented. This includes information obtained from a background questionnaire, their individual test scores in the main study, and their scores on the Extended Vocabulary Knowledge Scale (EVKS) administered as part of this study. Second, the results from the analysis of interview data are presented. The analysis was carried out on interview protocols using content analysis that categorized the knowledge sources participants reported resorting to in inferring the meanings of unknown or unfamiliar words as they read the experimental sentences and text. The analysis provides insights into what participants actually did when encountering unfamiliar words in English texts in circumstances where they did not have access to a dictionary or help from a teacher or a friend. To
demonstrate this, excerpts from some interviews are presented and qualitatively analyzed to show different approaches that ESL learners used in inferring lexical meanings while reading. Third, data collected from a Reading Strategies Questionnaire are quantified and tabulated to provide information about the participants' self-reported behaviours and strategies for dealing with unknown words in comprehending texts.

6.2 Profile of the Sample

The profiles of the 12 learners who participated are provided in Table 6-1. Table 6-2 lists the participants' scores on the RC, VS, DVK, and MK, which are extracted from the data for the main study, and their scores on the EVKS. The EVKS scores were obtained using the rating system described in Section 3.5.2.2, in order to provide information on the participants' relative depth of knowledge of the 10 pre-selected stimulus words in the experimental text. Ranking of the 12 participants in the study was in a descending order based on their scores on the DVK because the sole purpose of the present study was to assess the role of depth of vocabulary knowledge in reading comprehension, and the scores on the DVK, representing depth of vocabulary knowledge, proved in the main study to be a powerful predictor of participants' levels of reading comprehension.

It may be seen from Table 6-1 that the ages of the 12 participants averaged 26.4 years. Their duration of residence in Canada averaged 3.2 months. Among the 12 participants, 10 were native Korean speakers and two spoke Chinese as their L1. Except for one participant who was a high school graduate only, the 11 other participants all had some post-secondary education.
Table 6-1. Backgrounds of the participants in the follow-up study (n = 12)

<table>
<thead>
<tr>
<th>Par#</th>
<th>L1</th>
<th>Age</th>
<th>Gender</th>
<th>Educ. level</th>
<th>Duration of stay in Canada</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>Korean</td>
<td>41</td>
<td>female</td>
<td>univ grad</td>
<td>2 months</td>
</tr>
<tr>
<td>#2</td>
<td>Korean</td>
<td>22</td>
<td>female</td>
<td>univ 4th yr</td>
<td>10 months</td>
</tr>
<tr>
<td>#3</td>
<td>Korean</td>
<td>22</td>
<td>female</td>
<td>univ 4th yr</td>
<td>5 months</td>
</tr>
<tr>
<td>#4</td>
<td>Korean</td>
<td>25</td>
<td>female</td>
<td>univ grad</td>
<td>3 months</td>
</tr>
<tr>
<td>#5</td>
<td>Korean</td>
<td>25</td>
<td>female</td>
<td>univ grad</td>
<td>1 month</td>
</tr>
<tr>
<td>#6</td>
<td>Korean</td>
<td>24</td>
<td>male</td>
<td>univ 4th yr</td>
<td>1 month</td>
</tr>
<tr>
<td>#7</td>
<td>Korean</td>
<td>23</td>
<td>female</td>
<td>univ grad</td>
<td>8 months</td>
</tr>
<tr>
<td>#8</td>
<td>Korean</td>
<td>25</td>
<td>female</td>
<td>univ grad</td>
<td>1 month</td>
</tr>
<tr>
<td>#9</td>
<td>Korean</td>
<td>19</td>
<td>female</td>
<td>univ 2nd yr</td>
<td>1 month</td>
</tr>
<tr>
<td>#10</td>
<td>Korean</td>
<td>23</td>
<td>male</td>
<td>high school</td>
<td>4 months</td>
</tr>
<tr>
<td>#11</td>
<td>Chinese</td>
<td>38</td>
<td>male</td>
<td>univ grad</td>
<td>1 month</td>
</tr>
<tr>
<td>#12</td>
<td>Chinese</td>
<td>30</td>
<td>female</td>
<td>univ grad</td>
<td>1 month</td>
</tr>
</tbody>
</table>

Table 6-2. Participants’ individual scores on the RC, VS, DVK, MK, & EVKS (n = 12)

<table>
<thead>
<tr>
<th>Par#</th>
<th>RC (20)(^a)</th>
<th>VS (90)</th>
<th>DVK (160)</th>
<th>MK (30)</th>
<th>EVKS (80)</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>19</td>
<td>78</td>
<td>142</td>
<td>28.0</td>
<td>38.5</td>
</tr>
<tr>
<td>#2</td>
<td>19</td>
<td>62</td>
<td>135</td>
<td>28.0</td>
<td>32.0</td>
</tr>
<tr>
<td>#3</td>
<td>17</td>
<td>71</td>
<td>135</td>
<td>29.0</td>
<td>42.0</td>
</tr>
<tr>
<td>#4</td>
<td>20</td>
<td>86</td>
<td>133</td>
<td>29.0</td>
<td>49.5</td>
</tr>
<tr>
<td>#5</td>
<td>15</td>
<td>64</td>
<td>127</td>
<td>27.0</td>
<td>34.0</td>
</tr>
<tr>
<td>#6</td>
<td>16</td>
<td>85</td>
<td>125</td>
<td>29.0</td>
<td>41.5</td>
</tr>
<tr>
<td>#7</td>
<td>17</td>
<td>72</td>
<td>124</td>
<td>26.0</td>
<td>35.0</td>
</tr>
<tr>
<td>#8</td>
<td>16</td>
<td>49</td>
<td>121</td>
<td>18.0</td>
<td>44.0</td>
</tr>
<tr>
<td>#9</td>
<td>14</td>
<td>51</td>
<td>114</td>
<td>23.5</td>
<td>31.0</td>
</tr>
<tr>
<td>#10</td>
<td>16</td>
<td>55</td>
<td>110</td>
<td>23.0</td>
<td>19.5</td>
</tr>
<tr>
<td>#11</td>
<td>15</td>
<td>54</td>
<td>107</td>
<td>21.5</td>
<td>32.0</td>
</tr>
<tr>
<td>#12</td>
<td>12</td>
<td>41</td>
<td>94</td>
<td>26.0</td>
<td>29.0</td>
</tr>
<tr>
<td>M</td>
<td>16.33</td>
<td>64</td>
<td>122.25</td>
<td>25.67</td>
<td>35.67</td>
</tr>
</tbody>
</table>

\(^a\): the number in parentheses indicates the maximum possible score for the test.
The means and ranges in Table 6-2 are generally close to those in the main study (see Table 4-1), and therefore this subsample can be considered generally representative of the larger sample used in the main study.

6.3. Selection of the Test Words for the Individual Interviews

The following three criteria were used to select the test words for each individual interview.

1. Before being called on to explain the experimental sentences and text, the interviewee was asked to underline all the words that were unknown or unfamiliar to him or her. These underlined words generally served as the main source of test words in the interview.

2. When the interviewee's explanations of the meaning of any of the 10 target words in the EVKS indicated that he or she could not define the target word properly, this target word could be used as a test word.

3. In the process of presenting the general idea of the experimental sentences and text, when there were signs that the interviewee partially or totally misunderstood an experimental sentence or the experimental text possibly due to his or her misunderstanding of a particular word, this word could be chosen as a test word.

However, not all the words that met one of the above three criteria were used as test words. Because the time limit for each interview was not to exceed 45 minutes as promised in the letter of informed consent, some potential test words had to be dropped in some circumstances.
6.4 Analysis of the Interview Data

During the interviews, a set of generic questions (see Section 3.5.2.3) was used to form specific questions for each interviewee. I developed more specific questions based on these generic questions after test words were identified at the beginning of each interview.

In order to obtain insights into how these learners dealt with unknown words in their reading comprehension processes, the interview data were coded and analyzed both qualitatively and quantitatively using content analysis according to a framework developed specifically for this purpose (see the following section). The present analysis was designed to address the following two questions, the answers to which, together with the findings from the questionnaire data, will provide an answer to the third major research question posed in Chapter 1.

1. What knowledge sources did the learners use in inferring the meaning of unknown words in the experimental sentences and text?

2. What types of vocabulary knowledge did the learners use in the lexical inferencing process?

6.4.1 Framework for analysis

According to the working definition of depth of vocabulary knowledge vis-à-vis reading comprehension proposed in Chapter 2, and based on some features of Haastrop’s (1991) and de Bot, Paribakht and Wesche’s (1997) taxonomies of knowledge sources for lexical inferencing, the following framework was developed as the analysis tool for the present study.

1. Location of clues:

   a) Clues within the test word (see the analysis of Example 1 in Section 6.4.2);
b) Clues from the immediate context, i.e., the T-unit containing the test word;

c) Clues beyond the immediate context.

2. Knowledge of the world, which includes factual knowledge, attitudes, beliefs and prejudices.

3. Intralingual knowledge:

a) Phonological/orthographic forms. The learner compares phonological or orthographic similarity of a test word and other words in inferring the meaning of the test word.

b) Morphology. The learner uses morphological clues in a test word to infer its meaning. Such clues may include the stem, inflections, derivational affixes, other word formation devices, and the part of speech of the word.

c) Syntax. The learner uses syntactic clues from the T-unit surrounding the test word to infer the meaning of the test word. Such clues may include the test word’s position, and its collocational and other syntagmatic relations with words in the T-unit.

d) Meaning. The learner works on the appropriate meaning of a test word based on the meaning of the written context using, where applicable, clues indicating polysemy, antonymy, synonymy or other paradigmatic relations the word may have. Clues of this type may extend beyond the immediate context of the T-unit.

4. Interlingual knowledge:

a) L1. The learner uses his or her L1 knowledge in inferring the meaning of a test word. The main source of knowledge in this case is cognates.
b) L1. The learner uses the knowledge of a language other than his or her L1 and the target L2 to infer the meaning of a test word.

Some remarks are in order concerning this analytic framework. First, although the present framework looks similar to Haastrup’s (1991) taxonomy in its overall structure, the two differ in the following aspects: (a) the present framework is a two-level structure while Haastrup’s taxonomy has three levels, and (b) the organization and contents of some parts of the present framework differ considerably from those of Haastrup’s taxonomy. For example, although the heading intralingual knowledge was taken from the Haastrup taxonomy, the content (subcategories) of this section is very different from the corresponding section in the Haastrup taxonomy. In the present framework, all the categories under intralingual knowledge are grounded on the working definition of depth of vocabulary knowledge (see Section 2.3.2.3) that was specifically developed for this research in order to represent depth of vocabulary knowledge in the data analysis, whereas Haastrup (1991) did not consider depth of vocabulary knowledge a distinct factor in her data analysis.

Second, register and word frequency are, theoretically, also aspects of depth of vocabulary knowledge, which implies that these two aspects should be included in intralingual knowledge. However, a preliminary analysis of the interview data indicated that the two aspects were in fact not knowledge sources that the participants referred to. In order to keep the framework as simple as possible, these two aspects are therefore excluded.

Third, the concept, T-unit, has been adopted in the present framework. A T-unit is deemed to be either a simple sentence (main clause) without any subordinate clauses attached to it, or in the case of a complex sentence, an independent clause with any number of dependent
clauses associated with it (Hunt, 1965).

Fourth, while it is debatable whether conceptually "co-text" (Haastrup, 1991, p. 92) is a knowledge source, there is an advantage in including a separate category, location of clues, in the present framework. According to Haastrup's (1991) method of counting knowledge sources used in an inferencing attempt, an informant can be deemed to be using more than one knowledge source according to his or her response to a test word. For instance, if the informant used a syntactic clue from the immediate co-text surrounding the test word, he or she was regarded as using two knowledge sources: the co-text and the syntax. This means that co-text was categorized as a knowledge source additional to syntax. Although, as previously pointed out in Chapter 2, the concept of co-text as a knowledge source is questionable, from a practical point of view, this category can provide valuable information indicating how far afield the learner has found clues -- from the test word per se, from words surrounding the test word in the T-unit, or from beyond the T-unit. In the present analysis, therefore, the category of location of clues is included simply to locate clues used in learners' lexical inferencing. However, in the following sections, this category is not regarded conceptually as a knowledge source.

6.4.2 Definition of valid attempt and rate of success in lexical inferencing

The present analysis adopts Haastrup's (1991) principle that an attempt at lexical inferencing is taken as valid "if the item in question is discussed, and the discussion allows for a categorization according to the taxonomy" (p. 105). One valid attempt can include a single or multiple response, that is, it can be associated with the use of one or more knowledge sources. However, the data collection for this study used the form of individual interviews concentrating on
participants' retrospection and introspection about their lexical processing. Presumably, without a partner to work with, there was not much room for participants to carry on detailed and free-flow discussion of any lexical item whose meaning was being guessed in the course of their reading. Therefore, in counting valid attempts, I considered an attempt valid if a participant's responses to the interview questions indicated that he or she was trying to work out the meaning of the target word, and if the response contained sufficient information for categorization according to the analytic framework.

The following case, for example, was deemed a valid attempt at inferring the meaning of "readout" because the response given by the learner to the question on the test word "readout" indicated that he was trying to work out the meaning of the word with clues within the test word ("it sounds like "read-it-out""), clues from the immediate context ("So the word "digital" is important here"), his knowledge of the world ("It's like the display on our digital watches"), and his knowledge of word formation devices. Based on these clues and knowledge sources, he concluded that "readout" should mean "display" here, which was correct (see Example 1). This case was categorized as: 1a (clues within the test word), 1b (clues from the T-unit context), 2 (knowledge of the world) and 3b (morphology).

Example 1:

[T-unit: These portable meters give digital readouts that indicate the difference between a plant's temperature and that of the surrounding air.]

R¹: In the second sentence, you have one unknown word, "readouts".
L: Eh ... it sounds like "read-it-out". I thought it had something to do with digits, therefore I put it into display. It's like the display on our digital watches. Is it right?

¹ In this example and the other examples to appear later in this chapter, "R" stands for the researcher, or the interviewer, and "L" stands for the learner, or the interviewee.
I made use of my life experience to reach this.

R: Yes, you were right ... When you were guessing "readouts", did you consider any other meanings before deciding on "digital display"? Or was it your first reaction?

L: It was my first reaction. Because I've seen a lot of electronic clocks and watches, they are all related to numbers. So the word "digital" is important here.

R: So "digital" gave you some hint?

L: Yes. So it should mean "display".

The following case was not deemed a valid attempt because the learner’s response to the question indicated that he could not, or did not want to, guess the meaning of the test word "indispensable". Under the circumstances, the interview had to move on due to time constraints.

Example 2:

[T-unit: The greenhouse is a common feature in many gardens and is indispensable to the gardener, as it provides near-perfect conditions for plants, both edible and ornamental, all year round, irrespective of the weather.]

R: I’d like to take a look at some other words in the text now. The first word is "indispensable". Can you guess its meaning?

L: No.

R: Okay. Next ... (Transcript #18)

A note on categorization: In order to be categorized as using the knowledge source of meaning (3d), an inferencing case has to show that the learner was trying to use clues covered by the definition of the category in working out the meaning of the test word. A direct jump to the meaning or a wild guess at the meaning of the test word in responding to a question without any indication of a basis for inferencing was not considered a valid case, nor was it classified into the meaning category.

In total, 67 valid attempts were identified according to the definition of valid attempt provided in this section. There were also 47 cases considered invalid attempts because, although the questions on these words were asked, participants were either unable to work on these words,
or the responses to the questions did not contain sufficient information for categorization according to the analytic framework. Table 6-3 lists all the test words used in each individual interview. The 30 underlined words stand for the cases in which lexical inferencing attempts were successful. Based on the 67 valid attempts made by the 12 participants, the overall success rate was 45%.

In lexical inferencing, one attempt may involve a number of knowledge sources. In the foregoing Example 1, in order to infer the meaning of the test word "readout", the learner used morphological clues within the test word, as well as his world knowledge based on contextual clues within the T-unit containing "readout". In this case, two knowledge sources are deemed to have been activated, namely, morphology and knowledge of the world. Among the 67 valid attempts, there were 53 attempts involving only one knowledge source, 11 attempts involving two knowledge sources, and 3 attempts involving three knowledge sources.

### 6.4.3 Interrater reliability

I coded all the interview data. A random draw also selected a subset of two interview protocols to be coded by a second rater, an experienced ESL teacher. These two protocols contained 13 valid attempts, accounting for 19% of the total number of 67 valid lexical inferencing attempts by all the participants. Out of the 13 valid attempts, the two raters agreed on the categorization of knowledge sources for 11 test words. Thus, using the number of valid attempts as the basis for calculation, the interrater agreement on the sample was 85%. At a finer level, for the 34 codes for categorization contained in the two protocols, the two raters disagreed on two of them, which indicated that at the individual coding level the interrater agreement was 94%.
### Table 6-3. List of test words used in the individual interviews

<table>
<thead>
<tr>
<th>Par#</th>
<th>Test words: valid attempts</th>
<th>Test words: invalid attempts</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>contention, indispensable, irrespective(^a), free-standing</td>
<td>readout, contention, indispensable, irrespective, decorative, device, free-standing</td>
</tr>
<tr>
<td>2</td>
<td>potent, ornamental, prone, deteriorate, conduct</td>
<td>emotion, indispensable.</td>
</tr>
<tr>
<td>3</td>
<td>readout, contention, sound, functional, prone, deteriorate, conduct, device, free-standing</td>
<td>ornamental, decorative</td>
</tr>
<tr>
<td>4</td>
<td>potent, permanent, functional</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>potent, prejudice, contention, greenhouse, indispensable, edible, irrespective</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>contention, functional, free-standing</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>contention, ornamental, irrespective, functional, prone</td>
<td>readout, indispensable, decay, deteriorate, conduct, free-standing</td>
</tr>
<tr>
<td>8</td>
<td>potent, free-standing, conduct</td>
<td>ornamental, deteriorate, functional</td>
</tr>
<tr>
<td>9</td>
<td>emotion, indispensable, permanent, decorative, deteriorate, conduct, free-standing</td>
<td>decay, potent, readout, edible, irrespective, ornamental, prone, device</td>
</tr>
<tr>
<td>10</td>
<td>potent, cereal, sound, decorative, prone, conduct, free-standing, deteriorate</td>
<td>decay, meter, readout, contention, ornamental</td>
</tr>
<tr>
<td>11</td>
<td>readout, prejudice, contention, literally, deteriorate</td>
<td>decay, sound, indispensable, edible, irrespective, ornamental, free-standing</td>
</tr>
<tr>
<td>12</td>
<td>potent, portable, digital, contention, rejection, prone, digital, feature</td>
<td>cereal, decay, indispensable, edible, irrespective, ornamental, free-standing</td>
</tr>
</tbody>
</table>

Total: 67 cases  
Total: 47 cases

\(^a\) Underlined words indicate successful attempts.
6.4.4 Sources of knowledge used in lexical inferencing

The sources of knowledge that the 12 participants drew on in lexical inferencing are summarized in Table 6-4. Some findings are worth mentioning here. As far is location of clues is concerned, participants mainly used clues from within the T-unit containing the test word as well as clues from within the test word.

Table 6-4. Use of knowledge sources by all participants (n = 12)

<table>
<thead>
<tr>
<th>Code</th>
<th>Category</th>
<th>No. of times used</th>
<th>Success rate (%)&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Location of clues</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>within the test word</td>
<td>37</td>
<td>30</td>
</tr>
<tr>
<td>b</td>
<td>context within T-unit</td>
<td>40</td>
<td>65</td>
</tr>
<tr>
<td>c</td>
<td>context beyond T-unit</td>
<td>8</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>Knowledge of the world</td>
<td>16</td>
<td>56</td>
</tr>
<tr>
<td>3</td>
<td>Intralingual knowledge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>phono./orthographic forms</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>b</td>
<td>morphology</td>
<td>26</td>
<td>38</td>
</tr>
<tr>
<td>c</td>
<td>syntax</td>
<td>7</td>
<td>57</td>
</tr>
<tr>
<td>d</td>
<td>meaning</td>
<td>30</td>
<td>60</td>
</tr>
<tr>
<td>4</td>
<td>Interlingual knowledge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>L1</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>b</td>
<td>Ln</td>
<td>0</td>
<td>-</td>
</tr>
</tbody>
</table>

<sup>a</sup> Figures under this heading were derived based on the total number of times that a particular knowledge source was activated. For example, among the seven valid attempts involving use of syntax, four were successful. Therefore, the success rate of using the category of syntax was 57%.

Under the category of intralingual knowledge, participants relied most heavily on two knowledge sources, i.e., meaning and morphology. The other two knowledge sources, phonological/orthographic forms and syntax, were minimally employed. Altogether, intralingual knowledge as a knowledge source was activated 68 times. In contrast, world knowledge was
referred to only 16 times, and another main category, interlingual knowledge, did not appear to be activated at all.

6.5 Use of Depth of Word Knowledge in Lexical Inferencing: Some examples

This section analyzes some lexical inferencing cases selected from the interview transcripts to reflect approaches to inferencing that participants adopted. In accordance with the working definition of depth of vocabulary knowledge for the present research, all the four categories under the heading intralingual knowledge in the present analytic framework are regarded as aspects of depth of vocabulary knowledge. Therefore, cases that involved use of these four knowledge sources are deemed to be inferencing attempts using depth of vocabulary knowledge. In total, 57 cases met this criterion.

An interesting phenomenon was that learners used different approaches to infer the meaning of the same test word. The six examples provided below all involve the test word "free-standing". In these cases, learners adopted different approaches to the inferencing task such as: (a) resorting to word formation knowledge using clues within the test word; (b) identifying word meaning in context using contextual clues, especially meaning of the context; (c) using syntactic cues within the T-unit to determine the meaning relations between the test word and other words; and (d) a combination of the above. However, whichever approach was used, it was often critical to have a fairly good understanding of the meaning or grammar of the context, particularly the T-unit in question, so that contextual clues could be effectively used in lexical inferencing. Having a superior knowledge of morphology was also helpful. Occasionally, morphological knowledge single-handedly helped to solve a problem, as happened in some of
the cases below (Examples 5, 6, and 8), when clues outside the test word could not be found or utilized by the learner. An interpretation of the inferencing attempt follows each episode\(^2\).

More examples are provided in Appendix F.

[T-unit: aluminum requires special fixing devices or free-standing units.]

**Example 3**

L: ... Free-standing units, I don't know what it means. I didn't know, but I can guess from here. Free-standing units ... it can stand by itself.
R: Un huh, very good. So how did you guess this word? What's the basis for your guess, because you guessed it correctly.
L: Oh. Because, see, "requires special fixing devices, or free-standing units". You can stand up by itself or you can put it together with special fixing devices. That's about it.
R: So you just used this part "special fixing device" as the information source to guess the second part?
L: Yeah. Because it says "or". Gives you contrast meaning, maybe. That's my guess.
R: So you didn't guess from the word formation itself.
L: No, no. (Transcript #48)

In the above attempt, the learner used syntactic clues from the context of the T-unit. Based on the contextual clues, she concluded that she should look for some meaning contrast to "special fixing device" because the phrase and "free-standing units" were connected by "or". She therefore came up with the meaning "it can stand by itself". Her attempt, categorized as 1b (clues within the T-unit context), 3c (syntax) and 3d (meaning), was successful.

**Example 4**

R: So last word "free-standing" here. You know this word?
L: I don't know.
R: Oh, you don't know this one?
L: But I can guess.
R: Okay. Give me your guess, please.

\(^2\) All the episodes provided here were originally spoken in English.
L: Aluminum is pretty flexible. So we can make a shape whatever we want. That means "free-standing".
R: Un huh.
L: Or movable.
R: So could you give me sort of the definition of this word "free-standing"? So what kind of thing can be called free-standing?
L: ... (murmuring)
R: You still haven't got a clear idea what "free-standing" means here.
L: Ah! There's stick, there are many sticks. They can, like, they can get together, and they can make a shape.
R: Make a shape. So...
L: Yeah, 'cause to make one unit, to make one shape, certain shape, they need a special type of ... They need a number ... they need enough sticks.
R: Okay.
L: So they grows the unit that is called the free-standing unit.
R: Okay I see what you mean. (Transcript #45)

This was a failed attempt although the learner tried to use her world knowledge of usual ways of people comparing advantages and disadvantages of two things. It appears that the learner's mistaking "free-standing" for "making a shape whatever we like" because aluminum was flexible and movable was likely due to her oversight of the function of the coordinator "or" that connects "free-standing units" and "special fixing devices" because her previous attempt in inferring the meaning of "devices" was successful. In that previous attempt she demonstrated she understood the first part of the T-unit well. This attempt was categorized as: 1b (clues within the T-unit context) and 2 (knowledge of the world).

Example 5

R: Okay. Can you guess the meaning of "free-standing" here?
L: Device or free-standing unit ... it means some kind of machine or ... I know free, I know standing ... It means can stand without other help.
R: So free-standing means it can stand free without being supported, can stand by itself?
L: Yes. (Transcript #73)

This learner relied on clues within the test word to work out the meaning of free-standing. It appears that she used her knowledge of word-compounding to reach the conclusion
that free-standing means standing by itself. This attempt was categorized as: 1a (clues within the test word) and 3b (morphology).

**Example 6**

L: Free-standing ... free-standing ... It can be stand just by alone.
R: Can stand alone? Good. How did you work out the meaning? What things here lead you to the guess?
L: I can guess. Free-standing ...
R: You made the right guess. How did you work it out?
L: Free ... standing.
R: Free and standing, just based on the word itself? So that means without support?
L: You're right. (Transcript #35)

This learner, too, used word formation clues within the test word to reach the correct meaning of "free-standing". The attempt was categorized as: 1a (clues within the test word) and 3b (morphology).

**Example 7**

R: Now what about free-standing here? It cannot be a bus-station here?
L: Two words (referring to devices and free-standing units), some kind of instrument. One is, one is more big. 'Cause screw is a ... screw made these. There's some kind of connecting.
R: So which one is bigger? Free-standing units or device?
L: Of course devices, because screw is big. So I think device is bigger.
R: You mean free-standing unit is a small device?
L: Device is big and free-standing is small.
R: Free-standing is small? It's like an instrument, you mean?
L: Yeah. I think devices is large instrument. (Transcript #64)

This learner did not succeed in the attempt probably because a) she did not comprehend the T-unit well and b) she ignored the coordinator "or" which suggests some contrast in meanings of the two phrases this "or" coordinates. Therefore, her inferencing was not in the right direction to begin with. This failed attempt was categorized as: 1b (clues within the T-unit context), 1c (clues beyond the T-unit context), 2 (knowledge of the world) and 3d (meaning).
Example 8

L: Free-standing is like ... device ... like something like option ...
R: Option?
L: Make some special ... material ... call special fixing. But the kind of special fixing, yeah I think, I don’t know exactly what ... I can guess free-standing, just free, free-standing, don’t need another fixing.
R: So free-standing means it can stand free, doesn’t need another fixing?
L: Yeah, another helping.
R: Okay. Like another support?
L: Yeah support. (Transcript #71)

This learner started with the meaning of the first part of the T-unit and tried to use it as a clue to infer the meaning of free-standing to no avail. Failing that, he simply concentrated on the test word itself and activated his knowledge of word compounding to obtain the correct meaning of the test word. This attempt was categorized as: la (clues within the test word), lb (clues within the T-unit) and 3b (morphology).

6.6 Comparing Learners with Different Depths of Vocabulary Knowledge

The purpose of this comparison was to determine whether, in inferring meanings of unknown words in the reading process, those with greater depth of vocabulary knowledge differed from those with less depth of vocabulary knowledge in their rates of inferencing success and in their patterns of drawing on knowledge sources. I divided the 12 learners into three proficiency groups, high, intermediate, and low, according to the ranks described in Section 6.3. The ranking was based on the 12 participants’ scores on a depth-of-vocabulary- knowledge test, the DVK. Each proficiency group consists of four participants. That is, learners occupying the first through the fourth ranks were designated as high proficiency learners (HP), those occupying the fifth through eighth ranks as intermediate proficiency learners (IP), and those occupying the ninth
through twelfth places as low proficiency learners (LP). Because the grouping and ranking were based on their scores on the DVK, the HPs were thus deemed learners with greater depth of vocabulary knowledge, and the LPs learners with less depth of vocabulary knowledge.

6.6.1 Rates of success in lexical inferencing by the HPs, IPs, and LPs

In the individual interviews, the HP group made 21 valid lexical inferencing attempts, among which 14 produced successful guesses. The success rate of the HP group was, therefore, 67%. The IP group made 18 valid lexical inferencing attempts, among which eight were successful. The success rate of the IP group was, therefore, 44%. In comparison, the LP group made 28 valid attempts at lexical inferencing, among which eight were successful. The success rate of the LP group was, therefore, 29%. Based on these figures, it can be seen that the success rate for the HP group was 2.3 times as high as that for the LP group, and that the success rate for the IP group was very close to the average success rate of 45% for the 12 participants collectively.

6.6.2 Knowledge sources used in lexical inferencing by the HPs and LPs

Tables 6-5 and 6-6 compare the use of knowledge sources by the HP and LP groups according to the framework for data analysis. The category of interlingual knowledge is not included in this comparison since it was in fact not a knowledge source the learners referred to in the inferencing process. The total number of valid attempts made by each group (21 for the HPs and 28 for the LPs) was used as the basis for deriving percentages so that comparisons could be made. A comparison of Tables 6-5 and 6-6 leads to the following findings.

1. Proportionally, the HPs made more use of contextual clues (86%) than the LPs (68%),
but the LPs made more use of clues within test words (64%) than the HPs (43%).

2. Concerning the use of intralingual knowledge, the HPs mainly concentrated on two knowledge sources, i.e., meaning and morphology, while the use of the other two knowledge sources under the same heading, i.e., syntax and phonological/orthographic forms, was minimal. In contrast, although meaning and morphology were also the two most frequently used knowledge sources for the LP group, the LPs employed the four knowledge sources more evenly. Between the two most heavily used knowledge sources, meaning and morphology, the LPs relied more on morphology (39%) than on meaning (32%), while the HPs relied much more on meaning (67%) than morphology (24%).

Table 6-5. Use of knowledge sources by the HP group (Participants #1-#4)

<table>
<thead>
<tr>
<th>Category</th>
<th>No. of times used</th>
<th>% of 21 VA'sa</th>
<th>Success rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location of clues</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>within the test word</td>
<td>9</td>
<td>43</td>
<td>44</td>
</tr>
<tr>
<td>context within T-unit</td>
<td>16</td>
<td>76</td>
<td>81</td>
</tr>
<tr>
<td>context beyond T-unit</td>
<td>3</td>
<td>10</td>
<td>67</td>
</tr>
<tr>
<td>Knowledge of the world</td>
<td>6</td>
<td>29</td>
<td>67</td>
</tr>
<tr>
<td>Intralingual knowledge</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>phonological/orthographic forms</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>morphology</td>
<td>7</td>
<td>24</td>
<td>57</td>
</tr>
<tr>
<td>syntax</td>
<td>2</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>meaning</td>
<td>14</td>
<td>67</td>
<td>79</td>
</tr>
</tbody>
</table>

a. As explained in Section 6.4.2, since participants may use more than one knowledge source in a lexical inferencing attempt, it is possible for the percentages of use of some categories to exceed 100%.
Table 6-6. Use of knowledge sources by the LP group (Participants #9-#12)

<table>
<thead>
<tr>
<th>Category</th>
<th>No. of times used</th>
<th>% of 28 VA's</th>
<th>Success rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location of clues</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>within the test word</td>
<td>18</td>
<td>64</td>
<td>22</td>
</tr>
<tr>
<td>context within T-unit</td>
<td>14</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>context beyond T-unit</td>
<td>5</td>
<td>18</td>
<td>40</td>
</tr>
<tr>
<td>Knowledge of the world</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>25</td>
<td>43</td>
</tr>
<tr>
<td>Intralingual knowledge</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>phonological/orthographic forms</td>
<td>5</td>
<td>18</td>
<td>0</td>
</tr>
<tr>
<td>morphology</td>
<td>11</td>
<td>39</td>
<td>36</td>
</tr>
<tr>
<td>syntax</td>
<td>4</td>
<td>14</td>
<td>50</td>
</tr>
<tr>
<td>meaning</td>
<td>9</td>
<td>32</td>
<td>33</td>
</tr>
</tbody>
</table>

3. Comparing uses of the meaning category by HPs and LPs, it can be seen that although meaning was one of the most heavily employed categories by both groups, the proportion of its use by the HPs (67%) was much higher than that by the LPs (32%). On the other hand, the LPs’ proportional use of morphology was more frequent (39%) than the HPs’ was (24%).

4. In addition, on 18% of the inferencing occasions, the LP group used phonological/orthographic forms as a knowledge source while the HP group did not use this source at all.

5. In individual comparisons of the success rates for each of the knowledge sources used in the inferencing process, HPs achieved a much higher success rate than the LPs in every single category. This difference is especially striking in two categories, syntax and meaning, where the HPs’ success rates were at least twice as high as those of the LPs. It is also noteworthy that although the category of phonological/orthographic forms was
used as a knowledge source in 18% of the LPs' inferencing attempts, none of their attempts to use this category of clues turned out to be successful. The success rate for the LPs using morphological clues was also low (36%), relative to that achieved by the HPs (57%).

6.7 Analyses of the Data from the Questionnaire on Reading Strategies

The Questionnaire on Reading Strategies, designed as a small-scale survey on ESL learners' self-perceived behaviours and strategies in dealing with unknown words in English texts, was given to the 12 participants at the end of their individual interviews. It was arranged thus in order to avoid the possible influence of the questions in the questionnaire on participants' behaviours in performing the lexical inferencing tasks. The participants were asked to complete the questionnaire on the spot. All the participants willingly completed the questionnaire as requested. The questionnaire asked two main questions:

1. How often do you do each of the following when you meet an unfamiliar word in reading an English text?

2. What kind of information do you use when trying to guess the meaning of an unfamiliar word in an English text?

Nine options were given under the first question, and six options were given under the second question. The details of all these options can be found in Tables 6-7, 6-8, 6-9 and 6-10. A scale of four frequency levels was provided alongside each option so that for each option, which indicated a behaviour or strategy, the respondent could pick one frequency level that most suitably reflected his or her situation.
6.7.1 How did learners approach unknown words?

The purpose of the first question (from Harley & Hart, in preparation) was to identify the participants' perceptions of their most frequent behaviours when they encountered unknown words. Table 6-7 summarizes the questionnaire responses from the 12 participants. As Table 6-7 shows, guessing meaning from context was reported as the most frequent and popular behaviour among the respondents. About 67% of the sample indicated that they often guessed word meaning from the context. Another 25% stated that they sometimes guessed the meaning of an unknown word from the context. When these two groups are combined, 92% of the sample reported using "guessing from the context" as a primary means in dealing with unknown words in reading an English text.

The participants also reported looking up an unknown word in a unilingual English dictionary as a frequent behaviour. Half of the sample (50%) noted that they would often consult an English-only dictionary when they encountered an unknown word. Another 33% indicated that they sometimes used an English-only dictionary to get the meaning of an unknown word. When these two figures are added, 83% of the sample were in favour of using an English unilingual dictionary to obtain the meaning of an unknown English word. In comparison, only 33% of the sample acknowledged that they often sought help from a bilingual English-Korean/Chinese dictionary for the meaning of an unknown word, and 42% indicated they sometimes consulted a bilingual dictionary for a word meaning. In the aggregate, 75% of the sample were in favour of using a bilingual English dictionary to get a word meaning at least sometimes.
Table 6-7. Frequency of learners’ self-reported behaviours in dealing with unknown words while reading

<table>
<thead>
<tr>
<th>Behaviour</th>
<th>Often</th>
<th>Sometimes</th>
<th>Rarely</th>
<th>Never</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Look up the word in an English-Korean/Chinese dictionary</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>b) Look it up in an English-only dictionary</td>
<td>6</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>c) Guess its meaning from the context</td>
<td>8</td>
<td>3</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>d) Ignore it</td>
<td>1</td>
<td>7</td>
<td>4</td>
<td>-</td>
</tr>
<tr>
<td>e) Ask the teacher for assistance</td>
<td>1</td>
<td>3</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>f) Ask a friend if they know it</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>g) Look for clues to meaning in the word itself</td>
<td>3</td>
<td>7</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>h) Make a note of the word (i.e. write it down)</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>i) Other</td>
<td>-</td>
<td>1*</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

* One learner stated that he compares the spelling of the word with other words he knows.

Making a note of unknown words while reading English texts was also a popular behaviour among the participants. About 42% of the sample reported often doing this, and another 42% acknowledged sometimes doing this. Only one respondent (8%) reported rarely doing this, and one other reported never doing this. Therefore, in aggregate, 83% of the sample reported relatively frequently writing down unknown words while reading English texts.

Another noteworthy finding was the proportion of respondents indicating that they looked for the meaning of an unknown word from clues within the word. About 58% of the sample stated that they sometimes did this, and 25% indicated that they often did this. Thus, in the aggregate, 83% of the sample reported sometimes or frequently referring to clues within the
word itself when they met a new English word.

About 58% of the sample acknowledged that they would sometimes ignore an unknown English word in a reading process. One respondent, accounting for 8% of the sample, confided that, when reading an English text, she often ignored words that were unknown to her. This respondent ranked last in three (the RC, VS, and DVK) of the five test score sets in Table 6-2, and ranked 11th in the EVKS and the 7th in the MK among the 12 learners.

In comparison, the other two strategies, i.e., asking the teacher for assistance and asking a friend for assistance, were infrequently employed. Only one learner (8%) indicated she often asked the teacher for assistance when encountering an unknown word. Three other learners, or 25%, indicated they sometimes did so. The remaining 67% stated that they rarely or never asked for the teacher's help in dealing with a new word. As for the strategy of getting a friend's help in understanding an unfamiliar word, one learner (8%) indicated that she often did so, and another 33% of the sample reported that they sometimes would resort to this strategy. However, the remaining 58% of the sample indicated that they either rarely or never tried to get help from their friends in solving the problem. These two strategies, therefore, were reportedly the most infrequently used ones. When a scale of four points is assigned to the four frequency categories (often=4, sometimes=3, rarely=2, and never=1), a mean ranking of self-reported behaviours emerges as shown in Table 6-8.

Table 6-8 shows that when the 12 learners encountered an unknown word in an English text, they would most likely try to work out the meaning of the word by guessing from the context; their next most likely strategy was to check out the meaning of the word in a unilingual English dictionary, and the least likely was to ask a friend or a teacher for a solution.
Table 6-8. Ranking of frequency of learners’ self-reported behaviours in dealing with unknown words while reading (n = 12)

<table>
<thead>
<tr>
<th>Rank</th>
<th>Behaviour</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>c) Guess its meaning from the context</td>
</tr>
<tr>
<td>2</td>
<td>b) Look it up in an English-only dictionary</td>
</tr>
<tr>
<td>3</td>
<td>h) Make a note of the word</td>
</tr>
<tr>
<td>4.5</td>
<td>g) Look for clues to meaning in the word itself</td>
</tr>
<tr>
<td>4.5</td>
<td>a) Look up the word in an English-Korean/Chinese dictionary</td>
</tr>
<tr>
<td>6</td>
<td>d) Ignore it</td>
</tr>
<tr>
<td>7.5</td>
<td>f) Ask a friend if they know it</td>
</tr>
<tr>
<td>7.5</td>
<td>e) Ask the teacher for assistance</td>
</tr>
</tbody>
</table>

6.7.2 What knowledge sources did learners use in lexical inferencing while reading?

Tables 6-9 and 6-10 summarize the strategies that the participants identified in the second part of the questionnaire, in which the strategies were initially arranged in a random order. In Table 6-9, these strategies have been re-ordered by starting the list with bottom-up, or local processing, approaches and finishing the list with top-down, or global processing, approaches, in order to present a clearer picture.

Table 6-9 shows that the majority of the participants reported using all six strategies fairly frequently. What is most noticeable is their reported frequent use of the two global processing strategies (e and f). About 83% of the sample stated that they often made use of the meaning of the paragraph or text as a whole to guess the meaning of an unknown word (strategy e), and 17% sometimes opted for the same strategy. In a similar pattern, 75% of the sample indicated that they often used their background knowledge of the topic of the text to guess the meaning of an unknown word (strategy f), and 25% reported sometimes using the same strategy.
This suggests that all 12 participants used these two strategies fairly frequently.

Table 6-9. Learners’ self-reported strategies for inferring the meaning of unknown words in reading

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Number of learners</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Often</td>
</tr>
<tr>
<td>a) I examine the unknown word to see if it contains any grammatical clues to tell me what part of speech it belongs to</td>
<td>1</td>
</tr>
<tr>
<td>b) I examine the unknown word to see if any part of it is familiar in meaning</td>
<td>4</td>
</tr>
<tr>
<td>c) I look for grammatical clues in the surrounding sentence to help me guess the meaning of the unknown word</td>
<td>1</td>
</tr>
<tr>
<td>d) I use the meaning of other words in the same sentence to help me guess the meaning of the unknown word</td>
<td>6</td>
</tr>
<tr>
<td>e) I make use of the meaning of the paragraph or text as a whole to guess the meaning of the unknown word</td>
<td>10</td>
</tr>
<tr>
<td>f) I use my background knowledge of the topic of the text to guess the meaning of the unknown word</td>
<td>9</td>
</tr>
</tbody>
</table>

Strategies (d) and (b) were also considered useful by the majority of the participants. Half of the sample reported *often* using strategy (d) to work out the meaning of an unknown word, and one learner (8% of the sample) *sometimes* opted for this means. However, the other 42% of the sample stated that they *rarely* used this strategy. In comparison, 33% of the sample indicated that they *often* used strategy (b), and 50% reported *sometimes* resorting to this strategy.

In contrast, looking for grammatical clues, either at the word level or sentence level,
appeared to be regarded as a secondary strategy, as suggested by the responses to items (a) and (c). For strategy (a), only one participant reported that he often used this strategy. About 67% of the sample noted that they *sometimes* used this strategy, and the other 25% indicated that they *rarely* used this strategy. For item (c), only one respondent reported that she often used this strategy, 75% of the sample stated that they *sometimes* used this strategy, and 17% indicated that they *rarely* used this strategy.

When a scale of four points is assigned to the four frequency categories (often=4, sometimes=3, rarely=2, and never=1), a mean ranking of the learners' perceived strategies for inferencing the meaning of unknown words in English texts emerges as follows (Table 6-10):

The two global processing strategies (e) and (f) appear as the two top ranks among the six strategies compared, whereas strategies (c) and (a), which make use of grammatical clues to help infer the meaning of unknown words, occupy the two bottom ranks in the table.

**Table 6-10.** Ranking of learners' perceived strategies for inferring the meaning of unknown words in reading (n = 12)

<table>
<thead>
<tr>
<th>Rank</th>
<th>Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>e) I make use of the meaning of the paragraph or text as a whole to guess the meaning of the unknown word</td>
</tr>
<tr>
<td>2</td>
<td>f) I use my background knowledge of the topic of the text to guess the meaning of the unknown word</td>
</tr>
<tr>
<td>3</td>
<td>b) I examine the unknown word to see if any part of it is familiar in meaning</td>
</tr>
<tr>
<td>4</td>
<td>d) I use the meaning of other words in the same sentence to help me guess the meaning of the unknown word</td>
</tr>
<tr>
<td>5</td>
<td>c) I look for grammatical clues in the surrounding sentence to help me guess the meaning of the unknown word</td>
</tr>
<tr>
<td>6</td>
<td>a) I examine the unknown word to see if it contains any grammatical clues to tell me what part of speech it belongs to</td>
</tr>
</tbody>
</table>
CHAPTER SEVEN
THE FOLLOW-UP STUDY: INTERPRETATION OF FINDINGS

7.1 Overview
This chapter discusses the findings of the follow-up study described in Chapter 6 and provides an answer to the third research question posed for this research: How do ESL learners make use of their depth of vocabulary knowledge in the process of comprehending a general academic text?

7.2 Interpreting the Results of the Interview Study
This section is organized in general accordance with the framework for the qualitative analysis, the details of which can be found in Section 6.4.1. Because grouping of the learners was based on their relative scores on the DVK, the high proficiency (HP) and low proficiency (LP) ESL learners are deemed to represent different depths of vocabulary knowledge.

7.2.1 Success rates
Collectively, the success rate of lexical inferencing for the 12 participants was 45%. By depth-of-vocabulary-knowledge groups, the success rate for the HPs was 67%, and that for the LPs was 29%. In other words, the success rate for the HP group was 2.3 times as high as that for the LP group. Other researchers (e.g., dos Santos & Sanpedro Ramos, 1993; Haastrup, 1991; Morrison, 1996) obtained similar results in their studies when they divided their learners according to L2 proficiency levels. The basis for grouping learners in this study differs from that used in these other studies; in the present study depth of vocabulary knowledge was used as the
basis for division, whereas the others divided their groups according to general L2 proficiency levels. Despite this difference, the various findings are still ostensibly in agreement because vocabulary knowledge is extensively implicated in practical language skills (Meara & Jones, 1988).

The striking disparity in success rates of learners with different depths of vocabulary knowledge in this case prompts one to argue that there is a positive relationship between learners' depth of vocabulary knowledge and their ability to succeed in inferring meanings of unknown words in reading. The greater the depth of their vocabulary knowledge, the more likely that learners will succeed in inferring the meaning of additional vocabulary when reading English texts. In describing the Matthew effect on reading comprehension, Stanovich (1986) argues for a recognition of reciprocal causation in the reading comprehension process. He contends that vocabulary promotes reading comprehension, which in turn promotes vocabulary knowledge. Therefore, a learner with better vocabulary knowledge will achieve better reading comprehension, which in turn will help the learner acquire more vocabulary knowledge. Because lexical inferencing is an important part of the reading comprehension process, the present finding strongly suggests that the Matthew, or rich-get-richer, effect in L1 reading (Stanovich, 1986) is also applicable to lexical inferencing in L2 reading.

7.2.2 Location of clues

Results of the data analyses show that both clues within test words and from surrounding contexts were important and helpful in performing lexical inferencing tasks. On the one hand, clues within test words were frequently used. On the other hand, contextual clues were also used
extensively. This indicates the importance of contextual clues (Categories 1b and 1c in Table 6-4) because 72% of the 67 valid lexical inferencing attempts were made with the help of various types of contextual clues. This finding conforms to Haastrup’s (1991) and Morrison’s (1996) findings that the co-text, which is equivalent to "contextual clues" in the present study, was the most extensively used source of clues by learners at both the HP and LP levels.

However, clues were not used in the same manner by all the learners in the present study. The HP learners used contextual clues on 86% of their inferencing occasions. This percentage decreases to 68 for the LP learners. On the other hand, the LPs used test-word clues on 64% of their inferencing occasions. This percentage decreases to 43% for the HPs. Nevertheless, it is evident (see Tables 6-4, 6-5, and 6-6) that, for both the HPs and LPs, the success rates of lexical inferencing based on contextual clues were much higher than those utilizing only clues within test words.

These findings appear to indicate that there is a positive relationship between learners’ depth of vocabulary knowledge and their ability to notice and use different types of contextual clues in inferring meanings of unknown words in reading. The greater the depth of vocabulary knowledge, the better the learner can make use of the context. This better ability to utilize contextual clues, in turn, can result in a higher rate of success in lexical inferencing. At the same time, these findings also suggest that LP learners generally concentrate as much on clues within the target word as on those embedded in the context (see also Section 7.2.4).

Laufer (1992a, 1996) argues that, to attain a minimum acceptable level of reading comprehension (56% in Laufer’s studies), a learner needs to have a threshold of 3,000 word families. Since scores on breadth and depth of vocabulary knowledge were found to be
substantially correlated in the present research (see Chapters 4 and 5). A threshold of depth of vocabulary knowledge may also exist for lexical inferencing and reading comprehension. Once over this hypothetical threshold level, the learner should be able to make more effective use of contextual clues as well as clues found within unknown words in an English text.

7.2.3 Use of main knowledge sources

In general, two main knowledge sources were found to have been drawn upon as participants performed their lexical inferencing tasks. In the 67 valid inferencing attempts, *intralingual knowledge*, which conceptually represents depth of knowledge, was activated 68 times. In comparison, the only other knowledge source that was used, *knowledge of the world*, was activated only 16 times. This finding suggests that depth of vocabulary knowledge was an important knowledge source in facilitating learners to succeed in their lexical inferencing tasks.

7.2.4 Use of sources representing depth of vocabulary knowledge

Four categories of knowledge source, phonological/orthographic forms, morphology, syntax, and meaning, were established under the heading of *intralingual knowledge* to represent depth of knowledge of test words. The analysis revealed that for the 12 learners collectively, meaning was the most heavily used knowledge source, activated 30 times in the 67 valid attempts. The second most-heavily used source was morphology, which was activated 26 times. The use of two other knowledge sources, phonological/orthographic forms and syntax, was lower.

In terms of the rank order of frequency, the pattern of using these four intralingual knowledge sources by the HP group conformed in general to the collective pattern described
above (see Table 6-5). However, LP learners appear to have used these knowledge sources in a different way: they relied fairly heavily on morphology, and meaning was the second most-used knowledge source for this group (see Table 6-6).

These findings suggest that the HPs’ strategies of lexical inferencing were more top-down than those of the LPs, who on most occasions employed bottom-up approaches to lexical inferencing. The findings also suggest that, while learners with greater depth of vocabulary knowledge have the ability to frequently resort to meaning as a major knowledge source in the lexical inferencing process, learners with less depth of vocabulary knowledge are often reduced to looking at the form of the target word itself. This is perhaps because the ability to use meaning as a knowledge source often requires knowing relatively well the meaning of other words surrounding the target word and learners with less depth of vocabulary knowledge are often "hampered by a lack of knowledge of other words in the context of a target word" (Harley, 1996, p. 5). This finding is in accord with Cziko’s (1980) report that low proficiency learners attend more to graphemic form and less to contextual meaning, but it disagrees with the claim by dos Santos and Sanpedro Ramos (1993) that LPs tend to over-use top-down strategies. Findings from recent L1 research (Stanovich, 1990; Stanovich, Cunningham, & Feeman, 1984) explains why the HPs’ approaches to lexical inferencing seem more top-down than the LPs’ do: skilled readers rely as much on graphemic cues as less skilled readers, but because skilled readers use up less cognitive capacity than less skilled readers in doing so, they still have the capacity to attend to other types of cue simultaneously.

The present finding that LPs relied heavily on the form of test words in meaning inferencing corroborates Haynes’ (1993) findings: In the absence of adequate contextual clues
in lexical guessing, adult ESL learners have a tendency to analyze the form of unknown words, matching their graphemic units to words in their lexical memory. However, for LPs, these strategies, as Haynes (1993) pointed out, often lead to mismatches and therefore failures. As shown in the present study (see Table 6-6), none of the five valid attempts involving only graphemic matching in the category of phonological/orthographic forms was successful. These five attempts were all made by LPs. The success rate for the attempts involving word form analysis under the heading of morphology was only 32% (6 successful cases out of 19 valid attempts that involve various types of word-form analysis), well below the overall average rate of 45% for the 12 participants. This further suggests that a threshold of depth of vocabulary knowledge may exist for lexical inferencing (see also Section 7.2.2), and that unless the learner has reached this threshold level, his or her strategies for lexical guessing will largely be bottom-up ones which will result in few successes.

7.2.5 Use of interlingual knowledge in lexical inferencing

Throughout the analysis of the interview data, there was no evidence that knowledge of learners’ L1 or L2 was at work, presumably due to the fact that neither of the participants’ L1s, Korean or Chinese, is cognate with English. In contrast, both Haastrup (1991) and Morrison (1996) reported some use of interlingual knowledge by their informants. This was perhaps due to the fact that the L1s of the informants in Haastrup’s (1991) and Morrison’s (1996) research (Danish and English, respectively) were cognate to some extent with the target language of their respective studies (English and French, respectively).
7.3 Relationship between Participants' Scores on the DVK and RC

In this follow-up study, the 12 participants were divided into three groups according to their different depths of vocabulary knowledge based on their scores on the DVK. The analyses based on this grouping led to the argument that different results from the different groups were due to their differences in depth of vocabulary knowledge.

As an additional exploration, a second grouping was attempted based on the 12 participants' scores on the RC. This grouping was made in order to find out how different the results of the two groupings would be and, therefore, to determine whether or not, with this 12-person sample, scores on the DVK still represented levels of reading comprehension well. The results of the grouping according to the two sets of scores turned out to be very similar. The original top third participants, who formed the HP group, remained in the same group when scores on the RC were used. The only difference was a switch in grouping for two participants, Participants #5 and #10. All the other 10 participants remained in their original depth-of-vocabulary-knowledge groups. This fact suggests that scores on the DVK, representing some paradigmatic and syntagmatic aspects of depth of vocabulary knowledge, have a positive and close relationship with levels of reading comprehension. At least for these 12 people, the higher one's score on the DVK, the more likely he or she will achieve a higher score on the RC.

Meara and Jones (1988) were actually referring to vocabulary size when they noted the heavy implication of vocabulary knowledge in language skills. Findings of this study have now provided reasonable grounds to state that depth of vocabulary knowledge is positively and closely associated with reading comprehension.
7.4 Learners' Self-Reported Behaviours in Dealing with Unknown Words While Reading

The first question in the questionnaire on reading strategies asked: "How often do you do each of the following when you meet an unfamiliar word in reading an English text?". Eight ways of dealing with a new word were provided for the participants. The results show that "guessing the meaning of the word from the context" was stated to be the most popular behaviour among the 12 respondents. Meanwhile, it was also found that consulting dictionaries was a popular behaviour where this was allowed, and that respondents mostly preferred a unilingual English dictionary to an English-Korean/Chinese one. These results suggest: (a) that lexical guessing from the context is a useful process, whose importance has not only been identified by researchers but also accepted by L2 learners themselves, as indicated by the results of this study, and (b) that, as might be expected, dictionary use has an important place in dealing with unknown words.

7.5 How Do the Survey Results Relate to the Findings from the Interview Study?

In order to determine whether any relationship exists between learners' self-perceptions of dealing with unknown words and their actual behaviours in lexical inferencing, a comparison was made of the interview data and their responses to the second survey question. The comparison identified some discrepancies between the two parts of the data.

The second survey question asked: "What kind of information do you use when trying to guess the meaning of an unfamiliar word in an English text?" Six pre-written answers were provided for the respondents to choose from (see Tables 6-9 and 6-10). These six options
correspond approximately to the knowledge sources specified in the analytic framework in the following way, after some original categories of knowledge sources were broken down and re-organized in order to match the six questionnaire options (see Table 7-1).

**Table 7-1. Questionnaire options and their corresponding knowledge sources**

<table>
<thead>
<tr>
<th>Strategy (Questionnaire option)</th>
<th>Approximate correspondence to use of knowledge source</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) I examine the unknown word to see if it contains any grammatical clues to tell me what part of speech it belongs to</td>
<td>Morphology (part of speech)</td>
</tr>
<tr>
<td>b) I examine the unknown word to see if any part of it is familiar in meaning</td>
<td>Morphology (stem, affixation, &amp; compounding)</td>
</tr>
<tr>
<td>c) I look for grammatical clues in the surrounding sentence to help me guess the meaning of the unknown word</td>
<td>Syntax (T-unit)</td>
</tr>
<tr>
<td>d) I use the meaning of other words in the same sentence to help me guess the meaning of the unknown word</td>
<td>Meaning (clues within T-unit)</td>
</tr>
<tr>
<td>e) I make use of the meaning of the paragraph or text as a whole to guess the meaning of the unknown word</td>
<td>Meaning (clues from global context)</td>
</tr>
<tr>
<td>f) I use my background knowledge of the topic of the text to guess the meaning of the unknown word</td>
<td>Knowledge of the world</td>
</tr>
</tbody>
</table>

The frequency ranking of the six questionnaire options was in the following descending order: e, f, b, d, c, and a. This indicates that using the meaning of the global context to process the meaning of an unknown word was what the respondents considered their most frequent behaviour, and using knowledge of the world their second most frequent. In comparison with
their original frequency ranks, ranking of these two categories according to the respondents' actual frequencies of use appeared to be quite different from the frequency ranking of their perceptions as to what information they often used in inferring lexical meanings when reading English texts. To provide a comparison, Table 7-2 presents the frequency ranking (in a descending order) of the six options for the second question in the original questionnaire, together with their approximate corresponding categories of knowledge source used for the interview data analysis, number of times these knowledge sources were activated in the 67 valid attempts, and their new ranking according to the frequencies of activation of these knowledge sources.

Table 7-2. Lexical meaning inferencing: Comparing what the learners perceived with what they actually did (n = 12)

<table>
<thead>
<tr>
<th>Ranking of learner perceptions</th>
<th>Que. op. a</th>
<th>Approx. corresponding knowledge source(s) (KS)</th>
<th>Frequency of use of the KS</th>
<th>Ranking: Actual use b</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>e</td>
<td>Meaning (using clues from global context)</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>f</td>
<td>Knowledge of the world</td>
<td>16</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>b</td>
<td>Morphology (stem, affixation &amp; compounding)</td>
<td>21</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>d</td>
<td>Meaning (using clues within T-unit)</td>
<td>23</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>c</td>
<td>Syntax</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>a</td>
<td>Morphology (part of speech)</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

a. Que. op. = Reading Strategy Questionnaire option.
b. Ranking: Actual use = Ranking of the frequencies of the actual use of the six corresponding categories of knowledge sources
It may be observed from Table 7-2 that although the bottom rank remains the same for the self-reports and the actual uses, the top five categories now rank differently: The top category (option e) for the self-reported ranking has now become the fifth in actual use, and the fourth (option d) in the self-reported ranking has now ascended to the top.

Thus although the respondents believed that lexical meaning processing using clues beyond the immediate sentence context and using world knowledge to infer lexical meaning were the two strategies they assumed they most frequently used, what they actually did revealed a different picture, as shown by the actual frequencies of their use of corresponding knowledge sources. The results suggest that the respondents' actual approaches to ESL academic reading comprehension may not be as top-down as they themselves have perceived, and that they, in fact, rely substantially on the immediate context and the test word itself in inferring lexical meanings because, as indicated in Table 6-4, in 37 of the 67 valid attempts, they referred to clues within the test words for an indication of their meaning. In addition, among the six knowledge sources approximately corresponding to the six questionnaire options, "morphology (stem, affixation & compounding)" was activated frequently (21 times) by the learners, which also showed their substantial reliance on bottom-up inferencing strategies.

7.6 Answering Research Question Three

The following summary will serve as an answer to the third major research question posed: How do ESL learners make use of their depth of vocabulary knowledge in the process of comprehending a general academic text?

1. The analysis of the interview data reveals that, in a lexical inferencing attempt, a learner
may use a single knowledge source, but may also use a variety of knowledge sources to facilitate the processing of lexical meaning. As the results have shown, learners frequently used knowledge sources representing different aspects of depth of vocabulary knowledge at the same time, and often simultaneously sought clues within the test word itself as well as from the context of the T-unit containing the test word. Occasionally learners would also look for clues beyond the immediate context in order to work out the meaning of a test word.

2. Depth of vocabulary knowledge plays a fundamental role in enabling learners to make lexical inferences. In the 67 valid inferencing attempts, *intralingual knowledge*, which conceptually represents depth of vocabulary knowledge in this study, was activated 68 times. In comparison, *knowledge of the world*, the only other knowledge source that the learners appeared actually to draw upon, was activated only 16 times.

3. Among the four sub-categories of knowledge sources established to represent different aspects of depth of vocabulary knowledge, *meaning* appeared to be the most heavily used knowledge source. The second most heavily used source appeared to be *morphology*. The other two sources were both minimally referred to. This finding at least indicates that the meaning and morphological aspects of depth of vocabulary knowledge were explicitly recognized as more important than the other two aspects, syntax and phonological/orthographic forms, in the lexical inferencing process for these L2 learners.

4. Both contextual clues and clues within test words are important in lexical inferencing because these two types of clue were both frequently utilized. This finding indicates that, while clues within the T-unit containing the target word were important in facilitating
lexical inferencing, the ability to make use of clues within target words was also fundamental. In the present study, close to half of the clues utilized were found within the test words themselves. Therefore, whether or not being able to utilize test-word clues will make a considerable difference in the lexical inference results. From this angle, the learner's depth of vocabulary knowledge appears, to a great extent, to dictate whether he or she will be able to make lexical inferences.

5. Comparing the HPs with the LPs, first, proportionally the HPs made more use of contextual clues than the LPs, but the LPs made more use of clues within test words than the HPs. This finding suggests that, although both groups employed "interactive" approaches to lexical inferencing, the HPs' approaches were more top-down than those of the LPs were. Second, meaning and morphology were the two most heavily used knowledge sources in inferring meanings of unknown words in the case of both HP and LP learners. However, HPs drew much more heavily on meaning than they did on morphology. In comparison, LPs relied slightly more on morphology than they did on meaning as a knowledge source. Furthermore, when morphology and phonological/orthographic forms were treated as a combined knowledge source focusing on word form, its frequency of use was much higher than that of meaning as a knowledge source. This finding suggests that, while learners with greater depth of vocabulary knowledge have the ability to use meaning as a major knowledge source in the lexical inferencing process, learners with less depth of vocabulary knowledge are often reduced to looking at the form of the target word itself, thus greatly reducing their opportunities to utilize clues and knowledge sources that involve knowledge of other words in the T-unit containing the test word or words outside the T-unit. Third,
according to the success rates for both the HPs and LPs, there appeared to be a positive relationship between learners' depth of vocabulary knowledge and their ability to succeed in inferring meanings of unknown words in reading. The greater the depth of vocabulary knowledge, the more effective the learner will be in lexical inferencing.

6. There was no evidence that knowledge of learners' L1 influenced their lexical guessing processes. This was probably because neither of the participants' L1s, Korean or Chinese, is cognate with English.
CHAPTER EIGHT

IMPLICATIONS AND CONCLUSIONS

8.1 Overview

In this chapter, a brief summary of major findings from the present research is first presented. Second, implications of the present findings are considered for further research and for educational practices in L2 classrooms. Finally, some limitations of the research are described.

8.2 Summary of Major Findings from the Present Research

The two studies in the present thesis research have established a number of important findings.

8.2.1 Correlation between vocabulary knowledge and reading comprehension

With respect to the correlational relationships among vocabulary size, depth of vocabulary knowledge, and reading comprehension, the results of the main study indicated strong correlations among the scores on the VS representing vocabulary size, the DVK and MK representing depth of vocabulary knowledge, and the RC representing reading comprehension. The intercorrelations among scores on the DVK, VS, and RC were especially high, indicating a strong association among the three tests, and therefore among the three components in the vocabulary knowledge-reading comprehension chain. These findings suggest that at least two aspects of depth of vocabulary knowledge in an L2, i.e., meaning and collocation, which are the two components of the DVK, are strongly, and positively, associated with reading comprehension. However, the follow-up study was only able to confirm "meaning" as a vitally important knowledge source in the lexical inferencing process as the frequency of using syntax
as a knowledge source, which includes collocation, was low (seven occurrences out of 68 valid attempts) and there was little evidence that any of these seven occurrences involved knowledge of collocation.

The strong association found between vocabulary knowledge (both breadth and depth) and reading comprehension in some ways supports the instrumentalist hypothesis in reading research (Anderson & Freebody, 1981), which claims that vocabulary knowledge directly affects reading comprehension, although the present research, without looking into the question of causation, showed only that breadth and depth of vocabulary knowledge and reading comprehension are closely, and positively, related to one another.

8.2.2 Predictive powers of vocabulary knowledge tests
An examination of the predictive powers of the VS, DVK, and MK via multiple regression analyses confirmed that scores on the DVK, which represented meaning and collocation components of depth of vocabulary knowledge, and on the VS, which represented vocabulary size, were both relatively unique, and distinctive, predictors of reading comprehension scores, and that the DVK scores had the capacity to improve the prediction of the RC scores over and above the estimation accomplished by the VS scores. However, the scores on the MK, representing some morphological sub-components of depth of vocabulary knowledge, failed to provide meaningful prediction of the RC scores. This may be due to the possibility that the MK failed to measure what it was designed to measure (see Section 5.3.2), or that one's morphological knowledge is simply not indicative of reading comprehension levels.

Nevertheless, these findings still suggest that depth of vocabulary knowledge has a
fundamental role to play in the relationship of vocabulary knowledge and reading comprehension because the present study has confirmed the DVK's capability of contributing significant and unique prediction of reading comprehension scores.

However, these findings should be interpreted with caution because of some constraints: Two independent variables, namely, VS and DVK, were highly correlated ($r = .82, p < .05$) and the sample size ($N = 74$) was relatively small (see also Section 8.5.2).

8.2.3 Lexical inferencing

The topic of lexical inferencing was investigated in a follow-up study, resulting in the following especially meaningful findings among its many results.

8.2.3.1 The rate of success and the Matthew effect in reading. A comparison of the HP group and the LP group revealed that the HPs' success rate was 2.3 times as high as that for the LPs. This result conforms to those from previous relevant studies (e.g., Hastrup, 1991; Morrison, 1996). The disparity in success rates in the present study reflects a positive relationship between learners' depth of vocabulary knowledge and their lexical inferencing ability: The greater the depth of their vocabulary knowledge, the more likely that learners will succeed in lexical inferencing while reading. The Matthew effect contends that vocabulary knowledge promotes reading comprehension, which in turn develops vocabulary knowledge (Stanovich, 1986). This rich-get-richer effect in L1 reading appears, from the present analysis, to also be applicable to lexical inferencing in L2 reading.
8.2.3.2 **Lexical inferencing strategies.** The results from the analysis of the interview data indicated that learners with different depths of vocabulary knowledge employed clues and knowledge sources differently. First, the HPs made more use of contextual clues than the LPs, who relied heavily, but not necessarily successfully, on clues within test words. This finding indicated a positive relationship between learners' depth of vocabulary knowledge and their ability to utilize contextual clues in lexical inferencing.

Second, of the two most frequent knowledge sources that learners drew upon, namely, morphology and meaning, HPs made more use of meaning while LPs more frequently used morphology. LPs sometimes used phonological/orthographic forms as a knowledge source, which deals with partial familiarity of word forms, whereas the HPs did not refer to that source at all. These findings suggest that, while HPs were able to rely on meaning as a major knowledge source in lexical inferencing, LPs were often reduced to looking at the form of the test word itself, because using meaning as a knowledge source often entails drawing on good knowledge of other words surrounding the test word, which was often too challenging a task for the LPs. This corroborates Cziko's (1980) finding that low proficiency learners pay more attention to word forms and less attention to contextual meaning, perhaps because the HPs normally expend less capacity than the LPs in attending to graphemic cues, and therefore they have sufficient capacity to attend to other types of cue, such as semantic cues, at the same time (Stanovich, 1990; Stanovich, Cunningham, & Feeman, 1984).

The different patterns of use of clues and knowledge sources thus made the HPs' strategies of lexical inferencing appear to be more top-down than those of the LPs. These different patterns can be linked to their different depths of vocabulary knowledge because of the
basis on which the HPs and LPs were identified and ranked in the present study.

8.2.4 Learners' perceptions and actual approaches to lexical inferencing

In the small-scale questionnaire survey, the learners stated that the two strategies they most frequently adopted were lexical meaning processing using clues beyond the immediate context and using world knowledge to infer lexical meaning. In other words, they perceived their approaches to lexical meaning processing as global and top-down. However, what they actually did while inferencing, as shown by the frequencies of their actual use of corresponding knowledge sources in their individual interviews, was more bottom-up. In their individual interviews, they mostly relied on local information to infer the meaning of test words. The two types of local information that the learners used most frequently in the inferencing process, namely, morphological clues within test words and semantic clues in the T-unit containing the test word, cover two aspects of depth of vocabulary knowledge, morphology and meaning, which once more demonstrates the fundamental role that depth of vocabulary knowledge plays in the lexical inferencing process.

8.3 Implications for Future Research

The findings from the present research indicate some directions for future research.

8.3.1 Relationships among sub-components of depth of vocabulary knowledge

The investigation into this relationship was carried out by assessing the intercorrelations among a host of variables representing several components of the vocabulary-reading chain, namely,
vocabulary size, depth of vocabulary knowledge, and reading comprehension. In the course of the data analysis, it was found that correlations between scores on the part-of-speech component of the MK and the two components of the DVK, namely, word meaning and collocation, were both just under .50 \((p < .05)\). This finding indicated that, although three components of the vocabulary-reading comprehension chain, namely, vocabulary size, depth of vocabulary knowledge, and reading comprehension, generally correlate highly with one another, within the dimension of depth of vocabulary knowledge components do not all correlate strongly with one another.

It is apparent that there is a need for further investigation of the relationships among various aspects within the dimension of depth of vocabulary knowledge, especially the relationships among aspects of meaning, collocation and part-of-speech knowledge, in order to explain why scores on the part-of-speech component of the MK did not correlate well with scores on the word meaning and word collocation parts of the DVK. Was it attributable to the likelihood that the MK was too easy for some participants (see Section 5.3.2), or due to the possibility that a learner’s explicit morphological knowledge is simply not closely correlated with reading proficiency? These questions are worth exploring.

### 8.3.2 Vocabulary tests as predictors of reading comprehension levels

The results of the present research manifested the distinctiveness and utility of the DVK as a measure to predict reading comprehension levels. The results also confirmed that the VS is a useful measure for the same purpose.

However, the DVK is still a research instrument since the number of its available
versions is very limited, and all the available versions are at the same difficulty level. In comparison, the Eurocentres Vocabulary Tests, as discussed in Chapters 1 and 2, are more mature and sophisticated in that there are tests in the same format for a variety of difficulty levels, and a large number of versions are available at each proficiency level.

For the DVK to become a practical measure to be applied effectively in educational settings, more research and development are called for. First, more versions of the test for the current difficulty level need to be developed and assessed. Second, tests need to be developed for at least two other difficulty levels, one to suit learners at the low-intermediate proficiency level and below, and the other to cover advanced learners, as the current test fares well with learners from intermediate to high-intermediate proficiency levels. The modification of part of the DVK in the present research demonstrated that it is feasible to further develop the DVK, in order to make it practically useful.

8.3.3 Vocabulary tests as predictors of other language skill levels

The present research investigated the value of vocabulary knowledge tests in predicting reading comprehension levels. The research did not touch upon using scores on vocabulary knowledge tests to predict proficiency for other language skills. Since "vocabulary knowledge is heavily implicated in all practical language skills" (Meara & Jones, 1988, p. 80), learners with a superior vocabulary knowledge should perform better in all language aspects than learners with limited vocabulary knowledge. This theoretical assumption is worth testing. Future studies should determine whether vocabulary knowledge tests can predict proficiency of other language skills, such as listening and writing. Laufer and Nation (1995) made a meaningful start in this direction.
Their investigation found that, for intermediate ESL learners, there is a close relationship between vocabulary size and lexical richness in writing. However, the scope of their study was limited to the relationship between vocabulary size and one kind of writing.

8.3.4 The threshold for lexical inferencing

An important finding from the present investigation was that HPs had a much higher success rate in lexical inferencing than LPs did. This finding corroborates the outcomes from a number of previous studies (e.g., Haastrop, 1990, 1991; Morrison, 1996). The finding confirmed that effective lexical guessing can only take place when learners have reached a certain threshold level of vocabulary knowledge. Haastrop (1990) suggested that there is a threshold for lexical inferencing. Based on her empirical results, Laufer (1992a, 1996) contended that 3,000 word families should be an adequate threshold for academic reading comprehension in English. The present research suggests that a threshold for lexical inferencing may be established based on depth of vocabulary knowledge since breadth and depth of vocabulary knowledge are substantially correlated. Taking this a step further, it should be possible to find a threshold for general academic reading comprehension based on depth of vocabulary knowledge.

Two questions thus arise. First, what depth of vocabulary knowledge is appropriate for the threshold? Second, will reading comprehension and lexical inferencing share the same threshold of vocabulary knowledge, or should there be different thresholds for the two somewhat different, but closely related processes? Lexical inferencing, as a form of word identification, is an important part of the reading comprehension process. However, lexical inferencing is ostensibly a more challenging process than the general reading comprehension process, as
suggested by the results of the present research, which found that 6 of the 12 participants failed to attain the 56% success rate for lexical inferencing in the follow-up study\(^1\). In contrast, of the 74 learners in the main study, only five failed to reach the 56% level in reading comprehension. The learners in the main and follow-up studies were all recruited under the same criteria, i.e., their vocabulary size must be over the 3,000-word threshold based on their scores on the VS. Therefore, while a 3,000-word-vocabulary appeared to work well as a threshold for general academic reading comprehension, it does not appear appropriate as a threshold for lexical inferencing, which seems to require a more demanding threshold. But what might the appropriate level be for this threshold? This question calls for further research.

8.4 Implications for Second Language Education

Findings from the present research have educational implications in the following areas.

8.4.1 Vocabulary knowledge tests for placement

The results of the present research indicated that both the DVK and VS are potentially useful as additional placement measures or even as replacements for reading comprehension tests in educational practice. However, if the DVK and VS are to replace reading comprehension tests in placement testing, two conditions need to be met. First, more studies with appropriate sample sizes and population representation need to be conducted to empirically confirm that scores on these two tests can reliably predict levels of reading comprehension. Second, more versions of

\(^1\) The standard of 56% was used in Laufer’s studies (1992a, 1996) for judging between readers and non-readers. The passing mark in reading comprehension was 55% at the university where Laufer’s studies were based.
the DVK and VS need to be developed for practical testing needs. Until then, these two measures can at best remain as research instruments.

8.4.2 Importance of depth of vocabulary knowledge in reading comprehension

The present research showed that depth of vocabulary knowledge is a fundamental component in the vocabulary knowledge-reading comprehension chain. ESL learners, especially less advanced learners, relied heavily on their vocabulary knowledge in lexical inferencing, which is an important part of reading comprehension. The research also found that learners with higher scores on vocabulary knowledge tests were generally proficient in reading comprehension. All this points to the benefit, importance, and necessity of improving the depth of learners' vocabulary knowledge in their ESL learning. As the present research showed, vocabulary knowledge can no longer be regarded as a simple matter of recognition of word meaning. Vocabulary knowledge not only has breadth, as represented by the number of words known to a learner, but also has depth. The working definition of depth of vocabulary knowledge (see Section 2.3.2.3) established in the present research involves a good number of aspects representing various types of linguistic knowledge.

The importance of depth of vocabulary knowledge in reading comprehension has been established in the present research based on empirical evidence. Now it is up to educators -- curriculum designers and ESL teachers -- to incorporate this factor into their ESL syllabi and teaching activities. It would also be highly desirable to communicate this importance to ESL learners so that they will be aware of the benefit of paying attention to aspects of vocabulary knowledge beyond superficial meaning recognition while learning English words, and learn to
be able to do so effectively.

8.4.3 Appropriate use of lexical inferencing

Recently, encouraging learners to guess word meanings from context has been a popular practice in second language instruction. In the questionnaire in the follow-up study, 11 of the 12 participants indicated that they frequently made guesses when they encountered unknown words in ESL reading. However, as the test results showed, some of these learners had a less than satisfactory level of reading comprehension and vocabulary knowledge. Is guessing, or inferencing, an effective means of processing lexical meanings for learners at every level?

Both Haynes (1993) and Haastrup (1990) believe that unless learners have reached a certain proficiency level they would not be able to achieve much success in lexical inferencing. Haastrup (1991) suggested that there is a threshold level which is critical for successful lexical inferencing. Results of the present research appear to support Haastrup's contention. As revealed in Section 7.2.1, the HPs made more successful inferences than the LPs. This finding argues that learners should not be urged to make lexical guesses from context at the early stages of learning a new language, though of course they may want or need to do so in some circumstances. The present interview data showed that even if LPs were willing to guess when they came across unknown words, their success rates were generally low compared with those of HPs. This implies that premature guessing attempts by LPs tend to be fruitless.

Inferencing by no means equals making wild guesses. Making lexical inferences involves use of a variety of knowledge sources which, as demonstrated by the results of the follow-up study, include not only depth of vocabulary knowledge, whose significant role in reading
comprehension is by now evident, but also other sources such as world knowledge. Meanwhile, it also involves use of contextual clues and clues within target words. Using these clues effectively is a skill that requires long-time honing; it involves developing vocabulary knowledge as well as other aspects of linguistic knowledge in a well-rounded way. Otherwise, as Haynes (1993) put it, even if sufficient contextual clues are present, LPs will still be unable to make correct guesses simply because they have difficulty in comprehending other words in the surrounding text that provide clues and because they often mismatch word forms and meanings.

8.4.4 Skills in dictionary use

The results of the survey showed that where possible learners would frequently consult unilingual English dictionaries or bilingual English-Chinese/Korean dictionaries to work out meanings of unknown words while reading English texts. This finding points to the importance of mastering skills in dictionary use, which can make a difference in reading comprehension. Effective use of dictionaries is not as easy as it seems to be (Hartmann, in press; Nesi & Meara, 1994). It is therefore desirable that training in dictionary use be part of educational activities in ESL classrooms so that learners will be able to use dictionaries effectively to support their reading.

8.5 Limitations of the Present Research

The present research was an initial inquiry into the role of depth of vocabulary knowledge in general academic reading comprehension. As such, limitations were certain to exist. The following sections discuss some limitations that have been identified (see also Section 8.2.2).
8.5.1 Population representation

As a variable control measure, the sample of the present research was basically university-level ESL learners with either Korean or Chinese as their L1. Since neither of these two languages is linguistically related to English, cognate influence of the participants’ L1 was kept to the minimum in the research. Thus, this control measure reduced complications in interpreting the results. However, this measure has, at the same time, limited the scope for generalization of the research findings. As the matter stands, all the findings obtained from the present research, however important they may be, need to be interpreted with caution. In other words, some qualification limiting the population applicability needs to be attached to any findings from the present research. Further studies with samples of more varied representation are definitely desirable if the present findings are to be applied to any populations other than Korean and Chinese university students.

8.5.2 Test reliability and accuracy

During the data collection period, the adaptation of instruments to meet time constraints may have resulted in less-than-ideal quality in the data. In order to recruit as many volunteers as possible for the research, the testing in the main study was proposed not to exceed 1.5 hours. The time-limit led me to shorten the length of the RC. Had the RC been longer, as it was originally designed, it could have measured reading comprehension more accurately, and its results could have been more statistically reliable, as indicated by the estimation of reliability on the RC based on the Spearman-Brown Formula (see Section 4.3.1). Furthermore, had the RC been kept at its original length of 30 items, there would have been a wider range of scores,
which might in some way have improved the results of the analyses.

8.5.3 Selection of the head words for the new part of the DVK

In the main study, DVKold, or scores on depth of general vocabulary knowledge, proved to be a more powerful predictor of reading comprehension levels than DVKnew, or scores on depth of knowledge of adjectives appearing in the RC. However, this conclusion was based only on a rather limited number of stimulus words chosen from the RC texts. In the present study, the 10 words used as the head words for the RC-text-related test, or Items 31-40 of the DVK, accounted for only 4% of the total number of words appearing in the texts of the RC. DVKnew might have indicated greater predictive value had a larger percentage of words from the RC texts been incorporated into the RC-text-related test.

8.5.4 The form of qualitative data collection

Two forms are currently popular for collecting data on lexical meaning processing, namely, individual interviews (e.g., Read, 1989) and pair-think-aloud protocols (e.g., Haastrup 1991; Morrison, 1996). The present research adopted the form of an individual interview because it did not require training sessions for the learners and because, as the interviewer, I could control the pace of the interviews, as it had been promised that each interview session was not to exceed 45 minutes.

However, the interview process revealed a shortcoming which could perhaps be rectified by using pair-think-aloud protocols; that is, it revealed limits on information flow. Generally speaking, all the data obtained from an interview depends on the questions asked. When a
question was not asked at all due to time constraints, there could be no answer to that question. In the present research, quite a number of lexical inferencing attempts were deemed invalid as there was not sufficient information provided by the learner for classification according to the analytic framework. This happened simply because on some occasions follow-up questions were not asked to clarify the interviewee's responses. If properly trained, informants would be able to provide more useful data in a pair-think-aloud session. One advantage of using the pair-think-aloud procedure is that information flow can be relatively easily maintained when two peers are discussing an unknown word, and therefore more information will be obtained.

8.6 Concluding Remarks

The central purpose of the present research was to explore the relationships among vocabulary size, depth of vocabulary knowledge, and reading comprehension. The project was accomplished through a main study of 74 ESL learners with a Chinese or a Korean L1 background and then a follow-up study with 12 participants recruited from among the 74 learners. The results from the main study generally support the two hypotheses stated in Chapter 2. The follow-up study has provided insights into ESL learners' lexical inferencing processes and identified some patterns of use of clues and knowledge sources by learners with different depths of vocabulary knowledge. The findings from the present research have meaningful implications for future research on L2 vocabulary knowledge, lexical inferencing and reading comprehension, as well as for administration of ESL placement testing, syllabus design and instructional practices for L2 classrooms.

In conclusion, the present thesis has shown with empirical evidence that depth, or quality,
of vocabulary knowledge plays an important role in ESL reading comprehension. Results from three aspects of the present research support this claim:

1. Intercorrelations among most of the variables measured in the main study were high, which suggests that the levels of depth of vocabulary knowledge and reading comprehension are closely, and positively, associated with each other.

2. Scores on the DVK, which was the main measure representing depth of vocabulary knowledge in the research, added a unique and distinctive portion of prediction of RC scores over and above the prediction afforded by the vocabulary size scores, which suggests that scores on depth of vocabulary knowledge can be a useful predictor of reading comprehension scores.

3. L2 learners involved in the follow-up study relied heavily on their vocabulary knowledge in making lexical inferences, which suggests that depth of vocabulary knowledge matters a great deal in the process of inferring lexical meaning, and hence, in reading comprehension.

Although the present research was able to show the importance of depth of vocabulary knowledge as a factor in the chain of vocabulary knowledge-reading comprehension in ESL, more research is desirable in order to clarify questions that arose in the course of the present research, especially questions about the threshold for reading comprehension and lexical inferencing (see Section 8.3.4).

Trying to explain the close links between vocabulary knowledge and reading comprehension is a complex and demanding task. In order to identify all the relevant factors and to clarify the relationships among these factors, more theoretical explorations and empirical
research are needed. The present research has been able to accord legitimacy to one fundamental and previously-neglected factor, namely, depth of vocabulary knowledge, but there are still dark corners. For example, what other components in the factor chain have been neglected? Are breadth and depth the only two dimensions of vocabulary knowledge? It will take a great deal more research just to answer these seemingly straightforward questions.
REFERENCES


Dear Friend:

Greetings! I am a doctoral student at the Ontario Institute for Studies in Education of the University of Toronto, currently doing my thesis research on the relationship between ESL learners’ vocabulary knowledge and reading comprehension, and how learners make use of their vocabulary knowledge to comprehend English texts. This study is important because it will help understand the important role vocabulary knowledge plays in reading comprehension. I would like to include you in this research.

The research will be in the form of a one-page background questionnaire, three tests on vocabulary knowledge, and a test on reading comprehension. The total testing time will be about 1.5 hours. A few weeks after the test, a very small number of participants will be randomly selected to attend an interview. If you are selected, you will be invited to individually discuss with me how you use your vocabulary knowledge in reading comprehension. The interview will last about 45 minutes. I can assure you the test and the interview will have no impact on your academic record and the data from the tests and the interview will be kept strictly confidential. When the tests and the interview are completed, each participant’s name will be replaced by a code, and I will be the only one having access to these codes.

I sincerely hope you will support this research by participating in the study. If you agree to participate, please fill out the form below and return it to your teacher. You will be free to withdraw from the study at any time. If you have any questions, I can be reached at ______. Thank you for your cooperation.

Sincerely,

David D. Qian

CONSENT FORM

Name________________________________________ Birthdate________________________________________

First Language_________________________ If Chinese, __ Mandarin, or ___ Cantonese

Number of years in Canada _____ Number of years you have studied English _____

Check (✓) here:

___ I am willing to participate in the above study conducted by David Qian.

___ I am not willing to participate in the above study conducted by David Qian.

Date ___________________________ Signature ___________________________

Telephone Number (_______) _________________________________
APPENDIX B: BACKGROUND QUESTIONNAIRE

Please tell us something about yourself.

1. Name ___________________________. 2. Year of Birth ___________________________.
3. I am originally from ___________________________ (name of home country).
4. My first language is ___________________________ (If Chinese, ___ Cantonese ___ Mandarin)
5. I can speak the following languages: ___________________________.
6. I have been in Canada for ___ years ___ months.
   I will stay in Toronto until ___________________________, 199__.
7. I have been learning English for ___ years since ___________________________.
8. At home, my parents mainly speak (Please ✓ one):
   ___English ___Cantonese ___Mandarin ___other (specify ___________________________).
9. I am currently living (Please ✓ one):
   ___with parents ___by myself ___with relatives ___with friends ___other
10. The highest level of education I have completed (Please ✓ where applicable):
    1) ___ high school ___ grade 10/ ___ grade 11/ ___ grade 12/ ___ OAC.
    2) ___ college in ___________________________ (name of country).
       My major was/is ___________________________.
       The length of program was/is ______ years ______ months.
    3) ___ university in ___________________________ (name of country).
       My major was/is ___________________________.
       The length of program was/is ______ years ______ months.
       My highest degree obtained is ___________________________.
       ___ I graduated. ___ I am currently a 1st/2nd/3rd/4th year student.
11. My plan for future education or career:
    ___ I intend to go to a vocational/technical/business school in Canada.
    ___ I intend to go to a community college in Canada.
    ___ I intend to go to university in Canada:
       Program Level: ___ undergraduate ___ masters ___ doctoral
       ___ I intend to look for a full-time job as ___________________________.
       ___ Other. Please specify ___________________________.


## APPENDIX C: DEPTH-OF-VOCABULARY-KNOWLEDGE TEST
(ITEMS 31-40)

### 31 common
| complete | light | ordinary | shared | boundary | circle | name | party |

### 32 tight
| close | rough | uncomfortable | wet | schedule | pants | surface | wood |

### 33 potent
| persuasive | powerful | weak | unknown | drug | weather | arguments | telephone |

### 34 brilliant
| dull | shining | outstanding | splendid | office | problem | success | knot |

### 35 portable
| movable | deep | light | fixed | person | device | recorder | trees |

### 36 surrounding
| central | internal | linear | encircling | neighborhood | areas | trees | math |

### 37 prodigious
| vast | expensive | ordinary | amazing | sum | prison | appetite | person |

### 38 digital
| numerical | valuable | binary | body | liquid | computer | keyboard | wind |

### 39 sound
| logical | healthy | bold | solid | snow | temperature | sleep | dance |

### 40 powerful
| moderate | strong | effective | practical | computers | desks | weapon | patients |
APPENDIX D: MORPHOLOGICAL KNOWLEDGE TEST

Explain the meaning of the following 10 words in English, or translate them into your first language. If the underlined part of the word has changed the part of speech of the base word, please indicate this by using appropriate terms provided below.

Possible terms: *noun, verb, adjective, adverb, preposition*

<table>
<thead>
<tr>
<th>Examples</th>
<th>Meaning or translation of the word</th>
<th>The change in part of speech caused by the underlined part</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: homeless</td>
<td>without home</td>
<td>n. $\rightarrow$ adj.</td>
</tr>
<tr>
<td>B: impossible</td>
<td>not feasible</td>
<td>No change (adj. $\rightarrow$ adj.)</td>
</tr>
<tr>
<td>1 unknown</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 prerequisite</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 extraordinary</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 useful</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 findings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 prejudgment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 memorable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 evaporation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 rejection</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 awareness</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX E: QUESTIONNAIRE ON READING STRATEGIES

People have different ways of dealing with new words in a second language. There is no “right” way or “wrong” way. This questionnaire asks you what you actually do, not what you think you should do, in reading an English text. Please complete it to help me understand how you deal with words you don’t know.

1. How often do you do each of the following when you meet an unfamiliar word in reading an English text? (Please check one box for each item.)

<table>
<thead>
<tr>
<th></th>
<th>often</th>
<th>sometimes</th>
<th>rarely</th>
<th>never</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Look up the word in an English-Korean dictionary</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>b) Look it up in an English-only dictionary</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>c) Guess its meaning from the context</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>d) Ignore it</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>e) Ask the teacher for assistance</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>f) Ask a friend if they know it</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>g) Look for clues to meaning in the word itself</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>h) Make a note of the word (i.e. write it down)</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>i) Other (specify)</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
</tbody>
</table>

2. What kind of information do you use when trying to guess the meaning of an unfamiliar word in an English text? (Please check one box for each item.)

<table>
<thead>
<tr>
<th></th>
<th>often</th>
<th>sometimes</th>
<th>rarely</th>
<th>never</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) I examine the unknown word to see if it contains any grammatical clues to tell me what part of speech it belongs to</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>b) I use the meaning of other words in the same sentence to help me guess the meaning of the unknown word</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>c) I use my background knowledge of the topic of the text to guess the meaning of the unknown word</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>d) I look for grammatical clues in the surrounding sentence to help me guess the meaning of the unknown word</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>e) I examine the unknown word to see if any part of it is familiar in meaning</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
<tr>
<td>f) I make use of the meaning of the paragraph or text as a whole to guess the meaning of the unknown word</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
<td>[ ]</td>
</tr>
</tbody>
</table>

Thank you for your help.
APPENDIX F: EPISODES FROM INDIVIDUAL INTERVIEWS

In these episodes, a uniform legend is applied: R stands for the researcher, or interviewer, and L stands for the learner, or interviewee. The T-unit which contains the test word in question is supplied, with the test word underlined, prior to presenting each episode. An interpretation of the inferencing attempt follows each episode. Unless otherwise specified, all the episodes provided here were originally spoken in English.

Episode #1

[T-unit: There is a greenhouse to suit every requirement, heated or unheated, small or large, permanent or portable, decorative or functional.]

R: Now what does "permanent" mean here?
L: Permanent ... I think it is opposite to portable.
R: So what word would you use if I take out "permanent" from here?
L: Ah, fixed.
R: Fixed, Okay. In your last exercise you used forever, lasting, but it doesn't mean this here.
L: Yeah.
R: Did you know this means "fixed" before you saw the text?
L: No, no.
R: So what words helped you to figure out this meaning?
L: Permanent?
R: Yeah.
L: Portable.
R: Why?
L: Because in other things, like "small or large", they have opposite meanings.
R: So it's basically based on the meaning of other words?
L: Yeah. (Transcript #69)

In this inferencing attempt, the learner mainly used as clues the meaning of other words in the T-unit context. Phrases such as "small or large" gave her the direction that the meaning of "permanent" should be "opposite to portable", i.e., its antonym. Therefore, she chose "fixed" as the meaning for "permanent" in this context, and successfully accomplished the inferencing task. This episode was categorized as: 1b (clues from the T-unit context) and 3d (meaning).

Episode #2

[T-unit: Prejudice is the result of powerful emotions, not of sound reasoning.]

R: ... Okay. What about this "sound reasoning"? Could you explain the meaning of these
two words separately?
L: "Sound" has two meanings.
R: What are they?
L: Sound is like something you hear. And another meaning of sound is healthy.
R: Healthy?
L: Prejudice is not related to the reasoning, ration, just related to emotion.
R: Related to emotion? Okay. So here, what meaning is used here?
L: Sound?
R: Yeah.
L: Something I can hear.
R: Something you can hear?
L: No, wait a minute! Second meaning.
R: Second meaning?
L: Yeah. Not healthy but good.
R: Good? Okay. Why do you think it's the second meaning here?
L: Eh, I thought about grammar. Usually when I use this sound as a noun or as a verb, it's not suitable ... here.
R: So this is not noun or verb?
L: No verb, no noun. And the structure, from here to here, it's the same. "The result of powerful"... powerful is adjective, so "powerful emotions", "not of sound reasoning", it has the same structure. So "not of sound" it means it's like this, so maybe it must be adjective. That's why I chose the second meaning.
R: So you figure this out based on grammar. You didn't think of the meaning here, just grammar?
L: Actually the meaning also, I thought about the meaning, also. But the grammar thing came to my mind first. (Transcript #45)

This learner used multiple knowledge sources (parts of speech, syntactic structure and polysemy) in inferencing the meaning of "sound" in this context. She knew that "sound" is a polysemous word which could mean "something you hear" or "healthy". She also knew that when it means "something you hear", it must be a noun, and when it means "healthy" it must be an adjective. Her analysis of the grammar of the T-unit indicated that "sound" here must be an adjective because a parallel phrase "powerful emotions" in the T-unit had "powerful" as the adjective, which helped her to decide that "sound" means "good" here based on an conversion of the meaning "healthy". The inferencing attempt, categorized as la (clues within the test word), 1b (clues within the T-unit context), 3b (morphology), 3c (syntax), and 3d (meaning), was successful.

Episode #3

[T-unit: Prejudice is the result of powerful emotions, not of sound reasoning.]
R: Okay, then could you explain "sound reasoning" here? What does "sound" mean here?
L: Sound means is a some rumor or some gossip.
R: Sound reasoning?
L: Sound means gossip.
R:Means gossip?
L: Eh, just... see if... it's correct evidence, not of sound reason. Prejudice is the result of powerful emotions... I think "sound reason" is good words. So you have to... you have to sound reason, you have to look for sound reason, so people can judge.

(Transcript #64)

In inferring the meaning of "sound" in this T-unit, this learner failed to make use of the contextual and grammatical clues as employed by the learner in Episode #3. The learner in the present case basically relied on semantic clues within the test word. It appears that at the first she incorrectly made an association of a primary meaning of "sound", "something heard", with "rumor" and "gossip", and then concluded that "sound reason" were good words and that "sound" had to be "gossip" in the context. Later, she seems to be moving away from the "gossip" idea and getting closer to the correct meaning. However, she never really hit on the right idea. This inferencing attempt, categorized as 1a (clues within the test word) and 3d (meaning), was deemed unsuccessful.

Episode #4

[T-units: Most greenhouses, however, come in two basic materials, wood and aluminum, each possessing advantages and disadvantages: wood is warmer in winter but is prone to decay unless regularly maintained; aluminum does not deteriorate but conducts the cold inside in winter...]

(The researcher asks the learner to guess the meanings of all the words she has underlined...)

L: Deteriorate?
R: Deteriorate.
L: It's... this meaning is "prone to decay". I think so.
R: Why do you think this?
L: Because this sentence compare wood and aluminum, and first the sentence explains about what is the benefit and the disadvantage of wood. Actually they said wood is prone to decay, and next they explained about aluminum, and said not deteriorate, so I can guess it's just comparing wood and aluminum. So it's meaning of, maybe this paragraph show what's the benefit of aluminum compared to wood. So they actually compare the disadvantage of wood and the... the point is prone to decay. So I think deteriorate is, meaning is prone to decay. (Transcript #63)
This learner used contextual meaning to work out the meaning of the test word "deteriorate". She made use of the meaning of the T-unit, and also went beyond the T-unit to look for the meaning of the previous clauses. Based on what she found about the purpose and meaning of these clauses, she concluded that in the context "deteriorate" should mean "prone to decay", which was deemed a successful attempt. This episode was categorized as: 1b (clues within the T-unit context), 1c (clues beyond the T-unit context), and 3d (meaning).

Episode #5

[T-unit: aluminum does not deteriorate but conducts the cold inside in winter;]

R: Okay. So next. I guess you figured out this meaning (referring to deteriorate) already. Can you use a word to replace this?
L: Go bad?
R: Go bad? Okay. If I ask you to find a similar word here in the text, can you do it?
L: Decay?
R: Decay. Okay. How did you figure it out?
L: Actually I have seen this word before. "De-" means something negative. So I can guess. And they are talking about decay. So from the previous context, I can guess this meaning. (Transcript #45)

This learner successfully solved the problem with the help of her morphological knowledge. Based on the meaning of the previous context, which was talking about decay, and with her knowledge that "de-" implied the test word had a negative meaning, she concluded that the meaning of "deteriorate" was similar to that of "decay". Usefulness of the meaning of context and knowledge of morphology was clearly demonstrated in this case. This episode was categorized as: 1a (clues within the test word), 1b (clues within the T-unit context), 1c (clues beyond the T-unit context), 3b (morphology) and 3d (meaning).

Episode #6

[T-unit: The greenhouse is a common feature in many gardens and is indispensable to the gardener, as it provides near-perfect conditions for plants, both edible and ornamental, all year round, irrespective of the weather]

R: Okay, let's leave it just like that. Now do you have a better idea of this word "irrespective"?
L: "Irrespective of the weather" means ... "irrespective" means just change, variety.
R: Variety?
L: Yeah, I think so.
R: Because in your exercise you said it means bad, ignored. Your explanation now is different, right?
L: Yeah.
R: So which is which?
L: Sometimes first time I thought, respect and respective, in this passage, means some bad meaning, negative meaning. So, irrespective of the weather, eh. strange weather.
R: Strange weather? Okay.
L: Variable, various weather. It can not be predict, or indicate. (Transcript #6)

This learner started her inferencing by relating "irrespective" to "respect" and "respective", obviously using her knowledge of affixation. This might have been a promising path had she known that English words sharing the root "respect", which in Latin means "pay attention to", vary in their meanings. However, this route was too challenging for her as she lacked the depth of the knowledge needed. Therefore, she could only conclude that "irrespective" in the context meant something bad or negative, perhaps due to her knowledge that "respect" has a positive meaning and that the initial affix ir- implies a contrary meaning. This episode was categorized as: la (clues within the test word) and 3b (morphology).

Episode #7

[T-unit: Prejudice is the result of powerful emotions, not of sound reasoning.]

R: So what does this word "emotion" mean?
L: Emotion ... moving ... like emotion ... no, no, no. Powerful emotion, like moving.
R: Moving, like from this place to that place?
L: Yeah. But this not a thing. We can't see.
R: You can't see. It's a movement, but you cannot see.
L: Yeah, yeah.
R: Does it have to do with people, or with instrument, or with any other thing?
L: This one? Not instrument, I think so. Maybe, like people's mind.
L: Yeah, thinking of people. People's thinking.
R: Do you know the Korean word for this?
L: Yeah, I know Chinese characters for this.
R: Can you write it down for me? ... [The interviewee wrote a Chinese character xin, standing for "heart" or "mind"] Okay. Close to it, but not exact.
L: Yeah, not exactly. (Transcript #71)

The learner believed "emotion" to be similar in meaning to "motion" probably because the two words look alike. She thought of "emotion" as a kind of abstract motion in the mind, or "people's thinking", as she stated. However, "emotion" is very different from "thinking", which was a point the T-unit tries to make when it contrasts "powerful emotions" with "sound reasoning". Therefore, the inferencing attempt, categorized as 1a (clues within the test word) and 3a (phonological/orthographic forms), could not be deemed a success.
Episode #8

[T-unit: The other surprising research finding was that heavily sweetened cereals proved about equally potent in causing decay whether they contained eight percent sugar or almost eight times that much.]

R: So, you just indicated you don’t know this word “potent”, right?
L: Potent?
R: Can you guess?
L: Potential.
R: Potential?
L: Potentiality.
R: Potentiality?
L: "Equally potent" means it might break, it might occur some disease according to sugar, like diabetics. (Transcript #35)

The learner was obviously drawing on her knowledge of morphology in this attempt. However, as the transcript shows, she did not comprehend the whole T-unit correctly, and could not effectively used the meaning of the T-unit to help her infer the meaning of the test word "potent". Relying on her knowledge of the meaning of morphologically related words in this case did not help her much. This episode was categorized as: 1a (clues within the test word) and 3b (morphology).

Episode #9

[T-unit: Prejudice means literally prejudgment, the rejection of a contention out of hand before examining the evidence.]

R: Okay. So here you indicate you don’t know the word “contention”?
L: Un huh.
R: Can you guess from the text?
L: Eh, I think it’s "context" or something.
R: "Context" or something? So if you are asked to use one word to replace "contention", you would choose "context"?
L: Eh, "content".
R: Content? Why do you think it’s content?
L: Mmn ... maybe it’s like something in ... There is a phrase, "out of hands". Something in my hands is content.
R: Okay. So because of this "out of hands"?
L: Yeah.
R: So you think that’s something in your hand, eh?
L: Or something in my idea. (Transcript #45)
In this attempt, the learner seemed to think "contention" belonged to the same word family as "context" or "content". She first thought of "context", but after taking into consideration the meaning of the phrase "out of hand", she proposed "content" as a substitute for "contention". The attempt was not successful as the meaning of "contention" has little to do with either "context" or "content": it is derived from the verb form "contend". This was in fact a difficult inferencing task for her since neither the test word itself nor its neighbouring phrase "out of hand" provided her with much help when she failed to understand the real meaning of "out of hand" as an idiom. Under the circumstances, the meaning of the whole T-unit, which should have provided her with some clue, was not utilized effectively. This episode was categorized as: 1a (clues within the test word), 1b (clues with the T-unit context) and 3b (morphology).
IMAGE EVALUATION
TEST TARGET (QA-3)