

English Applied Linguistics PhD Program
Graduate School in Linguistics
University of Szeged

**The written assessment of the vocabulary knowledge
and use of English majors in Hungary**

PhD dissertation

Doró Katalin

Supervisor:
Dr. Donald W. Peckham

Szeged
2008

TABLE OF CONTENTS

PART I – GENERAL INTRODUCTION	1
Chapter 1 – Introduction and research aims.....	1
1.1 Importance of the domain	1
1.2 Defining vocabulary	2
1.3 General purpose of assessment	3
1.4 The Hungarian context.....	5
1.5 Specific research aims.....	8
1.6 Structure of the dissertation.....	10
PART II – LITERATURE REVIEW.....	12
Chapter 2 – Terminological clarification of the construct <i>word</i>	14
2.1 Introduction	14
2.2 What is a word?.....	14
2.3 Multi-word units.....	18
2.4 One word – multiple meanings	20
2.5 Concluding summary	21
Chapter 3 – Vocabulary knowledge.....	22
3.1 Introduction	22
3.2 Vocabulary knowledge frameworks.....	23
3.3 Aspects of vocabulary knowledge.....	29
3.4 Concluding summary	32
Chapter 4 – Assessing learners’ vocabulary	33
4.1 Introduction	33
4.2 Design of vocabulary measures.....	34
4.3 Tests of vocabulary knowledge.....	38
4.4 Assessing the vocabulary of written texts.....	51
4.5 Word frequency.....	56
4.6 Concluding summary	57
Chapter 5 – Vocabulary and its relationship with reading and writing.....	59
5.1 Introduction	59
5.2 Relationship between vocabulary and reading.....	60
5.3 Relationship between vocabulary and writing	73
5.4 Concluding summary	80
PART III – Empirical investigation	82
Chapter 6 – Research questions	84
6.1 Introduction	84
6.2 List of research questions.....	84
6.3 Explanation of research questions.....	85
6.4 Concluding summary	91
Chapter 7 – Methodology.....	92
7.1 Introduction	92
7.2 Participants.....	92
7.3 Research instruments.....	96
7.4 Procedure.....	102
7.5 Data handling and scoring.....	103

7.6 Piloting of research methodology	105
7.7 Concluding summary	107
PART IV – Results and discussion	108
Chapter 8 – Vocabulary knowledge as assessed by tests	109
8.1 Introduction	109
8.2 Study 1 – Descriptive vocabulary test data	111
8.3 Study 2 – Factors influencing test results.....	117
8.4 Study 3 – Relationship between test results	135
8.5 Concluding summary	141
Chapter 9 – Vocabulary knowledge and use	142
9.1 Introduction	142
9.2 Study 4 – Vocabulary knowledge and academic success	145
9.3 Study 5 – Vocabulary knowledge and use in written text production	157
9.4 Study 6 – Factors influencing the lexicon of written production tasks.....	164
9.5 Concluding summary	176
Chapter 10 – Study 7: Changes in vocabulary knowledge over one academic year.....	177
10.1 Introduction	177
10.2 Aim.....	178
10.3 Procedure.....	179
10.4 Results	179
10.5. Discussion of results.....	187
10.6 Concluding summary	188
Chapter 11 – Overall discussion of results.....	189
11.1 Introduction	189
11.2 Tests of vocabulary knowledge.....	189
11.3 Theoretical issues	190
11.4 Assessment implications	191
PART V – Conclusions	194
Chapter 12 – Review and conclusions	194
12.1 Introduction	194
12.2 Overview of the dissertation	194
12.3 Conclusions and implications.....	197
REFERENCES.....	199
LIST OF TABLES AND FIGURES	210
APPENDICES.....	213
Appendix A – Vocabulary Levels Test.....	213
Appendix B – Productive Vocabulary Levels Test.....	217
Appendix C – Lex30	219
Appendix D – Questionnaires	220
Appendix E – Written production tasks	225

PART I

GENERAL INTRODUCTION

CHAPTER ONE

INTRODUCTION AND RESEARCH AIMS

1.1 Importance of the domain

The last few decades have witnessed a boom in the interest towards vocabulary-related issues both from specialists and non-specialists including, but not limited to, applied linguists investigating first and additional languages, corpus linguists, language instructors, curriculum planners, textbook and dictionary writers, and language learners. This interest has resulted in numerous books and anthologies on second and foreign language (L2) vocabulary, with emphasis on the English language, some of the most influential ones being Singleton (1999) on the second language mental lexicon, Schmitt and McCarthy (1997), Schmitt (2000), Read (2000), Nation (2001) and Weigle (2002) on L2 vocabulary learning and assessment, and Granger (1998) on corpus studies in learners' English. Parallely, in recent years, many articles have been published that discuss previous theories and research, introduce empirical data or offer a place for academic debate. What emerges from this ever expanding literature is not only the growing body of data and knowledge related to L2 vocabulary studies, but also the growing number of questions and ambiguity in areas such as terminology, data collection methods and data interpretation.

Research has found that vocabulary knowledge plays an important role in reading comprehension (Laufer & Yano, 2001; Nassaji, 2004; Hunt & Beglar, 2005; Shiotsu & Weir, 2007), oral and written text production (Muncie, 2002), as well as overall proficiency test results (Laufer, 1992; Morris & Cobb, 2003; Zimmerman, 2004). Previous studies have also reported that receptive vocabulary size is a good predictor of effective reading and academic success in general. For example, Laufer (1997a) and Nation (2001) argue that 95–97% of the running words in a text need to be known in order to guess the meaning of the remaining words and, subsequently, gain adequate understanding of the given text. Other studies, including Hirsh and Nation (1992) and Hu and Nation (2000), raise this threshold level to

98%, still leaving about one unknown word in every five lines. This means a minimum of 3,000 word families for reading authentic general texts and 5,000 word families plus good academic or technical vocabulary coverage for reading university textbooks, academic articles and literary texts.

In short, it can be stated that without adequate vocabulary knowledge, conveying meaning and message in English for both general and academic purposes is likely to fail. When the target language is the medium of education, this failure may impose limitations on academic achievement. Yet, without clear understanding of how to measure students' vocabulary, how to interpret the data, and how measured knowledge relates to the use of the target language, the information so far gained on the topic is seriously limited.

The general goal of this dissertation is to understand, first, the multifaceted nature of vocabulary knowledge and how it is related to various measurement instruments; second, what the actual vocabulary knowledge and use of Hungarian students of English is; and third, how this knowledge can be turned into our better understanding of theoretical and assessment issues, while providing practical answers to instructors, students and syllabus designers.

1.2 Defining vocabulary

Even the seemingly simple question of what vocabulary is needed for a language learner will soon prove to be challenging to answer, as it brings with itself numerous other questions. We need to understand what we mean by *vocabulary*, by *word* and by *knowledge*. It is often tempting to compare L2 learners' needs or knowledge to those of native speakers. Many of the same questions raised above are also valid in this case. However, one can only estimate the size of native speakers' lexicon, as it is impossible to test all words in a language. Of course, not even an educated native speaker can ever acquire and use all the lexicon of the L1. Nation and Waring (1997), while reviewing vocabulary studies that aim to estimate the size of an English native speaker's lexicon say that "we should expect that native speakers will add roughly 1,000 word families¹ a year to their vocabulary size. This means that a five year old beginning school will have a vocabulary of around 4,000 to 5,000 word families" (Nation &

¹ For a definition of word families see Chapter 2.

Waring, 1997: 7). They estimate that an English speaking university student, therefore, would know about 20,000 word families.

Meanwhile, it needs to be underlined that knowing a word means much more than the superficial meaning of its form and most frequent meaning. One way of approaching the multifaceted nature of vocabulary knowledge is by looking at the various knowledge types or components that Richards (1976) or Nation (2001) proposes. In their interpretation, lexical knowledge includes the orthographical and phonological form of a lexical unit, its meanings, grammatical behavior, associations, collocations, frequency and register. Other studies (including Chapelle, 1998; Henriksen, 1999; Qian, 1999, 2002; Qian & Schedl, 2004) have proposed frameworks that describe lexical knowledge in terms of various dimensions, such as depth and breadth, lexical organization, partial and precise knowledge.

For purposes of this dissertation, which targets the written assessment of English majors' vocabulary knowledge and use, the discussion of their vocabulary focuses on what can be assessed and what is used in a written context. This is done with the understanding that some other aspects of vocabulary knowledge and use are assessable only through oral measures. Although many of the issues of vocabulary investigated here apply to oral language communication, speaking and listening will not be fully explored. Therefore, drawing upon various interpretations of the construct *vocabulary*, I have adopted my own definition with focus on assessment and use, rather than on teaching or acquisition. As such, the following definition of vocabulary underlies the discussion in this dissertation:

Vocabulary is the knowledge of forms, meanings and usages of words that allow people to understand and convey meaning in written texts. Vocabulary knowledge will be referred to as the types of sub-knowledge of forms, meanings and usages assessable by measurement instruments, while vocabulary use will be defined as activities in which written vocabulary is employed either receptively (reading) or productively (writing).

1.3 General purpose of assessment

If the question of what vocabulary knowledge is needed is reformulated in the context of Hungarian university English majors, we are getting one step closer to an answer regarding

vocabulary needs². Upon entering the university as L2 language majors, students are placed into an academic environment, where, on the one hand, the target language is studied more intensively and at a higher proficiency level than in secondary education, yet, on the other hand, students face the challenge of doing their studies in English, as it becomes the primary medium of instruction. They are soon expected to follow academic courses, read academic texts and show their knowledge using sophisticated language. However, the transition from secondary level language classroom to a higher level English major is not an easy one. Demands are significantly different on linguistic, cognitive and study skills levels. The difficulty of this transition between school levels in an English as a foreign language (EFL) context has not been widely studied, but can be expected to demand an excessively great task from students in the light of Smith's (2004) recent study among British first-year students of English. The study involved native speakers in five institutions in Britain, many of whom experienced difficulty with note-taking during lectures, type and amount of reading and writing expected of them. More specifically, they had to learn to change their copious note-taking strategies into producing reviewed and summarized lecture notes, to cope with both primary and secondary critical readings and the thinking process involved in academic writing. Since some of the demands expected of EFL students of English are similar to those of native speakers discussed by Smith (2004), it is not surprising if many of the first-year and even more advanced students are facing serious difficulties with these demands.

Several studies have shown that a large vocabulary size of general English and the knowledge of a high proportion of the Academic Word List (Coxhead, 2000) are vital for academic achievement. Therefore, students need to have a high language proficiency level and greatly expand their vocabulary during their university studies as a basis for any activity related to their studies, such as following language and content classes and conducting individual studies in the form of reading and writing. In order to do so, they need to become more independent learners than during their secondary school education, and they are expected to build a large lexicon including academic vocabulary.

Many scholars argue that testing should reflect this multidimensional nature of vocabulary competence (e.g. Zareva, 2005; Wolter, 2001; Nation, 2001; Singleton, 1999). Attempts have been made to include a number of knowledge dimensions into vocabulary

² Confront it with Hazenbun and Hulstijn (1996) who find 10,000 base words of Dutch necessary for non-native students' university studies in the Netherlands.

testing and continuously improve test designs to capture as much of the vocabulary knowledge and use and the underlying processes as possible (Nation, 2001). However, it needs to be stressed that there is no commonly accepted vocabulary test of English that can measure *the* vocabulary of a learner as such. Also, due to time limit and practicality, institutionalized testing sessions and research usually limit themselves to the use of one or maximum two tests of vocabulary knowledge; therefore, gain information on a limited number of aspects of the learners' lexical proficiency.

This dissertation seeks to move in this double, conflicting ground of theoretical and practical needs, on the one hand, calling for the clarification of research instruments and data interpretation, and, on the other hand, having the need of simple testing methods that are easily available and readily usable in language instruction and syllabus planning. Rather than designing alternative testing instruments, this dissertation wishes to employ data collection methods and research tools (tests of various formats and text analyzers) that have already been partly used in university contexts, both in Hungary and elsewhere, and that could be easily introduced in a wider context either as diagnostic measures of students' vocabulary or as tools to be integrated in instruction, or those that could be directly used by students during their individual language studies. It is of both theoretical and practical value to gain information on how these instruments can inform us about students' lexical knowledge and use of which, so far, we could only have partial insight or assumption on the basis of limited data and everyday observation of coursework and language exam results.

To achieve this aim, a multi-level approach is adopted: first, scores obtained on tests of controlled receptive, controlled productive, and free productive vocabulary knowledge are analyzed and correlated to each other and to factors that could influence this vocabulary knowledge. Second, test scores are also investigated in terms of their predictive nature of academic success in a target language medium education at the tertiary level, including reading ability, text comprehension, written text production and course grades.

1.4 The Hungarian context

Hungarian students who wish to enter the university as English majors need to pass the advanced level secondary school-leaving exam in English without an additional written or

oral entrance exam. This would theoretically ensure that they all possess a minimum of B2 level English knowledge as defined by the Common European Framework of Reference (Council of Europe, 2001), attested by the school leaving exam and/or a language proficiency exam of the same level (Csernoch & Korponayné, 2007). Many of them start their tertiary education after secondary school, therefore, bring a language proficiency level with them directly from the secondary school. However, in recent years there has been a tendency that students start their first year with rather intermediate (some of them lower intermediate) than advanced English language proficiency and, as a consequence, many of them struggle with their study load. This is not surprising if we consider the variety of students' English language backgrounds and the fact that there is no centralized curriculum and requirement for vocabulary acquisition in foreign languages in Hungarian secondary schools (Doró 2007a). While the 1995 National Core Curriculum (NAT) proposed the receptive knowledge of 1,600 words and the productive knowledge of 400 words for a foreign language studied at school at grade 10 (though it is not clarified what should be meant by receptive and productive knowledge and how to test it), the 2000 and 2003 editions of the National Curriculum do not prescribe any such figures. Csernoch and Korponayné (2005; 2007) analyze the lexical profile of the school leaving exams administered in previous years. They report that both the intermediate and advanced level exams contain a high proportion of very low-frequency words (above the first 10,000 words of the British National Corpus), many of which are not familiar to the students, therefore, cause problems of understanding. They also report that the comparison between the vocabulary of widely used textbooks and that of the exam tasks at the advanced level only provide a 78–90% overlap, meaning that even the partial knowledge of all the vocabulary introduced in textbooks leaves students with predictable difficulty in understanding exam tasks. But what actually the lexical knowledge that students bring with them to the university is and how it changes with time is yet to be discovered.

While we can read about a growing number of reports on the English as a second language (ESL)/English as a foreign language (EFL) vocabulary and writing skills of learners with different L1 backgrounds, very little is known at all about Hungarian learners of English in this respect. Beginning from the 1970s, a number of large-scale studies have been carried out which measure elementary and high school students' knowledge and skills in various subject areas, including foreign languages, but these provide limited vocabulary-related data.

Kádárné (1979), for example, discusses the English results of 9th and 12th graders involved in a representative international study. All subjects started learning English as a 3rd language upon entering high school and demonstrated having the most problems with listening skills and essay writing. Low scores on the essays (given topic with a list of words to include) had been foreseen by the researchers, as the subjects' instructors dedicated little attention to teaching written production and did not have clear ideas of how to include it in their syllabus. Váradi (1980a) reports on the same English written production task administered a few years earlier, in 1971, by carrying out a detailed error analysis of the texts subjects produced, but the aspects of the written production are not discussed in detail. The above mentioned reports on large-scale studies do not give detailed analysis of the lexical richness of the written productions or the lexical components of other tasks, nor do they compare them with other language exam scores or background data collected from the same students. Csapó (2002: 182–191), while discussing the English tasks students completed as part of a large-scale national data collection, notes that writing skills (the task was to write a letter on a given topic) received the lowest scores across vocabulary-related issues. Another large-scale study investigating the writing ability of 10th graders (letter writing about a dream holiday) show similar low scores on an analytical rating scale, especially in the case of students enrolled in vocational schools (Nikolov & Józsa, 2003, 2006).

In recent years a number of small-scale empirical studies have also discussed various vocabulary-related issues. Some of these are placed within a secondary school context, others involve university students as subjects (usually English majors). Barratt and Kontra (1996) target a small section of the L1 and L2 lexicon by comparing the color terms used by speakers of American English, Hungarian and Hungarian EFL university students. The authors analyze the similarities and differences between the color terms used by the different groups and reveal some discrepancies between the usage the subjects report on and what is found in bilingual dictionaries. Kiszely (2006, 2007) discusses secondary school students' writing done in Hungarian and English, while Doró (2007b) compares the lexicon of Hungarian students' essays with those written by American children. Dóczy (2007) also targets the secondary school level within the framework of a longitudinal study involving multiple types of students' lexical knowledge. Lehmann (2003) reports on a small scale study run with university English majors, which compares the subjects' receptive vocabulary using a self-

assessment 50-item test and the lexical frequency profile of a take-home essay, and in which the author finds no clear correlation between the two. Lehmann (2007), involving a larger study population, investigates the vocabulary knowledge of first and second-year students, with the general conclusion that they are lacking the size of lexicon needed for academic studies. A limited number of conference papers with direct focus on vocabulary-related issues in Hungarian universities have also been recently presented (e.g. Lehmann, 2006, 2007; Peckham, 2006, 2007), and a number of ongoing doctoral research is also targeting questions related to vocabulary. Moreover, the author of this dissertation herself has presented some preliminary findings using data subsets of the comprehensive data collection discussed in this dissertation (Doró 2007d, 2007e, forthcoming). However, this is the first time that the multifaceted nature of the vocabulary knowledge of Hungarian English majors is discussed in a comprehensive nature. The present study wishes to bridge the gap between the growing body of international literature on L2 vocabulary research and the need to have empirical data on Hungarian university students' lexical knowledge and use both for research and instructional purposes.

1.5 Specific research aims

As has been stated above, the empirical investigation of the present dissertation wishes to bridge the gap in the literature relative to Hungarian university students' lexical knowledge by assessing it from a number of perspectives. Rather than simply providing a description of the size of their lexicon, the question of the relationship between the various types of knowledge targeted by different tests, the possible reasons behind the different test results, the relationship between standard test scores and the actual vocabulary use in written production tasks will be addressed.

The first major research area targeting the vocabulary knowledge of English majors, including first and third-year students, emerges from the observation of Hungarian university instructors that students who wish to follow academic studies as English majors apply for the university with different levels of overall language proficiency, including vocabulary knowledge. Although it has been a practice at the study site to assess the receptive vocabulary of incoming students on the basis of one test, no information has been systematically gained

on other aspects of their vocabulary knowledge at their point of entry. They have neither been systematically followed up on their progress during their academic career. The present dissertation wishes to go beyond the scope of reporting the vocabulary size of a selected student population on the basis of one single testing method. It rather uses three tests of vocabulary knowledge to capture rich information on different subcategories of lexical knowledge (e.g. form, meaning, frequency, association). All three tests are based on the widely accepted view that frequency is a key aspect of a word, and high-frequent words comprise a fundamental part of a learner's lexicon on which less frequent words should be built.

As has already been pointed out, it is always desirable to assess various vocabulary knowledge types of learners. In real-life testing situations, however, due to efficiency and time constraints, usually a restricted number of tests or subtests are administered. Therefore, it is of both theoretical and practical value to analyze how strongly results on various tests correlate, in other words, to what extent receptive test results can predict the nature of the learners' productive lexical knowledge. Based on the findings related to this second major research area, I hope to be able to extrapolate results obtained on any part of the administered tests or subtest into the remaining knowledge types of learners' lexical proficiency.

The research carried out here aims to assess the lexis of not only incoming students, but also those at the end of their first and third year. These two groups have been selected for the fact that they are at the end of two important periods of their studies as English majors at the study site. Since three separate groups of students will be assessed (incoming-first-year, end-of-the-first-year, and end-of-the-third-year), it is possible to see changes in their results on the basis of apparent-time analysis. It is not the goal of the study to show a gain in students' lexicon over a three-year period, but rather assess what their knowledge is at various stages and what this vocabulary knowledge and use can inform us about possible problems in academic progress.

To balance for the need to gain an insight into the individual changes in participants' lexical knowledge, a small group of first-year students were selected and retested at the end of their first year, accompanied by a structured interview to give possible answers to the reasons behind the great variability of the change over one academic year. The results of this case-study are reported in the final chapter of data analysis and discussion.

Another key research area investigated in this dissertation is motivated by the fact that it would be desirable to know how vocabulary test results can predict academic success. It is a tendency that many students fail their classes or exams, and if motivational problems are not considered, it is more than probable that the lack of sufficient vocabulary and, therefore, inadequate understanding of oral and written input and the scarce quality and quantity of their language output, are key factors in their failure. Those students who encounter great difficulty in general language proficiency classes are also likely to be unsuccessful in content classes. The analysis of the test results is expected to show the minimum lexical competence needed for students at various stages in order to successfully complete their English language studies, and also identify a risk group who do not meet the lexical requirements and should go through an accelerated vocabulary building program.

The three tests of vocabulary knowledge employed in this dissertation have various formats and target various aspects of the complex nature of vocabulary knowledge. The two productive tests claim to measure vocabulary in a similar context to how it is recalled and used. However, it is also of importance to see how vocabulary is actually integrated into a written production task. Therefore, results obtained on various tests of lexical knowledge will be compared with the lexical profile of texts written by students under controlled circumstances. This is done to show whether students with a large vocabulary size integrate a larger proportion of less-frequent words in their essays than those with a limited vocabulary.

1.6 Structure of the dissertation

To conclude this introductory chapter, I shall briefly review the content and structure of this dissertation which can be divided into five parts. Part I includes one chapter, the present one that introduces the dissertation and the motivation behind it.

Part II is made up of four literature review chapters. Chapter 2 discusses issues related to the problematic nature of the definition of the construct *word* and related terms which form the basis of all vocabulary assessment questions. Chapter 3 focuses on what is included in vocabulary knowledge by discussing its various components and by looking at different, but complementary vocabulary knowledge frameworks developed by researchers in the last decade. Chapter 4 then provides some background to vocabulary assessment by discussing

various test formats and types, drawing upon some of the most often cited tests of vocabulary knowledge. Since one of the data collection instruments of this dissertation is a written production task, the chapter then discusses the various ways to assess the lexicon of written texts, including lexical richness and lexical profile. Chapter 5 takes a closer look at the relationship between vocabulary and reading and vocabulary and writing by reviewing some studies directly related to the empirical research of the dissertation.

Part III reports on the empirical investigation involving 342 university students, with detailed chapters on research questions (Chapter 6), data collection, processing and analysis (Chapter 7). The findings and discussion of the results are presented in Part IV, following the order and logic of the research questions (Chapters 8, 9 and 10). An overall discussion of the findings follows in Chapter 11.

Part V, which consists of Chapter 12, presents a review of the issues discussed, the results found, a concluding summary and some practical implications of the discussion to the implementation of test results and instruments into the syllabus.

PART II

LITERATURE REVIEW

Introduction

While in the past vocabulary used to be a neglected area of research within foreign and second language studies, the last few decades have shown an increase in the publications related to the lexicon of English as a second language (ESL) and English as a foreign language (EFL) learners, targeting issues such as the nature of vocabulary knowledge, assessment tools, teaching and learning strategies for both research and pedagogical purposes. While the body of literature related to vocabulary knowledge and use has had the aim of contributing to the understanding of these issues through theoretical discussions, large and small-case empirical investigations, in recent years more and more authors have voiced the problematic nature of interpreting the accumulated knowledge. This is due, on the one hand, to the massive size of the literature, and, on the other hand, to its fragmented nature of data and the variety of terminological use that make it difficult to gain an overall and universal understanding of key issues.

David Singleton, in his book on the second language mental lexicon, almost a decade ago wrote:

the task of giving a straightforward account of what research tells us about L2 lexical acquisition and processing is more difficult than ever. Not only has the recent vast increase in the amount and the range of L2 research focusing on lexical matters as traditionally defined rendered all hope of providing anything approaching an exhaustive survey of such research utterly vain, but the very conception of what is to be included under heading of lexicon has expanded to the point where almost any L2 acquisition/processing research can be seen relevant to the L2 mental lexicon (Singleton, 1999: xiii–ix).

What he chose as a strategy is to select from what he calls “lexical research avenues” the more relevant ones for his discussion. A similar strategy will be employed in this work, since it would be well beyond the scope of a dissertation and even that of single volumes to provide a detailed critical analysis of this ever growing body of literature. Instead, I will discuss the

main theoretical and practical issues to shed light on the complexity of problems that arise when assessing what can be called vocabulary knowledge.

Issues discussed in the following literature review include the concept and definition of the construct *word* so often used that it seems to be universally understood. Next, the problems of the boundaries of lexical units and their multiple meanings will be explored, as well as what is meant by the knowledge of a word in terms of various categories that can be identified such as form or meaning. Some of the vocabulary knowledge frameworks, which also emphasize the multifaceted nature of vocabulary knowledge, will be further reported on. In the last chapter of this literature review, assessment-related questions will be investigated when discussing various test formats, the underlying mental processes, the knowledge types they assess and the implication of the choices made between various instruments.

My purpose here is to explore what makes this body of research problematic to interpret without suggesting conclusive solutions to the questions raised. The literature that I review here intends to, on the one hand, provide a basic theoretical framework necessary to understand the complexity of issues related to the assessment methodology, and, on the other hand, examine some of the most recent empirical data reporting on the use of vocabulary tests and assessment techniques of the vocabulary in written texts. The scope of this part of the dissertation has been limited for a number of reasons. First, a balanced proportion between the discussion of the previously published material and the results of the empirical investigation carried out by myself had to be found. Second, I intend to review the most relevant issues in the literature as a whole, to illustrate particular problems, rather than detailing any one aspect of the issues or critique any one empirical study in its entirety. Third, the discussion of the literature will mainly, but not exclusively, focus on material published within the last 10–15 years, with some reference to earlier works that have become the basic readings of certain sub-fields of vocabulary studies. Fourth, I will not heavily rely on research about first language vocabulary acquisition or the lexical competence of early, balanced bilinguals, nor will provide a detailed review of vocabulary acquisition strategies and teaching methods. These issues will be brought in when I believe that they are necessary for the understanding of the questions discussed in connection with vocabulary assessment. Finally, in order to balance the seemingly missing links to certain broader issues of language testing or construct definitions, I will try to indicate literature that focuses on these aspects for further reference.

CHAPTER TWO

TERMINOLOGICAL CLARIFICATION OF THE CONSTRUCT *WORD*

2.1 Introduction

Before discussing the literature on lexical knowledge and assessment, it is inevitable to address the complex issue of what the lexicon consists of. The present chapter starts with an overview of how the construct *word* and related terms have been used in the literature and what aspects have shaped their usage. This chapter first of all attempts to summarize current perspectives on terms related to the lexicon, touching on some of the theoretical issues that emerge in this context. It then turns to elaborate on the types of multi-word units in English and their role in the lexicon. The chapter concludes with the discussion of multiple meanings of many lexical items.

2.2 What is a word?

The need for clarification is not a recent one, as Anthony (1975) already voiced the problematic nature of word meaning, the distinction between lexicon and vocabulary both from theoretical and pedagogical aspects. The question raised above may seem to be an easy one to answer by identifying a word in orthographical terms as a string of letters. This view of a word has been shaped and strengthened in two fundamental aspects: first, the non-specialists' view that language consists of words clearly separated by orthographical boundaries in written texts (Singleton, 1999); second, the growing influence of text- and corpus-based research on educational theories and practices (Gardner, 2007). Both aspects bring a number of problems with them. The view of a word as a sequence of letters separated by a blank space does not work for languages that operate with writing systems which do not clearly mark word boundaries. Furthermore, this idea of word boundaries is not easily transferable for oral language, where usually there are no clear pauses between each word. Although it is clear that identifying words on the basis of orthographical or phonological criteria is not without problems, simple text-based vocabulary research tools use word

boundaries as the basis for quick word counting of texts. Such research tools may use two alternative terms for *word*, the term *token* for lexical units separated by blank spaces and *type* for different word forms within a text. Therefore, the sentence *The cat is chasing the mouse* consists of six tokens, but only five word types. The less repetition a text contains and the greater variety of different word forms it operates, the higher the text's type/token ratio will be.

The previous introduction of the term *type* and *token* for the simple purpose of counting lexical units in text- and corpus-based research leads us to the discussion of the morphology-based definition of words. Gardner (2007) notes that many corpus linguists, while defining the morphological relationships between words, draw upon Francis and Kučera's definition of *lemma*, which they identify as "a set of lexical forms having the same stem and belonging to the same major word class, differing only in inflection and/or spelling" (Francis & Kučera, 1982: 1). This definition implies that lemma of the verb *go* includes *go*, *goes*, *going* and also the irregular verb form *went*. Gardner (2007), however, notes that irregular forms make the immediate, clear relationship between members of a lemma set opaque and psycholinguistically less valid. They also pose different learning problems. She adds that some corpus linguists, in addition to the same grammatical class requirements, while defining lemmas, include the same meaning requirement and use *lemma* as a synonym of *lexeme*. An example of the meaning-based definition of lexeme is shown in the following definition: "a group of word forms that share the same basic meaning (apart from that associated with the inflections that distinguish them) and belonging to the same word class" (Biber *et al.*, 1999: 54).

In research this semantically-based definition of criteria for a word is again not without problems unless researchers use both a morphologically and a semantically tagged corpus or a sophisticated collocational analysis program. As these are now only in their developmental phase, they are not available for everyday use in research and even less in pedagogical contexts. Without their use, however, frequency counts of words in texts often result in distorted or overgeneralized data. I shall come back to this issue in the next two sections when discussing multiple meaning and multi-word units.

The morphological aspect of defining words is crucial when trying to establish *word families* by grouping base forms or stems together with their inflected forms and derivatives.

Gardner (2007) uses the expression ‘transparent’ derivatives, while Read (2000) refers to ‘closely related’ derivatives. The two views are not exactly the same as Gardner incorporates the definition of Nagy *et al.* (1993) who includes in close relatedness both graphical similarity and meaning redundancy.

Grouping words into word families on the basis of the above mentioned criteria, however, does not necessarily mean that the knowledge of one member of the word family brings with itself the knowledge of other members, nor that learners are exposed to the stem first and inflected and derivated forms afterwards. Jiang (2000), for example, notes that while institutional learning favors the presentation of stems before other members of the given word family, reading authentic texts does not guarantee this order of exposure or learning. It follows that it is possible, especially without specific morphological training, that L2 learners are unable to identify or simply unaware of the stem and its affixes. Also, as Nagy *et al.* (1993) note, researchers often treat derivational prefixes and suffixes together, under the broader term *derivational affixes* without considering that the two pose different learning problems. While derivational prefixes tend to be paraphrasable (e.g. *non-*, *un-*, *in-* mean ‘not’) and keep word classes, derivational suffixes such as *-ness* or *-able* do change word classes and are difficult to conceptualize. The degree of transparency and difficulty of understanding the above mentioned morphological variables such as inflectional and derivational affixes are best summarized by Bauer and Nation (1993, cited in Schmitt, 2000) in the seven levels reported in Table 2.1.

This seven-level framework is based on research related to learners’ recognition and understanding of morphological variables in written texts. The authors claim that their findings can help research on learners’ vocabulary size, age-related morphological development and lexical storage. Gardner (2007), however, rightly points out a number of weak elements in the framework, the main one being the repetition of various affixes at different levels. She also voices the difficulty of the application of the framework on large corpora.

Table 2.1 Difficulty of inflectional and derivational suffixes summarized by Bauer and Nation (1993)

<p>Level 1: <i>Each form is a different word.</i> Learners have no concept of morphological relationships between words. Each derivate is considered a separate type.</p>
<p>Level 2: <i>Inflectional suffixes.</i> Base forms and their inflections are considered members of the same word family. Affixes in this category include the plural, third person singular present tense, past participle, <i>-ing</i>, comparative, superlative and possessive.</p>
<p>Level 3: <i>The most frequent and regular derivational affixes.</i> Affixes include <i>-able, -er, -less, -ly, -ness, -th, -y, non-</i> and <i>un-</i>.</p>
<p>Level 4: <i>Frequent, orthographically regular affixes.</i> Affixes include <i>-al, -ation, -ess, -ful, -ism, -sit, -ity, -ize, -ment, -ous</i> and <i>in-</i>, with restricted uses.</p>
<p>Level 5: <i>Regular but infrequent affixes.</i> Affixes include <i>-age, -al, -an, -ance, -ant, -ary, -atory, -dom, -eer, -en, -ence, -ent, -ery, -ese, -esque, -ette, -hood, -ian, -ite, -let, -ly, -most, -ory, -ship, -ward, -ways, -wise, anti-, ante-, arch-, bi-, ciurcum-, counter-, en-, er-, ex-, fore-, hyper-, inter-, mod-, mis-, neo-, post-, pro-, semi-, sub-</i> and <i>un-</i> Relatively easy to segment, but they do not add much to the number of words that is understood.</p>
<p>Level 6: <i>Frequent but irregular affixes.</i> Affixes include <i>-able, -ee, -ic, -ify, -ion, -ist, -ition, -ive, -th, -y, pre-</i> and <i>re-</i>. They are difficult to segment. Some already listed at earlier levels are here in their less transparent cases.</p>
<p>Level 7: <i>Classical roots and affixes.</i> Many frequent English prefixes are part of this level, examples are <i>com-, de-, dis-, ex-</i> and <i>sub-</i>. There should be explicit teaching of these affixes.</p>

A number of studies have been published on learners' awareness of affixes and their role in language proficiency. Schmitt and Meara (1997), for example, report that adult Japanese learners of English are more aware of English inflectional verb suffixes than derivational verb suffixes. Research indicates that the explicit attention to form and the direct teaching of the relationship between members of a word family facilitate morphological awareness, which brings with itself better reading skills (Carlo *et al.*, 2004; Schmitt & Zimmerman, 2002).

As I have reviewed in this section, the understanding of the basic term *word* in vocabulary research and leaning is not a straightforward task. In this dissertation I will use the term *word* to refer to orthographical units (unless the cited literature uses it in par with word family) and *word family* to refer to the base form and inflected and derivational forms as defined by the tests and research tools that I have been using. I will do this with the awareness

of weaknesses and limitations that the use of basic terminology may impose. In cases when the distinction between word forms and word families or multi-word units is not crucial, I will adopt the wider terms *lexical unit*, *vocabulary* or *lexicon*.

2.3 Multi-word units

The short overview of the construct of *word* on the basis of orthographical, phonological and morphological criteria has shown that it is indeed not an easy task to know what the term *word* refers to in the variety of academic contexts it is used in. But whether we cite it in the meaning of word stem, derivatives or inflected forms, as one member or a word family or as a synonym of word family, it is still usually one single word form that is being referred to. However, it is important to note that the English lexicon is rich in multi-word units. According to Moon (1997) a *multi-word item* is “a vocabulary item which consists of a sequence of two or more words (a word being simply an orthographic unit). This sequence of words semantically and/or syntactically forms a meaningful and inseparable unit. Multi-word items are the result of lexical (and semantic) processes of fossilization and word-formation, rather than the result of the operation of grammatical rules” (Moon, 1997: 43).³

Moon (1997) identifies three criteria which make multi-word units different from other, usually grammatically formed, strings of words. First, multi-word units are institutionalized in a language. Second, they are fixed, do not change in form and their elements cannot be substituted with a synonym. Finally, they are non-compositional as they cannot be interpreted on a word-by-word basis. The author distinguishes the following five types of multiword items: compounds, phrasal verbs, idioms, fixed phrases and prefabs, calling our attention to the fact that there are many overlaps between these categories, and also that there are a variety of other terms used in the literature to refer to multi-word items.⁴

³ In the literature many multi-word items are discussed under the wide term *collocation*. Collocation refers to the co-occurrence of two or more words, but even in its case there is great variability between its wider and most restricted uses. For detail on some of the most recent discussions on the matter see e.g. Howarth (1998), Nesselhauf (2003), Wray (2000, 2002), and Doró (2007c).

⁴ Definitions are diverse and varied and are based on a range of criteria. Two other, frequently used terms need to be mentioned here. *Multi-word expressions* are defined by Sag *et al.* (2002: 2) as “idiosyncratic interpretations that cross word boundaries (or spaces).” Many linguists adopt the term *formulaic sequence* which is, in Wray’s (2002: 9) definition “a sequence, continuous or discontinuous, of words or other elements, which is or appears to be prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar”.

a) *Compounds* are the largest and least stable group. They differ from single-word units only by being written as two or more orthographic words. Their instability is due to the fact that many times they are multi-word units on the basis of spelling conventions. Variability in hyphenation or lexicalization may change the status of compounds. An alteration in spelling conventions with time or between varieties of the same language may turn compounds into single-word units. This, again, poses problems when treating written language data or using computer-based research tools for even as simple steps as counting the number of words in a text. Compounding is a very productive and relatively easy process in English, therefore, learners often use it without being familiar with institutionalized spelling conventions.

b) *Phrasal verbs* are combination of verbs and adverbial or prepositional particles. These verbs are usually very frequent, monosyllabic words. Although some phrasal verbs are highly compositional in meaning (*write down*), many phrasal verbs are non-transparent, making their acquisition and use problematic for learners.

c) *Idioms* by Moon's definition are "multi-word items which are not the sum of their parts: they have holistic meaning which cannot be retrieved for the individual meaning of the component words" (Moon, 1997: 46). They are very much fixed expression (*kick the bucket*) and many have metaphorical or etymological basis.

d) *Fixed phrases* are other usually strongly institutionalized multi-word units. Many of them are highly frequent (*of course*) and have strong pragmatic functions, like greetings (*good morning*). Similes (*white as a sheet*) and proverbs (*early come, early go*) can also be included in this type.

e) *Prefabs* are "preconstructed phrases, phraseological chunks, stereotypes, collocations, or semi-fixed strings which are tied to discoursal situations and which form structuring devices" (Moon, 1997: 47). They are also called *lexicalized sentence stems* or *ready-made complex units*. It is easy to understand their surface meaning, and many of them are frequently used, their difficulty arises from knowing their right discoursal usage. Examples are *this leads us to, let me point out, to begin with*.

Although the majority of the assessment tools used in the empirical investigation treat vocabulary at the word level, it was needed to call attention to the importance of multi-word units. Both in the tests of productive vocabulary employed in the study and the vocabulary

used in reading or writing involve the co-occurrence of words. Multi-word units will be referred to in more detail when discussing the answers given in the association task, many of which use the stimulus as a part of a multi-word unit and the answer as another part of the same unit.⁵

2.4 One word – multiple meanings

Another issue that cannot be left without some consideration when discussing the term *word* and the various problems it brings in vocabulary instruction and research is the notion that many English words with the same or similar form have meaning variations. The assessment of vocabulary may be difficult due to two aspects, both of them related to the fact that there is often no on-to-one relationship between word form and meaning. *Polysemes* are variants of the meaning of a word they are related to (*She broke her leg, They broke the record*), while *homonyms* or *homographs* are words that are spelled the same way, but do not have clear meaning relationship. A major problem with polysemes from an assessment point of view is which meaning to take into consideration. Knowing one aspect of a word does not imply the familiarity with other meanings. Benssoussan and Laufer (1984, cited in Laufer 1997b) report that in their study of lexical guessing learners had most problems with words with multiple meanings and this category of words induced the greatest number of errors. Learners who know one meaning of a word tend not to search for other meanings even though the familiar meaning variation does not match the given context. Homonymy can also pose problems in corpus-based vocabulary studies. One may argue that homonyms are easily identifiable in context, however, rigorous grammatical and semantic tagging is needed in order to rule out errors in calculation of lemmas. This is a process which is not always readily available and which is rather time-consuming. Ambiguity in meaning is more relevant in high-frequency words, as these tend to have more meanings; therefore, the problem of form and meaning should be considered even in the case of low-proficiency students.

This issue of words that carry more than one meaning is not a marginal one, as counts of dictionary entries have indicated that over 40% of English words have multiple meanings (Nagy, 1997), averaging 2.3 definitions each. The author also concludes that words with a

⁵ For more on the knowledge and acquisition of formulaic sequences refer to e.g. Schmitt *et al.* (2004) and Dörnyei *et al.* (2004).

single meaning are either compounds or derivatives, or associated with a narrow and specialized domain. This issue will be directly addressed when discussing the assessment tools used for the empirical investigation of this dissertation.

2. 5 Concluding summary

In this chapter I elaborated on the importance of the word-concept in vocabulary theory and its significance in vocabulary-related research. While it may seem to be a universally accepted concept what researchers mean while referring to *words*, the definition is not so clear-cut. What we mean by lexical unit has a direct bearing on research design and data analysis, just as much as the fact that the lexicon is made up not only of single-word units, but many multi-word units of various types. One more assessment-related aspect of words was also briefly discussed, namely homonyms and polysemes, both undermining the one-to-one relationship between word forms and their meanings.

CHAPTER THREE

VOCABULARY KNOWLEDGE

3.1 Introduction

Some recent review articles have voiced strong criticism on the nature of terminological use not only related to the construct *word* and other terms discussed in the previous chapter, but also those of vocabulary, vocabulary knowledge and the underlying constructs. One piece of criticism in Read and Chapelle (2001) voices the following points:

Research on second language vocabulary development has been thriving for the last 10 years or more, as attested by numerous articles in applied linguistics journals, anthologies and single-authored volumes. An observation that emerges from a review of this literature is the ill-defined nature of vocabulary as a construct, in the sense that different authors appear to approach this from different perspectives, making a variety of – often implicit – assumptions about the nature and scope of the lexical dimension of learners' language (Read & Chapelle, 2001: 1).

It would be too strong to conclude that this ill-defined nature of terminological use is universally deliberate, yet, what Read and Chapelle note rightly points to the fact that the heterogeneity of terminological use and understanding makes it difficult to boil down the accumulated knowledge of vocabulary-related issues into a clear summary. Even after terminological clarifications, it is still not a simple task to define what lexical knowledge involves. In order to use units of learners' lexicon, one needs to have various types of information on the lexical units (not limited to their form and meaning), and to be able to retrieve the needed one and use it properly.

In this chapter I will introduce two possible approaches to the description of this knowledge. Some researchers in the past few years have attempted to form knowledge frameworks, including two, three or four dimensions of vocabulary knowledge such as precise or partial knowledge, receptive–productive distinction, quantity or quality of knowledge, and lexical organization. These have developed from the second, more traditional way of seeing lexical knowledge, which is listing all the inter-related sub-knowledge types or aspects involved, such as morphological and syntactic knowledge, form or meaning. Of course, all

these dimensions or aspects are appropriate for trying to provide a theoretical framework for vocabulary acquisition and use, but they are not clear-cut and the various aspects or dimensions are intertwined in acquisition, storage, retrieval and use.

3.2 Vocabulary knowledge frameworks

Richards (1976) was among the first researchers to enumerate the different types of vocabulary knowledge in his ‘vocabulary knowledge framework’, reported in Table 3.1.

Table 3.1 Richards’ vocabulary knowledge framework (1976)

•Frequency	knowledge about the degree of probability of meeting the word in speech or writing
•Register	knowledge about limitation on use according to function and situation
•Position	knowledge about syntactic behavior associated with the word
•Form	knowledge about underlying form and derivatives
•Association	knowledge about network of association between the given word and other words in the language
•Meaning-concept	knowledge about the semantic value of a word

From his work the view of listing various aspects of word knowledge (as will be discussed in Section 3.2) and also the need to form more compact frameworks have emerged. Starting from the 1970s, researchers have introduced various but complementary vocabulary knowledge frameworks (Richards, 1976; Read, 1993, 2000; Nation, 1990, 2001; Wesche & Paribakht, 1996; Chapelle, 1998; Henriksen, 1999; Qian, 1999, 2002; Qian & Schedl, 2004). Although these frameworks change in the number of dimensions they propose, they all highlight the multidimensional nature of vocabulary knowledge. Read (1988), Wesche and Paribakht (1996) and Qian (1999) agree that vocabulary knowledge should include at least the following two dimensions: breadth or size of the lexicon of which learners possess some kind of knowledge and depth or quality, the degree to which a learner knows a word. Qian (1999) notes that the dimension of depth should include the subcategories of meaning, register, frequency of use, morphological, syntactic and collocational properties, as well as pronunciation and spelling.

Henriksen (1999) proposes a distinction among the following three interrelated dimensions:

- a) partial vs. precise knowledge
- b) depth vs. breadth of knowledge
- c) receptive knowledge vs. productive use

The partial–precise dimension refers to the concept that vocabulary items are known to a certain degree, but not all aspects are fully mastered in every case (Haastrup & Henriksen, 1998). Henriksen’s dimension of depth should include knowledge of meaning, but also paradigmatic sense relations to other words in the lexicon (antonyms, synonyms, hyponyms, gradation) and syntagmatic relationships (collocations).

The receptive–productive distinction needs further discussion, as there is much less consensus in their use than what is seen in the case of the previous two dimensions. The terms *receptive* and *productive* vocabulary have been used in a variety of disciplines from language teaching to psychology. Waring (1997b) summarizes the variety of different ways the terms are generally discussed:

- receptive and productive vocabulary processes
- receptive and productive vocabulary abilities
- receptive and productive vocabulary skills
- receptive and productive vocabulary products

By *vocabulary processes* we refer to the mainly subconscious mental processes involved in the recognition, recall, retrieval, comprehension, and production of vocabulary items. *Vocabulary abilities* mean the understanding and control of language input and output. The use of these vocabulary abilities leads to the constant reorganization and increase of the learners’ vocabulary size. *Receptive and productive skills* are terms traditionally used in foreign language teaching to distinguish between receptive skills (listening and reading) and productive skills (speaking and writing). *Receptive and productive products* can be viewed through vocabulary tasks representing the skills of speaking and writing (Waring 1997b).

While reviewing vocabulary studies addressing the question of reception and production, Melka (1997) points out that the two terms have been widely used but rarely defined. Researchers have adopted the terms *active vocabulary*, *productive vocabulary*, *production* to mean one notion and *receptive vocabulary*, *passive vocabulary* and *reception* (the terms often used as synonyms) to refer to the other one. Melka concludes that the dichotomy of reception vs. production has been based on testing techniques or the intuition of speakers. Tests of lexical knowledge are often labeled as receptive/passive or productive/active. These tests clearly put different types of demand on test-takers who are asked to carry out various types of tasks, but it is not always clear what aspects of vocabulary knowledge are really tested. One attempt to overcome these difficulties of distinction has been the development of tests that require learners to report on various types of their lexical knowledge (for detail on testing vocabulary, see Chapter 4). The intuition of speakers mentioned by Melka refers to the fact that people often see passive vocabulary as words met before but not used in speech or writing, opposed to the active vocabulary well-known and often used by learners. This distinction, however, is very simplistic and does not cover the complex notions of reception and production outlined above. A similar simplistic view of passive vs. active vocabulary is reported by Hungarian university students (clearly influenced by terms often used in Hungarian language instruction as *aktív/passzív*) in Doró (2007a). Students refer to active vocabulary as the one “I can use in context”, “I use in speech”, “I know well”, “I do not have to check in the dictionary” or “the simple words I use every day”. They identify passive vocabulary as the one “I have learned from word lists”, “I understand while reading”, “I know less well”, “I need to check in the dictionary” or the “words rarely used”. From this short list it is clear that learner’s distinction between receptive and productive is even more varied than that of researchers and do not necessarily match the criteria found in the literature.

The problematic nature of determining the difference between receptive and productive knowledge is most thoroughly discussed in Waring (1997b). He reviews ten studies directly addressing this question. Let us here report some of the conclusions he draws while reviewing large-scale studies of the 1930’s and 40’s in a North American context. These studies are Morgan and Oberdeck (1930), Stalnaker and Kurath (1835) and Smith and Campbell (1942). These first large-scale studies use multiple-choice format as receptive tests

(also defined as recognition) and translation or word/sentence completion format as productive tests (also defined as recall). This clearly sets a tradition to associate these test formats with the knowledge type they are expected to test. Waring voices a number of problems with these studies: a) there is no solid conceptualization what the tests are actually attempting to measure; b) the focus is on test characteristics and not on the ability underlying the tested knowledge; c) the conclusion that the two types of tests employed are interchangeable for all practical purposes is questionable, as they are not testing the same knowledge source; and d) a high level of correlation between test results does not necessarily mean that the same underlying competence is being tested, also, knowledge types are qualitatively different, so as the underlying mental processes. These studies, although they date back over 60 years, still show a strong influence on the thinking about the relationship between receptive and productive knowledge and its assessment.

Of more recent studies discussing the relationship between the two knowledge types some employ the receptive and productive versions of the Vocabulary Levels Tests which are also used for data collection for this dissertation (see Appendices A and B). Therefore, these studies merit a more detailed discussion at this point. The receptive version (VLT, for detailed discussion of the test refer to Chapters 4 and 7) is a six option matching test and the productive version (PVL, also see Chapters 4 and 7) is a cued recall test with a sentence completion task. The test items are presented at five frequency levels in both tests, posing growing difficulty to the test-takers. Waring (1997a), testing Japanese university students, reports that the two tests confirm the gap between receptive and productive knowledge, but the difference between receptive and productive size is not a fixed percentage, rather a moving score depending on the proficiency level of the students and the frequency of words tested. He also calls attention to the fact that results must be affected by test sensitivity, the receptive version of the test being much more sensitive to partial knowledge. Also, test format and scoring method influence results that are then labeled as receptive and productive knowledge. Laufer (1998) using data from two groups of high school students in Israel concludes that “we can see that the growth of the two kinds of vocabulary knowledge is not identical. In one year the learners added about 1,600 word families to their passive vocabulary and 850 word families to controlled active lexis” (Laufer, 1998: 262). The conclusions of the two studies seem to closely match; however, a wide range of standard deviations calls our

attention to the fact that there is a great variability in the results, and any average will hide this aspect of the data. Also, Laufer's two groups used for comparison clearly differ only in terms of one extra academic year, but no other information, for instance, motivation, extra-curricular activities, language learning background, is discussed. In contrast, the three compared groups in Waring (1997a) are divided on the basis of English proficiency. The conclusion that can be drawn from these two recent studies is that there is a diverging gap between receptive and productive vocabulary. This means that a high-frequency item known receptively is more likely to be also known productively than a low-frequency word. This is the same message that was formulated 70 years earlier in the Morgan and Oberdeck (1930) study. More of this relationship will be discussed in Chapter 8 in light of the empirical data collected for this dissertation.

Seeing the difficult interpretation of the receptive–productive dichotomy, Melka (1997) proposes to “replace the idea of a gap between R [reception] and P [production] with other more realistic notions: familiarity and degrees or continuum of knowledge. This idea of a continuum better explains the fact that the boundaries between R and P are not fixed, but vary according to diverse linguistic and pragmatic factors” (Melka, 1997: 101). Many other researchers also see reception and production as a continuum. Meara (1996), however, opposes this idea and sees the distinction between the two as a result of different types of association between words. Passive vocabulary is seen by him as items which can be activated only by external stimuli, opposed to the active vocabulary of a learner which has ingoing and outgoing links with other words. His view has been criticized for being strongly association driven (Nation, 2001).

Whether reception and production are seen as two ends of a knowledge scale or as two points of a continuum, it is clear that knowing a lexical item has different aspects and degrees. When estimating one aspect of a learner's vocabulary, e.g. the receptive aspect, by one testing method, little will be known about other aspects of the same lexical item, for example, whether that item would be known in a productive test, and whether it is readily available and used during oral or written production. It is much easier to integrate new items into our receptive vocabulary and later recognize them when they occur again, than to use them actively in our own oral or written production (Laufer, 2005). These issues will be directly explored in Chapters 8 and 9 when receptive and productive test results are correlated and

when measured vocabulary knowledge are compared to the lexical items used in written production tasks.

In connection with this issue of the lexicon used for conveying written message, it is important to briefly review one more framework. Qian (2002, cited in Qian & Schedle, 2004: 30), drawing on earlier frameworks of vocabulary knowledge (Chapelle, 1994, 1998; Qian, 1999; Henriksen, 1999; Nation 2001), proposes a model consisting of four closely connected elements:

- a) vocabulary size – number of words of which the learner has mastered at least a superficial knowledge of meaning;
- b) depth of knowledge – knowledge of graphic, phonemic, morphemic, syntactic, semantic, collocational properties, frequency and register;
- c) lexical organization – storage and connections of the word in the mental lexicon;
- d) automaticity of receptive and productive knowledge – fundamental processes through which access to word knowledge is gained.

Of these four dimensions, the last two are aspects that are already present in Chapelle's (1994, 1998) framework, but not in that of Henriksen (1999) discussed above. Lexical organization refers to the nature of the storage of both words and morphemes in the mental lexicon. These elements are connected to each other on the basis of semantic or phonological properties. Chapelle (1994) notes, while reviewing earlier literature, that during change in the mental lexicon both reorganization and reanalysis take place. There has been a debate among researchers in recent years about how the mental lexicon is organized. This is a challenging question as it is possible to test links in the mental lexicon through the use of association tasks, which give information about the various types of links between lexical units, but it is impossible at this point to have conclusive answers even with the help of magnetic resonance images of the brain on how the lexicon is built⁶.

Automaticity, on the other hand, refers to the “phonological and orthographic encoding and decoding, access to structural and semantic features from the mental lexicon, lexical-semantic integration and representation, and morphological parsing and composing”

⁶ For more information on the mental lexicon and its organization, refer to e.g. Singleton (1999).

(Qian & Schedle, 2004: 30). These are processes that are parallelly activated during language use, but are difficult to assess separately in written measurement. These issues will be further treated while discussing lexical assessment instruments in the next chapter.

3.3 Aspects of vocabulary knowledge

Lexical knowledge is often referred to as a combination of various areas of sub-knowledge. Nation (2001: 27) revised Richards' model adding the receptive–productive distinction. He summarizes the aspect of vocabulary knowledge in forms of questions in the following model. Table 3.2 is reproduced in full in order to show how the various components are integrated to form a whole according to Nation's interpretation.

Table 3.2 Nation's word knowledge model (2001: 27)

		Receptive knowledge	Productive knowledge
Form	spoken	What does the word sound like?	How is the word pronounced?
	written	What does the word look like?	How is the word written and spelled?
	word parts	What parts are recognized in this word?	What word parts are needed to express the meaning?
Meaning	form and meaning	What meaning does this form signal?	What word form can be used to express this meaning?
	concepts and referents	What is included in the concept?	What items can the concept refer to?
	associations	What other word does this make us think of?	What other word could we use instead of this one?
Use	grammatical functions	In what patterns does the word occur?	In what patterns could we use this word?
	collocations	What words or types of words occur with this one?	What words or types of words must we use with this one?
	constrains on use (register, frequency)	Where, when, and how often would we expect to meet this word?	Where, when, and how often can we use this word?

Nation uses the receptive–productive distinction to separate the different knowledge sub-skills necessary to recognize or recall a lexical item. He makes a clear distinction between the two, although he is aware of the problematic use of the terminology when he notes that “like most terminology receptive and productive are not completely suitable because there are productive features in the receptive skills – when listening we produce meaning” (Nation,

2001: 24) In the following paragraphs the various aspects listed in Nation's model will be discussed in more detail.

Although for many specialists and non-specialist word meaning is the primary aspect of word knowledge, the first category in Nation's 2001 model is form, both oral and written. What Schmitt (2000: 45) claims is that "in some ways these can be considered among the most essential of the different kinds of word knowledge, because without the ability to recognize or produce a word, any other kind of knowledge is virtually useless". As has been discussed in the previous chapter, the same word forms may have multiple meanings. As a further difficulty imposed by form is the broad correspondence between the orthographical and phonological form of the same words in English. This implies that the oral form of the word does not necessarily bring with itself the immediate knowledge of the written form, or vice versa.

Vocabulary learning is often seen as the simple memorization of the meaning of the given lexical item, but acquiring word meaning is a complex process. Henriksen (1999) adopts the term *semanticization* instead of the more generic terms *acquiring* or *getting* word meanings, as she stresses the ongoing nature of semantic development, including the building of semantic relationships between words and building semantic networks in the mental lexicon. Aitchison (1994) also underlines the complexity of the process when defining three tasks involved in acquiring word meaning: a) *labeling* which means creating a link between concept, sign and referent), b) *packaging* which involves the process of grouping words together under similar concepts, and c) *network building* which refers to processes of categorization, abstraction, generalization and also the development of the understanding of hierarchical meaning relationships. According to this view, the acquisition of new lexical items or new meanings of the existing ones, involves two processes. Learners add new elements to the lexical store by labeling and packaging, and, at the same time, they reorder and change the existing links between items by network building (Henriksen, 1999)

The third aspect of vocabulary knowledge listed by Nation (2001) is use. This includes the knowledge of the grammatical functions, the collocations, the frequency and registers of the lexical items. It needs to be pointed out, though, that the knowledge of these aspects is only at the competence level and what is actually produced at the performance level is something different. Knowledge of how to use lexical items is not necessarily turned into

actual use in oral or written production task. Frequency is another issue that merits to be discussed in terms of vocabulary knowledge and which will be further explored in the case of knowledge testing (see Chapter 4). Schmitt and Dunham (1999), while reviewing earlier research on the role of frequency in vocabulary knowledge and use, note that frequency of occurrence has been shown to affect the time needed for lexical retrieval, fixation time during reading, word familiarity and learning. However, they felt the need to link information on absolute word frequency to the intuition that native and non-native speakers have of the frequency of certain words. They concluded that learners of English and native speakers of the same language proficiency rate the absolute frequencies of words in a very similar way, assuming that non-natives know all the words to be rated. However, they also called attention to the fact that it is extremely difficult to arrive at definite answers related to intuitions of learners due to the degree of the knowledge of the word pairs tested in similar research. McCrostie (2007) carried out a similar intuition research with two groups of native speakers: professional EFL instructors and undergraduate students. The study concluded that instructors did not have significantly better intuition results than undergraduates. Furthermore, both groups had difficulty judging the frequency of words in the middle frequency range, while better results were obtained in the case of high and low-frequency words. These two studies call attention to the need for cautiously taking for granted the ability to judge the frequency of words. To these findings one more aspect needs to be added. What McCrostie calls absolute frequency is a concept that is difficult to interpret in the light of the number of frequency lists that exist⁷ and in how these differ from the frequency of words in classroom language.

In summary, we can conclude that the areas of vocabulary knowledge listed by Nation (2001) imply that the knowledge of vocabulary is not a one-factor concept. It includes many types of sub-knowledge of which not necessarily all are known to the same degree. Furthermore, receptive knowledge may not directly lead to the productive knowledge of the same area, and productive knowledge may not necessarily be used in productive speaking or writing tasks.

⁷ See a review of the issue in terms of assessment in Chapter 4.

3.4 Concluding summary

This chapter has provided an introduction to the key issues of the theory of vocabulary knowledge. But, as Schmitt and McCarthy (1997) point out, while normally theoretical models inform test design, in the case of vocabulary research testing has a longer history than theoretical modeling. Therefore, they conclude the following:

test designers have long realized the difficulties in measuring everything a person knows about a word, so they have usually settled for measuring knowledge of a single meaning. This solution has gone on to influence general thinking; many people still believe that if you know a word's meaning (even if it is only one of several), you 'know' that word (Schmitt & McCarthy, 1997: 325).

Attempts have been made by researchers to include a number of knowledge dimensions into vocabulary testing and continuously improve test designs to capture as much of the vocabulary knowledge and use and the underlying processes as possible. With these considerations in mind, I now turn to the discussion of vocabulary testing.

CHAPTER FOUR

ASSESSING LEARNERS' VOCABULARY

4.1 Introduction

As has been pointed out in the previous chapters, lexical competence is viewed, by both researchers and language instructors, as a central issue in language learning; therefore, tests of vocabulary knowledge (most often referred to as vocabulary tests) are being widely used in a variety of contexts. The discussion in the literature review so far has shown that the construct of vocabulary knowledge is a complex one. Knowing a word can range from the partial recognition of its form or meaning to its fluent and pragmatically appropriate use in oral and written discourse. Can the assessment of vocabulary knowledge account for such variability in the possible interpretation of vocabulary knowledge? In order to answer this question, the following chapter will start out discussing the design options for measuring vocabulary proposed by Read (2000) in a three-dimension format. His framework includes both the tendency of assessing vocabulary knowledge as a separate component (by counting, classifying individual word forms), without reference to the context or the grammatical or discourse functions of the selected words, and the other option of assessing vocabulary as part of a larger construct. His three-dimension framework is then placed within the larger framework proposed by Read and Chapelle (2001) who discuss all issues that need to be taken into consideration while designing or using a vocabulary testing instrument. In the second part of this chapter an introduction is given to those test formats and vocabulary tests which are most often cited in the literature and which could be taken as a possible instrument for the present investigation. Reference is made to both the advantages and possible drawbacks of these measures. Finally, in the third part of this chapter assessment methods of vocabulary use in learners' written text production are discussed, touching also on issues related to corpus-based vocabulary studies.

4.2 Design of vocabulary measures

In testing vocabulary, similarly to testing other areas of language knowledge, the criteria of reliability, validity, practicability and purpose need to be considered (Bachman, 2000; Nation, 2001). Read (2000) proposes the following three dimensions to determine the nature of a vocabulary test: discrete–embedded, selective–comprehensive and context-independent–context-dependent. Schmitt (2000) notes that traditional tests have been mostly in the discrete, selective and context-independent end of the proposed continua. Read’s vocabulary testing dimensions are shown in Table 4.1.

Table 4.1 Read’s dimensions of vocabulary assessment (2000: 9)

<p>Discrete</p> <p><i>A measure of vocabulary knowledge or use as an independent construct</i></p>		<p>Embedded</p> <p><i>A measure of vocabulary which forms parts of the assessment of some other, larger construct</i></p>
<p>Selective</p> <p><i>A measure in which specific vocabulary items are the focus of the assessment</i></p>		<p>Comprehensive</p> <p><i>A measure which takes account of the whole vocabulary content of the input material (reading/listening) or the test-taker’s response (writing/speaking tasks)</i></p>
<p>Context-independent</p> <p><i>A vocabulary measure in which the test-taker can produce the expected response without referring to any context</i></p>		<p>Context-dependent</p> <p><i>A vocabulary measure which assesses the test-taker’s ability to take account of contextual information in order to produce the expected response</i></p>

The first dimension proposed by Read investigates whether vocabulary is being assessed as a **discrete** construct. This implies that what we see is what is usually called a

‘vocabulary test’, which measures lexical knowledge independently from other constructs, for example, grammatical knowledge. The majority of both breadth and depth tests are discrete measures of lexical knowledge (examples are discussed in Section 4.2). Vocabulary, however, can also be assessed as part of a larger construct, for instance, writing proficiency. In the case of **embedded** assessment vocabulary is treated as one element of the rating scale used to give the overall profile of learners’ performance.

The second dimension of Read’s framework distinguishes between tests of vocabulary knowledge that take into account only chosen lexical items or those that look at all items in input or output material. Read and Chapelle (2001) underline the fact that most conventional tests of vocabulary knowledge are **selective**, for they assess words selected on the basis of certain criteria. In the case of language teaching practice this usually happens when instructors would like to measure the knowledge of words previously introduced. But many large standard tests are also selective in nature, drawing on large frequency lists. Test-takers are presented with a list of words usually in isolation and are to be shown their knowledge of meaning or use of these target words. Examples are the Vocabulary Levels Tests used for this dissertation or the Vocabulary Knowledge Scale discussed in Section 4.2. **Comprehensive** measurement, on the other hand, analyzes all lexical items in learners’ oral or written performance. All measures of lexical richness discussed in Section 4.3 are good examples of comprehensive analysis. Lexical variation, for example, which shows the type/token ratio of a text, uses all lexical items in a given text for calculation.

The third dimension of Read’s framework takes into account the context in vocabulary assessment. **Context-independent** tests assess words in isolation, no contextual clue is provided to the test-taker. The receptive form of the Vocabulary Levels Tests is again a good example to this type of measurement. By contrast, in **context-dependent** instruments learners need to draw on the information provided by the input of the test or need to create output in which lexical items are appropriately used in the given context. For instance, in the C-test discussed by Singleton (1999) test-takers can provide the missing part of the target words by the help of the contextual clues provided in the texts. Many measures that target learners’ written performance are also context-dependent. An example is the Lexical Proficiency Profile discussed by Laufer and Nation (1995) and adapted for data analysis in this dissertation.

Read and Chapelle (2001) provide examples of six theoretically possible combinations of the three dimensions discussed above. They point out that the three dimensions can be viewed as being on a continuum, as different instruments can have a varying degree of context-dependency or embeddedness. They also conclude that “a comprehensive but context-independent measure is probably ruled out in principle” (7). I cannot fully agree with their statement in the light of the Lex30 instrument (Meara & Fitzpatrick, 2000) used for this dissertation. In this case, indeed, test-takers are asked to produce a string of words prompted by previously selected words which give minimal context to their output. The string of words written by the learners are treated as “texts” and analyzed with the help of computer-based text tools (for more detail on the Lex30, refer to Section 4.2).

While embedding the above discussed dimensions of Read (2000), Read and Chapelle (2001) propose a much more detailed framework for vocabulary testing which draws on Messick’s (1989) validation theory and includes validation, test design, mediating factors, validity considerations and test purpose. The framework is reported in Table 4.2.

Table 4.2 A framework for vocabulary testing proposed by Read and Chapelle (2001: 10)

TEST PURPOSE	Inferences	Uses	Intended impacts
VALIDITY CONSIDERATION	Construct validity	Relevance and utility	Actual consequences
MEDIATING FACTORS	Construct definition	Performance summary and reporting	Test presentation
	↘		↗
TEST DESIGN	Decisions about the dimensions Discrete - Embedded Selective - Comprehensive Context-independent – Context-dependent		
VALIDATION	Arguments based on theory, evidence and consequences		

Since the purpose of the present dissertation is not that of designing and validating a new instrument in order to assess vocabulary knowledge, the aspects mentioned by Read and

Chapelle in their framework are only briefly introduced in this section, with the aim of giving the issue of vocabulary measurement a theoretical perspective, and also of indicating the complex nature of the key factors that are often overlooked in the discussion of test selection and data analysis. According to the authors of the framework, the design of any assessment instrument must be based on the explicit description of the purpose it is to serve. Since learners' knowledge cannot be directly observed through the use of various tests, it can only be inferred from the proportion of correct responses. Read and Chapelle (2001) define **inferences** as "conclusions drawn about language ability or performance on the basis of how the test-takers perform on the test" (Read & Chapelle, 2001: 11). Inferences can be made at three levels: item level, sub-test level and whole test level. Embedded, context-dependent tests can draw on the sub-test level inferences, while discrete tests can also use whole test inferences by looking at overall test scores. The second component of the purpose dimension is **use**, which refers to the relevance and utility of tests. In the case of vocabulary tests, we can distinguish between instructional uses (placement, achievement or proficiency), research uses and evaluation uses. Tests designed for empirical research may have little practical outcome for practicing teachers; similarly, small-scale tests designed by instructors for specific student groups often provide little information to researchers. It is, however, possible that assessment tools have more than one purpose and are used in a variety of contexts. The third component of the purpose dimension is **intended impact and actual consequences**. This last component refers to the consequences of the test on test-takers, teachers, programs or wider communities. Research on test evaluation has been particularly focusing on the washback effect of certain major tests (Alderson & Wall, 1993; Wall, 1997).

The most crucial component of the mediating factors of Read and Chapelle's framework is **construct definition**. The authors discuss three major types of definition: trait definition, construct definition and the interactionist approach to definition. The **trait definition** views test performance as characterization of a learner's discrete form of structural knowledge. It usually proposes tests that are discrete, selective and context-independent. According to the second definition type, that of **behaviorists**, vocabulary is embedded into a larger construct, for example, writing or reading ability, and not tested separately, since single lexical items are not considered important. Tests that follow this view of definition are usually embedded, comprehensive and context-dependent. The third type of definition is an

interactionalist approach, which allows for a wider selection of test types. Vocabulary in this respect is tested in certain contexts of use, but the tests can be both embedded or discrete, and selective or comprehensive.

The second aspect of the mediating factors is **performance summary and reporting**. It refers to the way tests should be evaluated and learners' performance reported and interpreted. For example, results can be summarized as one overall score or as various sub-scores. These decisions have direct bearing on test-takers if the tests are used as placement or diagnostic tests. Large placement tests, such as the TOEFL or the Cambridge Placement Tests provide one single final score on the basis of which decisions about the acceptance or rejection of candidates for educational institutions or grants are often made. The third aspect of the mediating factors is **test presentation**. It implies that test designers carefully consider whom and why they are presenting a certain test type. This means that test type and audience should not be casually paired. Read and Chapelle call attention to the fact that certain tests, due to their popularity, are used for audiences different from the intended one, which may call validity into question.

The first three aspects of the framework discussed above (test purpose, validity considerations and mediating factors) have implications for **test format** and the **process of validation**. The three views of construct definition bring with themselves certain test formats, as has been pointed out above. Score reporting and test presentation also have implications for validation (for more on validity of tests see e.g. Chapelle, 1994; Bachman & Palmer, 1996; Bachman, 2000).

The framework of testing briefly reviewed above provides a theoretical basis for test selection and use. It is clear that many factors need to be taken into consideration when selecting a test appropriate for a given purpose. The following section will review some of the most widely used types of vocabulary tests.

4.3 Tests of vocabulary knowledge

After the brief introduction into the complex nature of test design and validity considerations, this section introduces some common test types that target the vocabulary knowledge of ESL/EFL learners as a separate knowledge component. It will be seen that

many tests are associated and referred to by their format rather than the vocabulary sub-knowledge or the vocabulary dimension they are testing. This practice is often misleading, as various formats can test different components of lexical proficiency, or have various purposes or ways of data evaluation. Some concrete examples of tests most often cited in the literature are discussed, with special emphasis on the three instruments used for data collection in this dissertation. It is evident that world-wide there is a great number of instruments used for local instructional purposes (including those designed by teachers for immediate classroom usage), large placement and diagnostic tests for institutional or research purposes. However, it is not within the scope of the present dissertation to enumerate all possible testing instruments cited in the L2 vocabulary research. Therefore, the first three sections concentrate on the tests employed in this study (Vocabulary Levels Tests, Productive Vocabulary Levels Test and Lex30), while other frequently cited test formats will be more briefly reviewed. This is done in order to have a clear view of what the selected tests are, what they measure, and how they are different from other tests that are frequently discussed in the literature reviewed. Also, since the present dissertation is focusing on the written assessment of vocabulary, reference will not be made to instruments that collect data orally, as many of the association formats or picture descriptions do.

4.3.1. Vocabulary Levels Test (VLT)

Multiple-choice tests are among the most popular language test formats. They are used not only to measure vocabulary, but also grammar, reading comprehension, listening comprehension, etc. Their obvious advantage over other test formats is their clarity for test-takers and easy scoring. Clearly the most often cited vocabulary test of the past two decades is the Vocabulary Levels Test (VLT). This test was originally designed by Nation as a diagnostic test to be used for instructional purposes. It was first published in 1983, then in 1990. When Nation (1990) became a major reference book, the test started to be used for research purposes. The original test was followed by a revised version (version A) and three additional versions (versions B, C, D) designed by Schmitt⁸. Meara (1996, cited in Zimmerman, 2004: 17–18) claimed that VLT was the “nearest thing we have to a standard

⁸ For a more detailed review of the various versions see Schmitt *et al.* (2001).

test in vocabulary”. The test has become the major research instrument of vocabulary measurement, and most other instruments are validated using the VLT as a basis. The different versions have been used in a number of studies (for example, see Schmitt & Meara, 1997; Beglar & Hunt, 1999; Laufer & Nation, 1999).

The VLT receives its name after the various frequency levels that its sections measure. Other than just providing one single score of learner’s vocabulary size, it gives a profile of their receptive vocabulary knowledge at five frequency levels or frequency bands: 2,000, 3,000, 5,000, 10,000 and academic vocabulary. The 2,000 level covers the most common 2,000 word families in English, the 3,000 frequency band the 2,001 through 3,000 most common word families and so on. The fifth band (representing 836 items) is approximately at the 3,000 to 6,000 word level and consists of the most academic words.

The words in the 2,000 through 10,000 levels were selected based on Thorndike and Lorge’s (1944) frequency count and were cross-checked against the 2,000 word level of West’s (1953) General Service List and the frequency lists of Kučera and Francis (1967). Target words on the academic level were all drawn from Xue and Nation’s (1984) University Word List (UWL), which was primarily compiled from the lists of Campion and Elley (1971) and Praninskas (1972), with some additions from the lists of Lynn (1973). The UWL is intended to be a list of general academic words which occur across a wide range of academic disciplines such as arts, science and law (Coxhead, 1998), and are the ones that university students reading articles and books during their English studies most frequently encounter (for more detail on the selection of tested words see Nation 1983, 1990; Beglar & Hunt, 1999). The UWL excludes technical vocabulary. Technical vocabulary is defined by Coxhead (1998) as vocabulary that is recognizably specific to a particular topic, field, or discipline. The latest test versions of the VLT use an improved listing of academic vocabulary which has been compiled from a new, carefully-balanced academic corpus, the Academic Word List (AWL) (Coxhead, 1998, 2000). The AWL has the advantage of giving a better coverage of academic texts than the UWL listing fewer words than the UWL including fewer word families (836 vs. 520).

The test is a six option matching test of words and their definitions. The test requires minimal reading as in the example reported in Figure 4.1 (for the full test see Appendix A).

- | | |
|---------------------|-------------------------|
| 1. <i>birth</i> | |
| 2. <i>dust</i> | _____ <i>game</i> |
| 3. <i>operation</i> | _____ <i>winning</i> |
| 4. <i>row</i> | _____ <i>being born</i> |
| 5. <i>sport</i> | |
| 6. <i>victory</i> | |

Figure 4.1 Sample test item from the Vocabulary Levels Test

The revised versions of the VLT, used for this dissertation, operate with ten clusters containing six words and three definitions to be matched, unlike the six clusters of the original VLT. This results in a longer, 150-item test versus the original 90-item one. Schmitt *et al.* (2001: 59) summarize the design purpose of each cluster as follows:

1. The offered options are words rather than multi-word items or definitions, making the test easy to read.
2. The definitions to be matched are selected in order to require minimal reading and to enable quick task completion.
3. The test requires only partial lexical knowledge. The option words in each cluster have very different form and meaning to help correct matching.
4. The clusters are designed to minimize guessing. Data on retrospective interviews reported in Schmitt *et al.* (2001) indeed support this claim, as only a limited number of students proved to be successful guessers on this test.
5. The words used in the definitions are always of lower-frequency than the option words. This ensures that the matching is not rendered difficult by the lack of knowledge of words in the definitions.
6. The most frequent member of each target word family is introduced in the test. This includes not only base forms, but some derivatives and affixes up to level five in the Bauer and Nation's (1993) hierarchy of affixes discussed in Chapter 2.

7. Option words in each cluster start with a different letter and have no similar orthographic or phonological form. Target words and their respective definitions also differ in orthography whenever it is possible. These considerations are crucial for reasons discussed in the section on word forms in Chapter 2.

The VLT is a context-independent, selective, discrete test, one that measures breadth or size of learners' receptive vocabulary. Validity of the test has been discussed in various recent studies. In terms of underlying mental processes it involves recognition of pre-selected test items. A preliminary validation study reported in Beglar and Hunt (1999) uses only Japanese students (n=496), both in secondary and tertiary education, and concentrates on the 2,000 and at the academic word levels. On the other hand, in their validation study of the new versions of the VLT, Schmitt *et al.* (2001) include students from a greater variety of educational, age, language and proficiency backgrounds (n=801). They conclude that both their student groups and native speakers perform well on the test. They point out that future research is needed in order to see how well the test works with other student populations and how guessing, which is partially allowed by the test, varies with test-takers' proficiency level. Other validation studies involving smaller numbers of subjects are reported in Cobb (1997, 2000), Tamaditsu (2001), Meara (2005), Laufer (2005) and Xing and Fulcher (2007).

4.3.2 Productive Vocabulary Levels Test (PVLТ)

The productive version of the Levels Test (PVLТ) was designed and first reported by Laufer and Nation (1995). Unlike the VLT, this productive (Laufer calls it 'active') test requires learners to produce words rather than recognize them. Therefore, the main underlying mental process involved is aided recall. Context is given in a sentence for all items. Similarly to a C-test, the first few letters of the missing words are given to provide a cue to the test-takers, but unlike in the case of a C-test, where half of a target word is provided, here only a minimal number of letters is given to disambiguate the cue. Sample items are reported below in Figure 4.2 (for the full test see Appendix B)

I'm glad we had this opp _____ to talk.

She found herself in a pred _____ without any hope for a solution.

Sudden noises at night sca _____ me a lot.

Figure 4.2 Sample test items from the Productive Vocabulary Levels Test

The overall structure of this test resembles that of the original VLT, sampling 18 items at five levels: 2,000, 3,000, 5,000, 10,000 word levels and the academic level. Test Version A uses the same items as the original VLT. Three additional versions exist, each using items of the parallel revised versions of the VLT. This is called a productive test as it measures the productive knowledge of form, meaning and syntactic properties of the target words, but receptive knowledge is also required to understand the context sentence. This makes this testing method more demanding and similar to real lexical usage. The productive version has the same frequency bands as the receptive Levels Test; therefore, parallel results are easily comparable. However, this test type is much less frequently administered and reported in research than the receptive version. Similarly to the VLT, the PVLTL allows for the rough estimation of test-takers' vocabulary size by calculating the proportion of right answers at each level. For example, if a learner scores 50% on the 2,000 level, it shows his mastery of around 1,000 words.

The validity and reliability of the PVLTL is discussed in Laufer and Nation (1999), based on the results of a small number of secondary school and university first-year students (n=79). They conclude that the PVLTL is a good indicator of the general proficiency level of EFL learners and it is practical and easy to use in classroom settings. They report continuous decrease in scores in the following order: the highest score at the 2,000 level, then follow the 3,000, academic, 5,000 and 10,000 word levels (the last one showing 0 or close to 0 scores in the case of secondary school students). While testing the equivalence of the four versions of the tests they also find that the correlations between the versions are significant enough to be used for test-retest purposes or for decision-making. They suggest using the VLT and PVLTL parallelly to investigate questions such as the development of the receptive and controlled productive lexical knowledge over time or the relationship between different types of lexical knowledge of the same learners and their change over time.

4.3.3 Lex30

The Lex30 association test was designed by Meara and Fitzpatrick (2000), and its validity has been discussed by Fitzpatrick and Meara (2004). The Lex30 uses thirty carefully selected low-frequency words and asks for multiple responses for each stimulus word, and measures the proportion of the words produced that do not fall into the first 1,000 most frequent English words (see Appendix C). This test elicits responses that are “more varied and less constrained by context than free production tasks” (Meara & Fitzpatrick, 2000: 22). While both Levels Tests require students to show their knowledge of pre-selected word items, the association format of the Lex30 asks students to write any word that comes to their mind when reading each of the stimuli. This unaided recall test reports on the frequency of response words as well as partial knowledge of form or meaning of the stimuli, since associative responses are possibly based on either form or meaning. The original format reported by Meara and Fitzpatrick (2000) is a pen-and-paper test, but similarly to other tests, a computerized version of this instrument has been designed at the University of Swansea, Wales. The computerized version asks for up to four responses to each stimulus and calculates the proportion of words typed in by the test-taker.

The association format of vocabulary tests are most frequently designed to assess the links between words in the mental lexicon. Answers to the stimuli are then categorized according to the relationship they have with the stimuli. The categorization most often used is syntagmatic, paradigmatic and clang responses. It needs to be clearly stated that the Lex30 was not designed to evaluate such links between the stimuli and the response words. It does not intend to analyze the links between the words provided either. It does provide some information on the links between words in the learner’s mental lexicon in general, assuming that less frequent words are prompted from the lexicon only if links are built between existing and new lexical items and the less frequent ones are quickly retrieved when needed (Fitzpatrick & Meara, 2004). A pilot study discussing answers to the Lex30 list of words in terms of association types has been reported in Doró (forthcoming). These association links will not be discussed in this dissertation as they would require a different set of research questions and research orientation.

4.3.4 Other test formats

Apart from the three tests discussed above, which all have a different test format, and are often associated and mentioned according to their test format rather than the knowledge type they intend to measure, there exist a number of other test formats. The most often cited instruments that target vocabulary as an independent construct are the yes/no tests, C-tests, translation tests and mixed formats assessing a variety of knowledge types parallelly.

4.3.4.1 Yes/no tests

The use of the checklist format of vocabulary testing has risen from the need of measuring vocabulary knowledge in classroom settings through an instrument which is easy to administer and to score. It requires a single judgment regarding the knowledge of the target items. The most often cited test using this method is actually called the Yes/No test which was developed and used by Meara and colleagues (Meara & Buxton, 1987; Meara & Jones, 1988; Meara & Milton, 2003). The checklist format allows for a large number of items to be tested in a short amount of time while scoring is also simple and time-saving. Target words are selected from a variety of frequency levels. In order to control for guessing, the Yes/No test also uses ‘pseudowords’, items that are phonologically and orthographically possible, but not real words of the target language. If test-takers check many pseudowords as known items, it is likely that they heavily rely on the strategy of guessing.

Another example of the yes/no format of vocabulary test is the Eurocenters Vocabulary Size Test 10KA (EVST). This test was designed by the Eurocenters chain of European language schools with the aim of finding a test format which is easy to administer and grade; therefore, saving faculty’s time (Read, 2000). It was first reported in Meara and Jones (1990). As it is computerized, the test-takers see words which appear on the computer screen and need to answer the question “Do you know this word?”. Since it does not require learners to demonstrate their knowledge of the meaning of the given words in any way, about one third of the tested items are pseudowords to balance results given by guessing. This is a test of vocabulary breadth and measures the knowledge of the 10,000 most frequent lemmas of English and can be used as a placement test (Nation, 2001). The advantage of the test is

that it takes less than 10 minutes to sit and results are immediately calculated by the computer. However, computer testing is optimal only if technical requirements are met.

The reliability and validity of the yes/no test format has also been discussed in the literature. Cameron (2002), in order to show the validity of the test, correlates the yes/no test with the VLT. She reports varying degrees of moderate correlation (at the academic level only a 0.20 correlation) and a high number of false alarms (pseudowords indicated as known items). Mochida and Herrington (2006), after reviewing contradictory results obtained in previous studies, also conclude that instructional differences, item differences and L1 background may have influenced validation results in previous reports. Therefore, they validate the yes/no test format more directly by using the same items from the VLT, unlike the original *Yes/No test* of Meara and Buxton (1987). This ensures that the two test formats are validated against each other and not the content of the tests. They call attention to the discrepancy between the easiness of administration of the test and the complexity of the interpretation of the scores. This is due to the use of pseudowords which distinguishes this test format from other tests of vocabulary knowledge. The interpretation of the false alarm rates has given rise to concern related to the reliability of this test format (Read, 2000). What Mochida and Herrington (2006) suggest is the introduction of easy, high-frequency words in place of the pseudowords to test false alarms and guessing.

To conclude, although the yes/no test format may seem to be a good and easily administrable test format, the difficulty of the interpretation of guessing and how it should be built into scoring make it less directly usable for a study similar to the one carried out in this dissertation.

4.3.4.2 C-tests

C-tests are usually not associated with vocabulary measurement, but are used to measure overall language proficiency (Eckes & Grotjahn, 2006; Dörnyei & Katona, 1992; Ittzésné, 1989). The C-test administered and viewed as one assessing vocabulary is discussed in Little and Singleton (1992), Singleton and Singleton (1998) and Singleton (1999). The test format employed for data collection asks test-takers to restore a text of which every second word has been partially deleted. The relevance of this test format in terms of information

gained about learners' vocabulary is thoroughly discussed in Singleton (1999) with detailed reference to the critique of Chapelle (1994) towards the test format's legitimacy in respect of lexical organization or processes. Chapelle voiced three major concerns with the C-test as a vocabulary measure: first, test items assess the surface knowledge of the orthographical, inflectional, derivational and semantic properties of the words; second, interpretation of the results in terms of test-takers' vocabulary size can be done only if items are selected more systematically; and third, responses do not reflect vocabulary organization. Singleton's view, in line with the interactionalist approach of vocabulary discussed in Section 4.1, is that of treating vocabulary not as a separate construct, but in context, as he advocates that "the viability of a separate lexical construct has to be seriously questioned" (1999: 269). He argues that it is no longer justifiable to restrict vocabulary research to measures involving knowledge of individual content words. It is, of course, more difficult to interpret data on more integrative tests than those assessing vocabulary size based on carefully selected content words. The relevance of the C-test has been tested by the above mentioned Singleton studies (Little & Singleton, 1992; Singleton & Singleton, 1998; Singleton, 1999) by correlating results to other data types, such as word-association, story-telling, translation tasks and retrospective interviews.

Although the aim of testing vocabulary knowledge not as a discrete component in accordance with the interactionalist definition of vocabulary is a valid attempt, the C-test remains difficult to interpret in a study like the one designed for this dissertation, which attempts to measure L2 vocabulary knowledge as a useful constraint on the basis of which it is possible to have a better understanding of vocabulary use.

4.3.4.3 Translation tests

Translation tests are traditionally viewed as ones assessing productive knowledge, as they ask test-takers to produce the L2 target equivalent. This is unquestionably still one of the most widely used vocabulary test formats in everyday classroom settings. It is very much context-independent and does not verify the ability to use the target item. Other than having a role in instruction, it is also used in research as an independent testing instrument or a tool to address the reliability of other test formats (see, for example, the case of the Vocabulary Knowledge Scale in the next section). Since translation tests involve the L1 of the test-takers

and no widely used test of this kind is available in the Hungarian context, this test format could not be considered for this dissertation.

4.3.4.4 Mixed test formats

The most widely used test of this mixed format is the Vocabulary Knowledge Scale (VKS). It was designed with the intention of measuring not only the size of the learners' lexicon, but also the depth dimension of knowledge in order to illustrate a developmental scale (Paribakht & Wesche, 1993; Wesche & Paribakht, 1996; cited in Schmitt, 2000: 175). It combines self-report tasks and production, and requires the test-takers to indicate on an acquisition continuum where their knowledge of the tested items might be (see Figure 4.3). As for the test format, it combines check-list, translation and free production formats. This test follows the partial–precise dimension of vocabulary knowledge, ranging from partial recognition of form and meaning to the use of the word in context-free production. It is, however, debatable how much of the depth of knowledge is assessed with the help of level 5, as practically even a word never seen before can be inserted in a neutral sentence. Although it has the advantage of allowing for partial knowledge, the test has a number of drawbacks. The lower level answers are not verified, the number of levels one should have on a knowledge continuum is not clearly identifiable, and different knowledge dimensions (recognizing, knowing the meaning, using) are mixed. The answers that refer to different stages of non-linear nature are scored on a scale of 1–5 (Wesche and Paribakht, 1996). Therefore, this linear scoring conflicts with the attempt to capture different degrees of knowledge. A final mean score is attributed to the learners' lexical knowledge which makes the interpretation of gains difficult.

1. I don't remember having seen this word before.
2. I have seen this word before, but I don't know what it means.
3. I have seen this word before, and I think it means _____ (synonym or translation).
4. I know this word. It means _____ (synonym or translation).
5. I can use this word in a sentence: _____

Figure 4.3 Test format of the Vocabulary Knowledge Scale test (Wesche & Paribakht, 1996)

Although this test format is an attempt to simultaneously measure different types and dimensions of vocabulary knowledge, its scoring makes it less straightforward and less informative to be employed in a study like the present one that wishes to assess different knowledge types separately, and still being able to distinguish them and refer to them separately. This is difficult in the case of this test, which unfortunately, seems to lose its benefits over other tests.

Waring (1997b, 2000, 2002) took the knowledge scale a step further to allow for the testing of multi-dimensionality of vocabulary knowledge. In his instrument called the Self Rating Task he uses simpler wording than in the previously discussed scale test, and a clearer distinction is made between receptive knowledge (understanding the meaning of the word) and productive use. The test asks students to identify a state of knowledge attributed to each tested item (A–E) (see Figure 4.4). Contrary to the knowledge scale format, this test does not require students to demonstrate their knowledge. Yet, as it has been pointed out in the case of the yes/no test, the checklist format can easily prompt answers on the basis of which over- or underestimation of learners' knowledge may result.

I do not know this word E		I think I understand this word ↓	I understand this word ↓
	I do not know how to use this word →	I think I understand this word but I don't know how to use it D	I think I understand this word and I know how to use it C
	I know how to use this word →	I understand this word but I don't know how to use it B	I understand this word and I know how to use it A

Figure 4.4 Test format of the Self Rating Task (Waring, 1997b)

4.3.5 Computer- and web-based tests

We cannot conclude this brief introduction of the various test formats without some consideration of the growing number of computer-adapted tests and the ones freely available on the Internet. In the last two decades, a growing interest in computer assisted language learning (CALL) could be seen. This has meant the promotion of autonomous language learning and practice, on-line communicative interactions, the use of lexical tools without an instructor's supervision or with minimal intervention from tutors or instructors, as research has shown that technical support can promote reflective and independent language learning (Goodfellow & Lamy, 1998).

Alongside with more general learning tools and on-line access to target language materials, proficiency tests with a lexical component have been designed, see for example the DIALANG assessment system which is based on principles of the Common European Framework of Reference (Alderson & Huhta, 2005). Explicit vocabulary-related projects and programs are also available in great number. Some of these target mainly native-speakers, others L2 learners preparing for exams (e.g. SAT, TOEFL), and many programs exist with the purpose of general vocabulary building of L2 speakers.

Another group of computer-based tools also need to be mentioned, namely tests of vocabulary knowledge that were either designed for computer use or have been adapted from the original pen-and-paper format. An example for the first case is the Eurocenter's Vocabulary Size Test which was designed for large group testing. The electronic use of originally paper-based tests is usually called for by the need for an easier and quicker administration process and data handling. Both versions of the Levels Tests are available freely in electronic format, and the Lex30 is also part of a selection of computerized vocabulary tests designed by the vocabulary research group at the University of Swansea, Wales. It needs to be stated, however, that results gained on tests in paper-based and computer-based formats are not always equal, as various aspects of the data collection methodology may change considerably⁹.

4.4 Assessing the vocabulary of written texts

After reviewing various tests of vocabulary knowledge, in this section I will discuss measures that do not focus on particular vocabulary items as tests do, but look more comprehensively at the vocabulary content of written texts. Since the empirical investigation in this dissertation does not include the overall assessment of the writing ability of L2 students¹⁰, this section will review some of the means L2 vocabulary researchers have used to assess the productive vocabulary employed in written texts and some of the statistical measures they have applied.

Productive written vocabulary can be measured not only through carefully designed tests that target learners' knowledge, but also from texts that reflect lexical use. As Nation (2001: 362) notes "vocabulary learning is not a goal in itself; it is done to help learners listen, speak, read or write more effectively. When testing vocabulary, it is important to distinguish between how well a word is known and how well a word is used." Laufer (1998, 2005) also underlines the importance of difference between vocabulary knowledge and vocabulary use. She states that while it is possible to improve L2 vocabulary, it is difficult to turn improved

⁹ For a more detailed analysis of the difference between paper-based and computer-based testing refer to e.g. Wang *et al.* (2007), Breland *et al.* (2005), Choi *et al.* (2003), Goldberg & Pedulla (2002) and Chin *et al.* (1991).

¹⁰ For measuring overall quality, linguistic accuracy, complexity, content, mechanics or coherence, see e.g. Polio (2001) and Lumley (2002).

vocabulary knowledge into use in writing. Nation (2001: 182) concludes that “there seem to be (at least) two important factors affecting productive vocabulary use: knowledge and motivation”. As vocabulary knowledge assessable through various tests is not necessarily transferred to vocabulary use in written performance tasks, it will be one of the key research questions of this dissertation to measure how lexical knowledge correlates with the words used in written essays.

4.4.1 Holistic and analytic scales

One possible and often applied way to measure written performance tasks is to apply holistic or analytic scales to assess the vocabulary component of texts (Tsang, 1996). While holistic rating provides one final score, analytic scales usually include components such as task achievement, vocabulary, structure and organization. Scales of various bands (most often 5 or 6) are given to each area to arrive at a composite score. Vocabulary assessment of this type includes not only the size of the lexicon activated for the given context, but also multi-word units, lexical errors, and appropriateness of style and register. The rater has a significant role in interpreting and using scales. Due to raters’ judgment involved in the assessment procedure, training and standardization of raters have to be ensured (for more on rating refer to Bukta, 2007).¹¹

4.4.2 Measures of lexical richness

Apart from holistic scales, the vocabulary use in written texts can also be characterized through various statistical measures which have the general term *lexical richness*. Since this method is more appropriate for the treatment of the data collected for the present empirical investigation, it will be discussed more in detail in the following section. Silva and Matsuda (2001: 98–100) list the following five ways to describe the richness of the vocabulary of texts:

- lexical variation = type/token ratio
- lexical density = content words/function words

¹¹ Bukta (2007) notes that raters report on having the most difficulty with judging the vocabulary component of essays and this aspect of rating is the most difficult one to standardize.

- lexical sophistication = the use of common words and rare words
- lexical errors = number of lexical errors or lexical errors/total errors
- lexical individuality = ratio of words unique to the writer

From the above list the aspects of lexical variation, density, errors and individuality have been widely discussed and used in vocabulary research (Nation, 2001; Read, 2000; Shaw & Liu, 1998). Type/token ratio is the most frequently quoted measurement of lexical richness of texts. Granger and Wynne (2000) call attention to the fact that the use of the type/token ratio in large learner corpora presents researchers with some considerable difficulty. First, due to frequent typing errors that remain in these collections of texts, all spelling variants of a word are counted as different types, unless corpus designers undertake the complex task of manual correction of all the compiled texts. Second, they point out that lemma/token ratio would be a more useful measure for pedagogical purposes, although this option is rarely offered by software packages like WordSmith Tools due to the unavailability and sensitivity of computer-aided lemmatizers. The authors conclude that “it is not safe to use crude type/token or lemma/token ratios with learner corpora” (Granger & Wynne 2000: 7). The problems concerning spelling variants or errors, however, are ruled out in the present empirical study as it does not use learner corpora, but shorter single texts entered to the computer by the researcher and corrected for spelling problems (see Section 7.4 in more detail about written data handling).

Laufer and Nation (1995) discuss some of the shortcomings of the measures such as lexical variation and lexical density and offer an alternative method to capture the lexical sophistication in texts. The *Lexical Proficiency Profile* (LPP) is a computerized measure that gives an analysis of any piece of written work. It describes texts in terms of vocabulary frequency bands by breaking them into four lists. These four lists are the following:

- the first 1,000 most frequently used word families
- the second 1,000 most frequently used word families
- the University Word List
- the remaining words (or off-list words)

The sources of these lists are the General Service List of English Words by West (1953) for the first 2,000 words, the University Words List by Praninskas (1972) and The New Academic Word List by Coxhead (2000) containing 570 word families. One of the major criticism of this vocabulary profile has been voiced concerning the date of publication of these word lists, however, the majority of the most frequent words have remained the same over the years. The LPF can be used to measure the difference between the lexical richness of texts written by native speakers and learners, to see the appropriateness of texts to be used with L2 learners, to measure the lexical richness of texts of different genres, or to capture the improvement in written vocabulary use.

Text tools are available which quickly calculate the LPF of any text, alongside with lexical variation and density. The online version of the text tool called VocabProfile is a user-friendly, regularly improved form of the original text tool, and within a few seconds it gives various types of vocabulary-related information about the analyzed text¹². A major advantage of this text tool is that words do not need to be manually tagged and that it gives information on academic vocabulary, a key aspect of the present research. Table 4.3 shows an example of the output calculated by VocabProfile for a learner's text collected during piloting for the empirical investigation discussed in this dissertation¹³.

Table 4.3 Lexical Proficiency Profile for a learner text calculated by version 2.5 of the VocabProfile

	<i>Families</i>	<i>Types</i>	<i>Tokens</i>	<i>Percent</i>		
<i>First 500</i>			(307)	(76.75%)	<i>Words in text(tokens):</i>	400
<i>K1 words (1 to 1,000)</i>	126	143	358	89.20%	<i>Different words (types):</i>	174
<i>Function</i>			(206)	(51.50%)	<i>Type/token ratio:</i>	0.44
<i>Content</i>			(152)	(38.00%)	<i>Token per type:</i>	2.30
<i>> Anglo-Saxon</i>			(109)	(27.25%)	<i>Lex density (content words/total):</i>	0.48
<i>K2 words (1001 to 2,000)</i>	18	19	21	5.25%		
<i>> Anglo-Saxon</i>			(14)	(3.50%)	<i>Pertaining to onlist only</i>	
<i>AWL words</i>	2	2	2	0.50%	<i>Tokens:</i>	381
<i>> Anglo-Saxon</i>				(0.00%)	<i>Types:</i>	164
<i>Off-list words</i>	?		19	4.75%	<i>Families:</i>	146
	146+?	174	400	100%	<i>Tokens per family:</i>	2.61
					<i>Types per family:</i>	1.12
					<i>Anglo-Sax Index:</i>	86.35%
					<i>Greco-Lat/Fr-Cognate Index:</i>	13.65%

¹² The original text tool called RANGE can be downloaded from Paul Nation's homepage at http://www.vuw.ac.nz/lals/staff/Paul_Nation, or it is available online on Tom Cobb's Complete Lexical Tutor webpage at <http://www.lextutor.ca>

¹³ For more detail on piloting, refer to Chapter 7.

Although the table provides an extensive amount of information, for the present empirical study the following data are of utmost importance: tokens and percentages for the first 500 words, K1 and K2 bands (the first 1,000 and the second 1,000 most frequent words, respectively), academic vocabulary, and off-list words. On the right hand side of the table, the number of words in the text, the type/token ratio and lexical density are calculated, these are frequently used measures of the lexical richness of the analyzed written production. The data in Table 4 show that almost 89% of the 400 word text consists of the first 1,000 most frequent words in English (out of which a high proportion, 77% are very frequent words), slightly more than 5% is drawn from the first 2,000 word band. Also, there are only two academic words in the text and almost 5% are off-list words (meaning not academic or highly frequent words).

4.4.3 Corpus studies

As a result of advancing technology, a new area of research within SLA/FLT has emerged (Granger, 1998, 2003; Szirmai, 2001, 2005): similarly to large native language corpora, learners' oral, but mostly written texts have been gathered as a rich source for text analysis. There exist raw and annotated corpora, for commercial purposes (e.g. the *Longman Learners' Corpus* and the *Cambridge Learners' Corpus*) and for academic ones (e.g. the *Hong Kong University of Science and Technology Learner Corpus*). Granger (2004) points out that the academic corpora greatly vary in size, content, annotation type, the degree of careful selection of texts, etc. Unfortunately, most academic corpora are not available for the research community, which makes them difficult to compare. The only carefully selected international learner corpus is the *International Corpus of Learner English (ICLE)*. What Granger concludes is that small corpora compiled by teachers or researchers can also be of utmost value if the purpose is to gain a better knowledge of a selected group of learners' skills.

Similarly to other countries, in Hungary the attempt has been made to build learner corpora to facilitate EFL research. Learners' texts have been compiled at the University of Pécs by Horváth (2000, 2001, 2007). The *Jannus Pannonius University Corpus* designed by Horváth is a growing written corpus of essays and university seminar papers. The advantage

of similar corpora is the great number of words they contain (which is hard to reach with shorter texts written under controlled circumstances) and the quick data analysis that is available thanks to computer research tools (Granger 1996, 2004, 2005). However, most of the essays gathered are written at home with the help of dictionaries (and probably sample papers and information gathered from other reference materials, which cannot be verified). Therefore, they do not give an accurate picture of the actively available skills, vocabulary or grammar of the learners. If background information on the learners or other language data (e.g. different test results, oral or written production tasks registered at different times from the same subjects) are not available, the compiled data can be used only to trace overall tendencies of the writing of the given student population. Although essays written under controlled circumstances in the previous years by the students at the University of Szeged are available, they could not be used to compile a corpus for this proposed project, as no information on the authors of the texts has been recorded.

4.5 Word frequency

Since many of the research instruments discussed above and corpus tools available rely heavily on frequency counts, it is inevitable to address some of the issues and possible problems that lie within this topic.

A major concern with the measurement instruments (both tests and text tools) based on older frequency counts is the changing nature of language and the possible revision of research tools thanks to the availability of more recent lists. Beglar and Hunt (1999), while validating the VLT conclude that future revisions of this test should be cross-checked using more recent frequency information based on, for example, *Collins' COBUILD Bank of English Corpus*¹⁴ and/or *The British National Corpus*¹⁵. They report that a preliminary comparison of the vocabulary items on the 2000 word level test with the frequency data in *Collins' COBUILD English Dictionary* (1995) and Longman's *Dictionary of Contemporary English* (1995) revealed that some of the target words originally drawn from the GSL are not in the top 2,000 words according to these dictionaries.

¹⁴ Available at <http://titania.cobuild.collins.co.uk/>

¹⁵ Available at <http://info.ox.ac.uk/bnc/>

However, even recent corpora need to be used with care. As Jackson (2007) points out, relying on word frequency without keeping in mind what the actual corpus is made of or should be used for can be misleading. It is important to note that two of the tests employed in the empirical investigation of the dissertation were originally designed by the same researchers and test the same vocabulary frequency bands. Furthermore, data obtained on the association test and the essays are analyzed with the help of the same research tool (VocabProfile) which makes the direct comparison and correlation of the data more valid than using various, more recent frequency counts parallelly.

4.6 Concluding summary

As we have seen in these three chapters of the literature review so far, vocabulary knowledge is a complex notion. Different dimensions of lexical knowledge have been discussed by researchers, one of which is the receptive–productive distinction. Scholars interpret these notions in a variety of ways; therefore, it is difficult to capture the true nature of the L2 lexicon. When designing or choosing a test of lexical knowledge, key points, such as word choice, item format, length of test, time and cost efficiency, the learners' L2 proficiency, type of vocabulary knowledge to be assessed, whether we want to measure achievement or proficiency, etc., need to be considered.

Various tests of vocabulary knowledge have been designed, the majority of which focus on one particular aspect of vocabulary and usually these tests are labeled as receptive, productive or knowledge scale, on the basis of the test designers' attempt to measure these particular aspects of the lexicon. However, it is not always certain whether the tests really measure these and only these aspects of the L2. As has been discussed in this review, there are many interrelated factors that must be taken into consideration when designing tasks and scoring procedures for assessment. There seems to be some agreement on what some test formats measure (e.g. multiple-choice tests are likely to measure some aspects of receptive knowledge). However, with the use of a larger variety of test types the picture has been widened. But even studies using the same or similar testing methods often report surprisingly varying outcomes and interpretations of the data. Other than the lack of clear definitions and descriptions of basic categories, what is missing from the literature are the longitudinal

studies addressing vocabulary-related issues. The cross-sectional data using a great variety of student populations in terms of size, L1 background, language-proficiency and different instruments make it challenging to draw general conclusive results.

Other than administering standard tests, the L2 lexicon of learners has also been measured in written production tasks alongside with other aspects of texts, such as grammar or cohesion. Written production tasks are good indicators of what learners use of their lexicon in a given written context. It is possible to measure the relationship between various vocabulary measurement tasks, including tests and written texts, so that in the future better assumptions can be made about other aspects of the L2 vocabulary when only one aspect of the lexicon is measured. Various studies have attempted to show this relationship, but the correlation studies report various results. This calls attention to the fact that many factors influence scores of vocabulary measurements, such as language proficiency, learning methods, motivation, target language practice, strategies of use, testing and scoring methods. Thus, it is of great importance that as many of these aspects as possible are clearly explained in the research reports to facilitate the interpretation of the collected data.

It has also been shown that, with the help of computer technology, compiling and analyzing texts have become easier than in the past. Large and smaller size learner corpora have been built similarly to native corpora, and simple research tools are available to both instructors and learners that can also be used in the everyday L2 learning process. Although various research instruments have facilitated data collection and analysis, it has been pointed out that the decision-making of the researcher at all stages of a study is crucial and may have direct impact on the outcome of research projects.

In order to account for the complex nature of vocabulary knowledge and the diversity of tests in use, a multidimensional approach of assessment was designed for the empirical investigation of this dissertation. This assessment employs three testing instruments differing in their format, type of lexical knowledge targeted, and the underlying mental processes. They will form the basis of assessment of vocabulary knowledge and use discussed in more detail in the following chapter.

CHAPTER FIVE

VOCABULARY AND ITS RELATIONSHIP WITH READING AND WRITING

5.1 Introduction

Following up on the previous chapters in which we reviewed issues concerning the notion of word, aspects of vocabulary knowledge and finally measures of vocabulary knowledge and use in written context, this chapter aims to provide an insight into the research concerning reading and writing in relationship with vocabulary knowledge. As will be pointed out, these two broader issues can be looked at from different angles. In the case of reading, for instance, it is important to discuss, on the one hand, the number of words necessary for effective reading, and, on the other hand, the rate of vocabulary learned through reading. Similarly, while evaluating the relationship between vocabulary and writing, we can discuss the vocabulary used by learners in written texts, the effect of their lexical proficiency on their writing, the role of genre and topic choice, the role of writing experience and use of materials such as dictionaries. These issues will be directly addressed in the research questions of this dissertation, therefore, need to be treated in detail in this chapter.

As I have already pointed out in Chapter 4, even a large vocabulary size assessable through tests of vocabulary knowledge is not a guarantee for the use of this vocabulary in context. It will be shown that vocabulary size, the accessibility of this vocabulary during language use, language practice and subject matter knowledge are interrelated, and the inadequacy in any one of these areas may result in difficult language use, pose restraints on subject knowledge gain and imply very limited vocabulary growth through reading and writing activities.

It has been noted in the introductory chapter that written language input and output play a significant role in target language medium studies. English majors are expected to not only follow classes conducted in the target language, but also to study from their written notes, textbooks, and to read authentic general and academic texts. Moreover, they are expected to read for pleasure, and to consult materials not directly related to their studies.

Gain in content knowledge is possible only if they understand the assigned readings without great difficulty. Since the minimum requirement towards understanding a text is the knowledge of the words it contains, a threshold level of vocabulary size has to be met. This minimum vocabulary size will be explained in the following sections.

Alongside with reading general and academic texts, students are expected to produce texts themselves in forms of summaries, essays, short answers to exam questions, research papers, etc. The relationship between vocabulary knowledge assessable through tests and the lexicon of written texts is, therefore, a vital question to be examined before we turn to the empirical study carried out among a Hungarian university student population. For this reason, in the second half of this chapter, after a review of the role the lexicon plays in text production, previous research that directly relates to the vocabulary knowledge and the lexical profile of the texts produced by learners will be discussed.

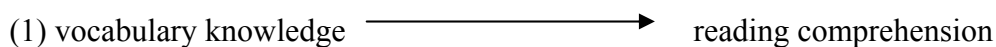
5.2 Relationship between vocabulary and reading

The relationship between vocabulary and reading can be investigated from two major points of view. The first one assesses the minimum vocabulary knowledge needed for successful reading, and the second one focuses on the vocabulary gained from reading. Although these two points of view may seem the opposite ends of a broad issue, they are interrelated to a great extent, thus, both need to be discussed in the following sections. As will be seen, if learners do not meet the lexical threshold for reading comprehension, they will have difficulty in understanding texts. As a result, it is unlikely that new vocabulary is learned to a great extent from this form of input. As a theoretical basis for these issues, the next section aims to review the various views concerning the causal relationship between vocabulary size and reading comprehension.

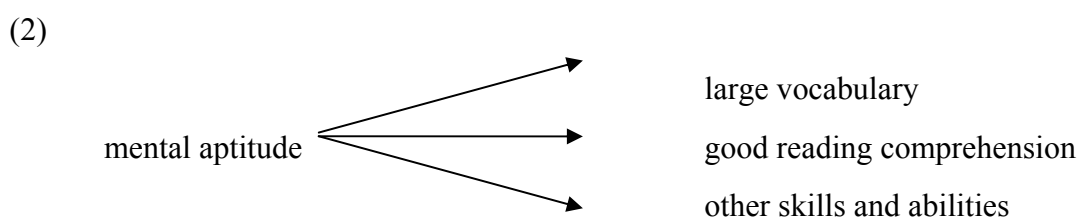
5.2.1 Causal relationship between vocabulary size and reading comprehension

Nation in his study (1993: 115–117) reviews some of the fundamental views concerning the relationship between lexical knowledge and reading comprehension. According to the **instrumentalist view**, there is a direct link between lexical size and reading

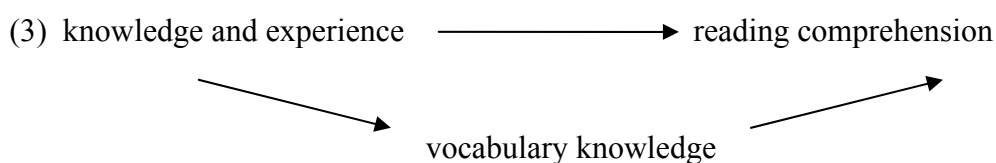
comprehension. Based on this rather simplistic view, good vocabulary coverage directly leads to good text comprehension, as shown by the following diagram:



According to the **aptitude view**, a mental aptitude, in other words, a good brain is the basis of both the learner's lexicon and reading comprehension, alongside with many other skills and abilities, but not all of them are language related. This view is illustrated by the following diagram:

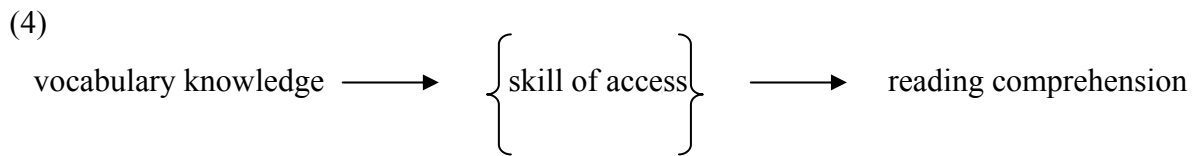


A third view takes a step further in bringing knowledge and experience into the picture. According to the **knowledge view**, vocabulary is not only a basis for reading comprehension, but also an indicator of content knowledge. What is called 'world knowledge', is all the subject knowledge which is necessary and which is deepened as a result of reading. Language knowledge, therefore, is not enough for text comprehension in itself.



A fourth one, the **access view**, similarly to the instrumentalist one, places a causal relationship between vocabulary knowledge and reading comprehension, with one crucial factor in the process, namely access to the vocabulary. This is indeed similar to the automaticity component of Qian's (2002) framework of vocabulary knowledge. If access to the vocabulary in the mental lexicon is difficult, or cannot be activated, comprehension is only partial.

This view can be illustrated the following way:



Nation (1993) concludes that the factors involved in these views are the following: a) mental aptitude; b) vocabulary knowledge (receptive size); c) skill in language use (reading comprehension in this case); and d) knowledge of the world (subject familiarity). He also notes that the relationship between the last three factors is a changing one, affected by a growing vocabulary size, reading practice and gain in subject knowledge. He summarizes learners' vocabulary development in relation to these three factors as the following four implications: a) skill in language use depends on learners' vocabulary size; b) knowledge of the world depends on the skills in language use; c) vocabulary growth is affected by knowledge of the world; and d) broad vocabulary growth is dependent on vocabulary strategies which are independent of subject knowledge (Nation, 1993: 118–124).

The first of these aspects, namely the vocabulary threshold level needed for reading is discussed in detail in Section 5.1.2 below. The threshold vocabulary necessary for academic studies in the target language will be assessed in this study by using the Vocabulary Levels Test that gives information on the size of the subjects' receptive lexicon and academic word knowledge.

The second aspect stresses the importance of fluency of access to the lexical items in the mental lexicon. This implies that known vocabulary items are reactivated for use through association networks. Automaticity and fluency in accessing words highly depend on practice. The presence or lack of richer and stronger associative links between less frequent words in the mental lexicon will be targeted in this dissertation through the use of the Lex30 association test.

The third aspect covers the need to learn the technical vocabulary of the field of study, as they include specific terms. Nation (1993: 122–123) suggests that technical vocabulary should be treated differently from other vocabulary items, because they cannot be learned the same way. Simply looking them up in a dictionary or learning a list of them relevant to the specific

field will not help the reader who is lacking subject knowledge background. Technical vocabulary must develop alongside with gaining information about the field.

The fourth aspect of relationship between the three aspects involved in vocabulary size and reading is illustrated by Haastrup's (1990, cited in Nation 1993: 124–125) model of inferring meaning from context (discussed more in detail in Section 5.1.3). Haastrup differentiates between 'top down' and 'bottom up' inference, noting that it is not a conscious decision from the part of the reader which strategy to employ, as it depends on the knowledge of vocabulary and subject matter brought into the reading activities. In the case of the '**top down**' meaning inference, readers have a very good familiarity with the topic, but show lack of language knowledge. This means that they can infer the meaning of many unfamiliar lexical items, but vocabulary learning is minimal. On the other hand, during '**bottom up**' meaning inference, we see good language knowledge, but little subject knowledge. In this case most of the vocabulary is known to the reader and unfamiliar word meaning can be inferred and learned.

At this point, before discussing some of the above mentioned aspects in more detail, the question needs to be asked which one of these two reading strategies university students enrolled in target language medium education employ. They are now getting introduced a variety of subjects; therefore, good topic familiarity necessary for the 'top down' process is highly questionable. That would imply using the second strategy, the 'bottom up' one, which requires, as a basis, excellent vocabulary knowledge. The question emerges here whether students have the threshold knowledge necessary for this second reading strategy. If not, can they at all infer and retain new word meaning? These questions will be directly targeted in the empirical research of this dissertation, first, by assessing the receptive vocabulary size of subjects and comparing it to threshold levels set for successful reading, and second, by addressing the question of how much students read and whether that can be enough for practicing vocabulary access and reading automaticity. If these basic requirements are not met, then, according to the theoretical framework reviewed by Nation (1993), neither subject matter knowledge, nor vocabulary size gain is easy, if not close to impossible.

5.2.2 Vocabulary coverage needed for effective reading

Research has found that both depth and breadth of vocabulary knowledge play an important role in reading comprehension, some of the most recent studies being Laufer and Yano (2001), Nassaji (2004), Hunt and Beglar (2005), and Shiotsu and Weir (2007). Reading, of course, is not simply about understanding every word in the text, as has been reviewed in the previous section. As Brutton (1981) rightly states, while reviewing earlier studies on reading strategies and vocabulary difficulty, comprehension can be seriously limited if learners have a restricted lexical knowledge for the given text and, as a result, are unable to note which the key words are and how to use contextual cues. In this case learners do not feel secure in transferring reading strategies from their L1, instead, they would like to know the meaning of each word, even though not all words are equally important for the global understanding of the text. Honeyfield (1977) compares the act of inferring the meaning of unknown words from context to a cloze test type gap filling exercise. What he calls attention to is the fact that the two seemingly different tasks pose very similar problems to learners, as a text with unknown lexical elements looks like a text with blank spaces. What needs to be added to his comment is that the unknown words in many cases may be mistaken for known lexical items due to inter- or intralinguistic factors (such as phonological, morphological, orthographical, grammatical and semantic factors, as well as, length and synformy)¹⁶, and, as a consequence, mislead the reader towards a false decoding of meaning (Laufer, 1997b). Table 5.1 shows how the rate of vocabulary coverage in a text changes the number of unknown items. Even as little as a few percentage points of change in the vocabulary coverage leads to a significant change in the number of unknown elements. While with 90% coverage every tenth token is unfamiliar for the reader (meaning more than one unknown word per text line), 98% coverage sharply decreases this rate to every fiftieth word. What may seem to be a small difference between a 98% and a 99% text coverage, indeed, is an enormous difference if turned into the number of unknown tokens in a text. This implies not only a quantitative difference in the number of unfamiliar words, but also a huge qualitative

¹⁶ By *intralexical factors* Laufer (1997b) refers to intrinsic factors related to the word's form and meaning; and by *interlexical factors* she refers to relationship between the word and other familiar words in the L2 or other languages. It needs to be pointed out that *interlexical* is a more general term than *interlanguage*, as it is not restricted to transfer between L1 and L2.

difference in reading, which may require different reading strategies, and what is more, greatly affect reading fluency and the amount of time dedicated to reading a text.

Table 5.1. Change in the number of unknown words with increase in text coverage

Text coverage %	Unfamiliar tokens in a text every
99	100th
98	50th
97	33rd
95	20th
90	10th
80	5th

Various studies have attempted to find a minimal coverage rate needed for successful reading. In other words, researchers have pointed out the need to have a certain percentage of the words known in a text for the successful guessing of unknown items from context. Recent research seems to stress a higher percentage of minimal coverage necessary for reading than earlier studies. In the following paragraphs I review some of the most relevant studies targeting this issue of coverage.

In Honeyfield (1977) the exact text coverage is not given, attention is rather paid towards what students with 80–90% gained through formal instruction do with authentic texts. The author discusses syntactic, semantic, pragmatic and topic familiarity which can facilitate the inferring of word meaning in context. A few years later Na and Nation's (1985) results suggest that 95% text coverage is a precondition for partial guessing of the meaning of the remaining words. The authors use passages in which a number of words are replaced with nonsense words, keeping the original length, part of speech, inflection and suffixes. Subjects included EFL teachers (n=59) with a variety of English proficiency background. Results of this study show that a higher density of unknown words makes guessing more difficult. They also point to a hierarchy of difficulty, verbs being the easiest to infer, followed by nouns, adverbs and adjectives. Subjects of higher proficiency level could guess unknown words easier than lower level subjects.

Hirsh and Nation (1992) increase this 95% coverage to 97–98%. Their reason for moving the threshold level by a few percentage points is the sharp drop in the number of unfamiliar words with only one percent change in the coverage at this high rate, as shown in Table 5.1 above. What they are interested in is the vocabulary size needed to read three

different short novels. They, therefore, turn percentage figures into vocabulary size needed to read without difficulty. Their results show that the 2,000 words of the *General Service List* (West, 1953), plus the proper nouns in the novels they used, provide roughly 92.5–95% coverage. However, their calculation suggests that 5,000 words are needed for 97–98% and up to 7,000 words for 98–99% coverage. They set the minimal lexical knowledge at 5,000 words for pleasure reading, moreover, they suggest the computer analysis of texts and the pre-teaching of key low-frequency words.

Hu and Nation (2000) take the matter one step further in checking subjects' comprehension of a text with three different text coverage densities, by elaborating on the methodology used in the Na and Nation (1985) study. Their subjects include adults preparing for under-graduate or post-graduate studies in an ESL context (n=66). The selection of participants for the study is based on the Vocabulary Levels Test results (minimum 80% at the 2,000 level). Subjects are presented with a text (657 words) with one of the following three coverage levels: 80, 90, 95%. Unknown words are provided by the introduction of nonsense words. Understanding of the texts is checked by using a multiple-choice test covering all main points and a cued recall test using a number of questions. The result provides experimental support for previous findings in terms of the increase of understanding with higher text coverage. Although some learners at the 90 and 95% coverage gained some understanding of the given text, the authors conclude that 98% coverage is necessary for unassisted reading of fiction. Their findings also support the predictable nature of understanding with the increase in text coverage.

As a summary, from the various percentage figures it can be concluded that a minimum of 3,000 word families are necessary to understand general English texts, and a higher 5,000 word family coverage, as well as academic vocabulary in order to follow academic texts. This needs to be complemented with the knowledge of a so-called technical vocabulary, lexical items related to specific topics and fields of science. Chung and Nation (2003) note that 38% of the running words in an anatomy text and 17% of the words in an applied linguistics text can be considered technical. Only a marginal part of these are included in the academic vocabulary lists, and, as a result, they pose an extra load of vocabulary necessary for reading these specific academic genres. If specific technical terms are not understood, then they are likely to already push the ratio of unknown words over the threshold

level. It follows that, for reading academic text, a sound basis of a high level of general vocabulary is even more essential than in the case of general texts. Moreover, as Sutarsyah *et al.* (1994) rightly point out, reading for academic purposes at a university level requires the knowledge of several fields and topics, both in terms of general background information and technical vocabulary. Also, these formal academic texts put a greater cognitive demand on the reader than general texts. These issues are key points in the case of the subject population used for the present dissertation, as during their English studies they are following classes in all areas of humanities, including, but not limited to, various topics in literature, linguistics, history, art, and philosophy. The technical vocabulary necessary to understand these academic texts is crucial to be learned early on, otherwise they add to the percentage of unknown vocabulary to such a degree that understanding texts is difficult even with a dictionary and other forms of help.

Hu and Nation (2001: 407) conclude that there are at least two possible interpretations of the threshold level of vocabulary coverage discussed above. According to the first one, it is seen as an “all-or-nothing phenomenon”, without which adequate text comprehension is impossible. The second view, on the contrary, considers the threshold level as a “probabilistic boundary”. If a learner does not reach this level, understanding will likely to be seriously challenged, while higher coverage will give better chances to the reader for successful decoding of the text. It needs to be stated again that vocabulary knowledge is not the only factor in reading comprehension, but inadequate coverage makes background knowledge and reading skills insufficient for unassisted reading.

One final point needs to be added. With the growing availability of technology for research and everyday pedagogical application, it is now possible to quickly check the word frequency profile of texts and the possible difficulty of reading comprehension due to lack of lexical coverage. Software programs are freely available on the Internet and on CD-ROM which greatly facilitate this process. Examples are the *Complete Lexical Tutor* (Cobb, 2008a) used for analysis in this dissertation, or the *Frequency Level Checker* (Maeda & Hobara, 1999). Recent research using information technology can improve the use of text coverage measures, as was done by Chujo and Utiyama (2005) who explored factors such as text length, genre and vocabulary size that may affect text coverage. With these in mind I now turn to another related topic, the lexical items learned while reading texts.

5.2.3 Vocabulary learned through reading

While the above discussed text coverage and minimum vocabulary size have a well-defined and limited body of research within vocabulary studies, the vocabulary learning that takes place through reading has a much wider literature. This section aims to review some of the key issues within this type of relationship between vocabulary and reading, with special attention to the factors that directly relate to the empirical investigation in this dissertation. Although the central topic of the present study is not the detailed review of vocabulary learning strategies and instruction, I need to dedicate some discussion to the potential of vocabulary learning from written input as this may be considered as one of the leading sources of vocabulary gain of a university language major population.

5.2.3.1 *Incidental vocabulary learning*

The view that incidental vocabulary learning takes place through reading, as a ‘by-product’ of the reading process has a long history in second language reading research (Coady & Nation, 1988; Nation, 2001)¹⁷. Waring and Nation (2004) review some of the most representative studies related to the incidental vocabulary learning through reading (e.g. Pitts, White & Krashen, 1989; Hulstijn, 1992; Horst, Cobb & Meara, 1998; Zahar, Cobb & Spada, 2001). Results point towards some retention of the tested vocabulary items, but the picture is very much mixed. Various results are partly due to the diversity of research methods employed (including text type, tests of vocabulary knowledge, number of encounters) and the type of knowledge tested after reading. It can be concluded that the topic of incidental vocabulary learning and its testing is a much less straightforward issue, as it might have been seen in early research. Na and Nation (1985) rightly stated the following ideas, but what is missing from the picture is the complex nature of inference of meaning in text and how inferred knowledge can be turned into lexical gain:

¹⁷ In Hulstijn’s definition incidental vocabulary acquisition is the “learning of vocabulary as the by-product of any activity not explicitly geared to vocabulary learning”, while intentional vocabulary learning is viewed as “any activity geared at committing lexical information to memory” (Hulstijn 2001: 271, cited in Rieder, 2003: 25).

There are many low frequency words and their occurrence is largely unpredictable, so it is not possible to learn them in advance. Thus dealing with them as they occur is the only feasible way of handling them. Because of their narrow range and the low probability of meeting them again soon, they do not deserve much effort in learning them. It is better to use context clues to infer their meaning than to spend time on learning the words themselves (Na & Nation, 1985: 33).

In recent years criticism towards early studies and conclusions has been voiced by authors in need of a fuller picture of incidental vocabulary learning. Rieder (2002) calls for a need of a distinction between text meaning and word meaning. She notes that the reader is at a text level when inferring the meaning of the unknown lexical item, and a subsequent shift is needed to the word level understanding in order for acquisition to take place. She stresses the fact that incidental vocabulary acquisition is a result of a complex network of processes and factors influencing acquisition. She concludes that “the conditions for the incidental acquisition of an unknown word which a learner encounters in a text appear to be the result of two complementary determinants: the *availability* of the *resources* required for specifying the contextual meaning successfully, and the *necessity* for *focusing* on this particular word” (Rieder, 2002: 67–68). The resources include the textual clues and the reader’s language and broader world knowledge. The necessity for word meaning specification is made up of textual conditions (e.g. position of the word), learner conditions (e.g. difficulty of text comprehension) and situational factors (e.g. goal for reading and time available). A key precondition for acquisition is the reader’s attention to the word. This means that if a word is ignored in the text or is mistaken for another word similar in form, acquisition will not happen. From this point of view it is crucial that learners accurately estimate whether a lexical item is familiar to them or not. Laufer and Yano (2001) investigated the accuracy of learners’ understanding of familiarity of words in context. 106 subjects with different residential and cultural backgrounds (China, Israel and Japan) all overestimated to different degrees (which shows a role of cultural differences) their understanding of lexical items. Subjects’ perceived knowledge did not match the actual knowledge tested through various tasks, which means they treated a large number of unfamiliar words as known items. This implies difficulty in coding the meaning of words or the serious misinterpretation of a text.

In another study Rieder (2003) investigates the relationship between implicit and explicit learning and incidental vocabulary acquisition. She calls attention to the unclear use of the terminology concerning implicit/explicit and incidental/intentional learning. She bases her

discussion and criticism on Ellis's (1994) model, in which incidental learning of meaning is characterized by explicit learning and the learning of form happens through implicit learning.¹⁸ The author feels the need to refer to the term implicit as 'non-conscious' or unaware, and incidental as 'un-intentional', not posing awareness restrictions to the term. In this sense incidental vocabulary acquisition can be characterized by both implicit and explicit learning processes. Through empirical investigation among German speakers of English she shows the need for a more differentiating model and not a simple implicit–explicit dichotomy. Her data show different levels of forms of explicitness in the case of both vocabulary learning and knowledge through reading.

If too many words are unknown in texts, there is a likelihood that students will not read them even with extra dictionary work. As Dycus (1997) warns, guessing should not be the only strategy used while reading and encountering unknown vocabulary. As he states, "guessing can easily become a strategy for frustration and demotivation, instead of for improvement and learning" (Dycus, 1997: 4). Within the limited Hungarian research on the subject, one of Peckham's (2006) findings among English major students similar to my study population needs to be stressed. Students were able to respond to open questions checking the understanding of a text with a high proportion of pseudowords. Some of these pseudowords even functioned as adjectives which authors of similar research have found to be the most difficult ones to guess (Na & Nation, 1985). The risk of extensive guessing to this point can be the development of the habit of inferring meaning without an outside control of insuring real meaning. Moreover, if limited reading is done by students, then we cannot talk about extensive reading necessary to develop advanced vocabulary and fluency in the target language.

While reviewing previous studies, Nassaji (2004) summarizes the following factors that may affect inferring unknown lexical items in a text: a) the nature of the word and the entire text; b) the degree of textual information available; c) the learner's ability to make use of extra-textual cues; d) the perceived importance of the word to understanding of the text; e) the degree of cognitive and mental effort involved in the task; f) the learner's attention to the details in the text; and finally g) the reader's preconceptions about the possible meaning of the

¹⁸ Ellis defines implicit learning as "acquisition of knowledge about the underlying structure of a complex stimulus environment by a process which takes place naturally, simply and without conscious operation" and explicit learning as the one characterized by "more conscious operation where the individual makes and tests hypothesis in a search for structure" (Ellis 1994: 1, cited in Rieder, 2003: 25).

word.¹⁹ Moreover, Laufer (1997a) concludes that L2 readers tend to heavily rely on words as landmarks of meaning, less on background knowledge and ignore syntax to a great extent. Furthermore, she notes that without a solid L2 language base learners are unable to use the metacognitive strategies developed in their L1.

A final crucial point is discussed by Hunt and Beglar (2005) concerning vocabulary learning and reading. They promote the involvement from the learner's part in activities that target implicit learning and the parallel use of explicit lexical instruction and learning strategies, including dictionary work. Their framework emphasizes extensive reading which gives the opportunity for meeting unknown vocabulary on repeated occasions, a necessary part of lexical acquisition discussed in the following section.

5.2.3.2 *The effect of repetition on vocabulary learning*

Studies in L1 and L2 incidental vocabulary learning have reported that the number of times an unknown vocabulary item is encountered in context also have a significant effect on the learning of the meaning of that item. The first research investigating the effect of repetition of L2 vocabulary acquisition was carried out by Saragi *et al.* (1978, cited in Webb, 2007: 47). They suggest that learners need to meet a word ten or more times in a reader in order to gain knowledge of meaning. They report great variability in their data, which is explained by the meaningfulness of the context and the degree of similarity of the word form in the L1 and L2. Horst *et al.* (1998), however, point out that the great number of words learned in the Saragi *et al.* (1978) study may be partly explained by the fact that the adult subjects have good skills for guessing vocabulary from context. Using a graded reader they also find great variability in the minimum number of encounters needed for knowledge gain. They point out that two factors have significant influence on vocabulary learning through reading, namely pictures in the book and the part of speech of the target words. Concrete nouns accompanied by pictures are the easiest to remember. Waring and Takaki (2003) suggest that over 20 encounters may be needed in order to gain partial knowledge of the word meaning, and be able to retrieve and use the given lexical item. In their study learners were

¹⁹ Note that the term *guessing* discussed in the literature up until the 1990's is not used by Nassaji, instead *inferring* and *deriving word meaning from contexts* are employed.

given the task to recognize and identify the meaning of words met in a reader three months earlier. They claim that eight or more times of encounter gives only a 50% chance to recognize the word form after three months, and the meaning of the words met less than 18 times were not recalled.

Webb (2007), while reviewing studies including Saragi *et al.* (1978), Hulstijn (1992), Horst *et al.* (1998) and Waring and Takaki (2003), concludes that previous results related to this issue cannot report conclusive findings due to two reasons. The first one is that previous studies fail to control the type of context in which new words are met. The second is that the great majority of research has focused on the meaning aspect of knowledge gain without studying the effect of repetition on orthographic, syntactic, associative knowledge and that of grammatical functions. What is clear is the fact that meeting a lexical item once in context or without context cannot lead to its long-time retention in most cases. If not learned through implicit or explicit strategies, it cannot become part of the learner's receptive vocabulary, and even less likely to be used in context.

The issues of implicit and explicit vocabulary learning will be touched upon in the empirical investigation in two ways. First, the amount of reading done in the target language, including the amount of reading related to university studies, will be investigated through a questionnaire to see whether it is possible to interpret implicit reading as a meaningful form of vocabulary gain. Second, questionnaire and interview data of a longitudinal case-study carried out with a group of subjects involved in the empirical investigation will target explicit vocabulary learning strategies. With these ideas in mind I now turn to the discussion of vocabulary in writing.

5.3 Relationship between vocabulary and writing

Similarly to the reading context, vocabulary is one of the key factors in written text production. However, while reading requires the use of receptive vocabulary, text production asks for the recall and use of vocabulary which is activated for a certain writing task. Before turning to the discussion of specific empirical studies employing the research instruments of lexical richness discussed in Chapter 4, it is crucial to summarize the role of vocabulary in written texts, with direct reference to the academic essay writing context. The next two sections, therefore, will review on the one hand, how vocabulary fits with other factors involved in text production; and, on the other hand, what the difference is between the views of text as a product versus text as a process and how this affects the focus on vocabulary from part of the writer and the researcher. The third part of this section will then discuss empirical studies which investigate lexical richness in student writing.

5.3.1 The role of vocabulary in text production

There are a number of components of the writing process, all of which need to be considered by the writer in order to arrive at clear, fluent and effective communication through the text produced, as illustrated in Figure 5.1. Vocabulary is only one of the components under the heading of word choice (including vocabulary, idiom and tone), combined with aspects such as syntax, grammar, mechanics, organization, purpose or writing, audience, the writing process and the overall content. Of course, the question emerges about how much attention can be or is dedicated to vocabulary choice during writing if so many aspects need to be taken into consideration. The answer is that we do not know. First, because there has been limited research concerning lexical choices during the writing process, second, because it is usually the written product that is under investigation, which cannot account for the choices made or the reasons behind them. In order to better understand the role of the above mentioned factors in writing, including vocabulary, two of the most often debated theoretical and methodological approaches will be discussed in the following section; writing as a process and writing as a product. Since the third major approach, namely writing as a

social activity (Johnson & Johnson, 1999) is not relevant in the academic context discussed in this dissertation, it will be not further explored.

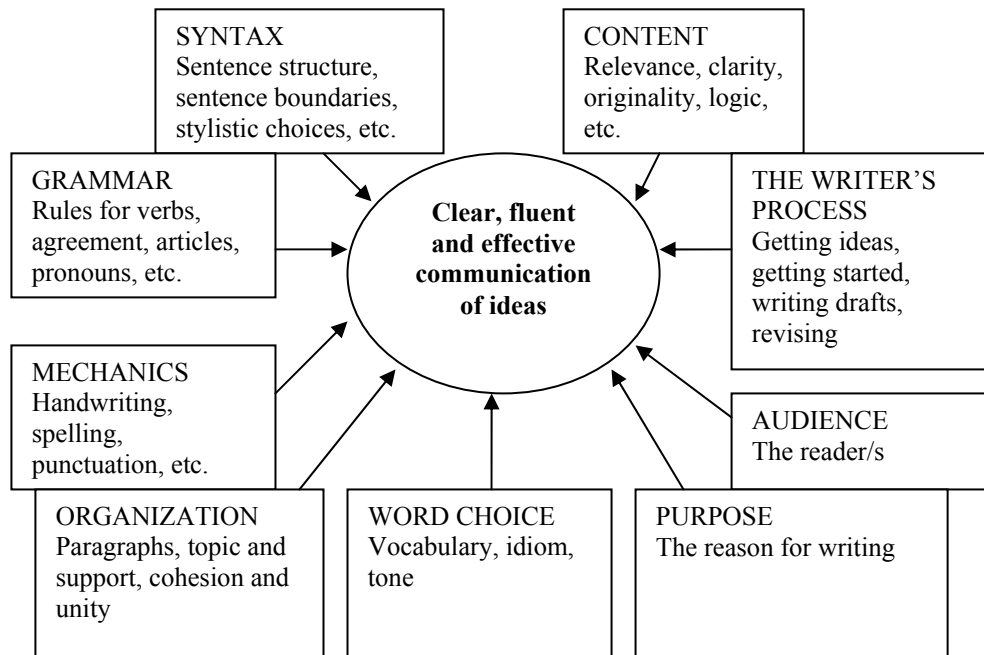


Figure 5.1 The component of writing (adopted from Johnson & Johnson, 1999: 344)

5.3.2 Writing as a process vs. as a product

In the early 1980s, Flower and Hayes (1981, cited in Witalisz, 2004: 15) designed a cognitive model of the writing process, which since then has remained the basis of further research (see Figure 5.2). Their model identified three major parts of the process, namely the writer's long-term memory, the task environment and the writing process itself, including a planning phase, the translation of ideas into verbal forms, and the reviewing and editing phase. The model underlines the non-linear nature of the writing process and the ongoing linguistic and cognitive processes. This model, however, was later criticized for not taking into consideration the expertise of the writer, as novice writers are more likely to be involved in a knowledge-telling process, while more skilled writers are involved in knowledge-transformation (Grabe & Kaplan, 1996, cited in Witalisz, 2004: 16–19).

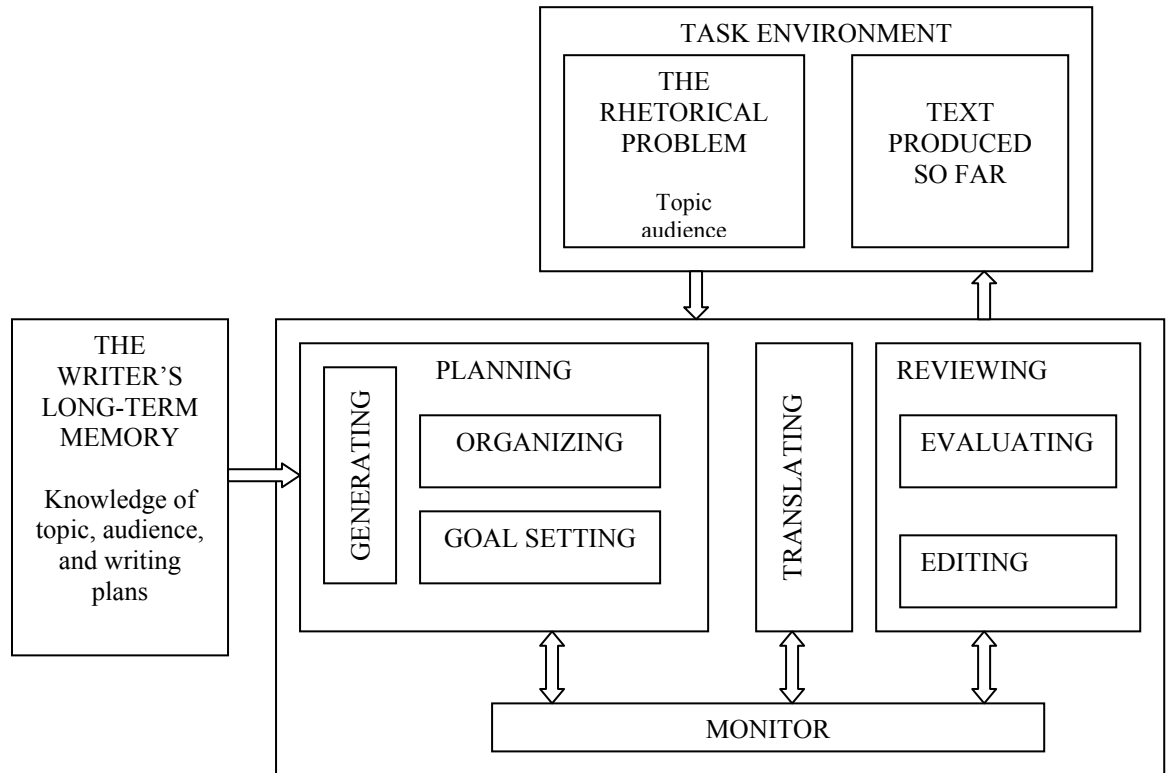


Figure 5.2 The Flower and Hayes writing process model (Flower & Hayes, 1981: 370)

It needs to be underlined that, while the process model is effectively transferred into the methodology of teaching writing to L2 learners (by stressing the importance of multiple drafts, different options of organization and feedback, various types of audience, self-discovery of the writer's own voice, etc.), it is less applicable to research in the field of writing for academic purposes for multiple reasons.

The first reason concerns the academic environment many of the essays used in research are produced in. As Horowitz (1986) voices, the process-oriented view does not take into consideration examination writing which is an essential part of academic studies, and, therefore, cannot be left without concern. Although exam papers may be regarded as "non-real" writing by advocates of the process writing model, students are often required to write under time-pressure which hardly ever allows for multiple drafts and thorough reviewing processes. This means that the idea of good writing as "involved" writing is not applicable in this case, as students are not likely to be involved in real knowledge transformation or be interested in the topic they are given. What he suggests is that "students rarely have a free choice of topics in their university writing assignments. Teaching students to write

intelligently on topics they do not care about seems to be a more useful goal than having them pick topics which interest them” (Horowitz, 1986: 143). Moreover, he suggests preparing students for multiple writing purposes, including exam writing, keeping in mind that in this case it is the product and not the process that is most often evaluated.

The second reason for being cautious with the advocacy of the writing as a process view is connected to this latest concern related to assessment. In research it is mainly the product that is evaluated and much less is dedicated to the assessment of the process. In the case of vocabulary, this means that the research tools applied to text analysis cannot take into consideration how and why those certain lexical items were chosen for the essays. It is undoubtedly difficult to combine both process and product orientation into the same research design and very often the one excludes the other. However, it is without question that even limited information on the process itself may give valuable information that helps us evaluate and understand the product. It is especially true in the case of vocabulary. For this reason the empirical investigation in this dissertation devotes attention not only to the vocabulary items used in exam papers written under time and topic constraints, and how this lexical use relates to the productive vocabulary knowledge measurable through tests, but also the importance paid by the students to vocabulary during writing alongside with factors such as organization or grammar. This will add to our understanding whether certain types of vocabulary are used by students because of difficulty in assessing less frequent vocabulary in their mental lexicon, or because vocabulary is not considered the most important of all the factors illustrated in Figure 5.1 that writers need to consider during text production. If the second one is the case, then students’ attention needs to be called to the importance of striving for the use of more sophisticated vocabulary that they possess but do not recall for use.

After this brief review of the role of vocabulary in the process and product of writing, studies using text tools employed for analysis of lexical richness of student essays will be discussed.

5.3.3 Previous studies of the lexical profile of learners' texts

In the last two decades the issues of L2 writing research have covered topics such as L2 writing theories, literacy development, reading–writing connections, research methodology, text interactions, writing assessment, curriculum and material design, and technology-assisted writing (Fujieda, 2006). What is surprising is the relatively little space dedicated to direct vocabulary assessment, as testified by articles in journals such as the *Journal of Second Language Writing* or *Written Communication*. Growing attention is dedicated to corpus building and corpus-based studies in L2, but these, as has been pointed out in Chapter 4, go beyond the scope of this dissertation. Instead of drawing large-scale conclusions of the vocabulary use of L2 learners, the attention in this section will be dedicated to text-based studies that use methodology discussed in Chapter 4, such as lexical profiling.²⁰ A great number of studies in recent years have focused on learners' vocabulary use, but with focus on the lexical errors of learners with different L1s (e.g. Augustín Llash *et al.*, 2005; Witalisz, 2004; Granger, 2003). Since the present dissertation does not intend to investigate the errors in learners' written production, the following section aims to review a few studies directly related to the question of lexical use in texts produced by students in secondary or tertiary education from the pool of studies directly or indirectly related to the broad issue of vocabulary and writing.

Morris and Cobb (2004) explore the potential of the lexical profile of essay samples as a predictive measure of academic success of foreign language trainees. Subjects (n=122) were of native or near-native-like proficiency in English in Quebec, with a variety of age and L1 background. A 300-word sample was extracted from entrance exam essays of each subject and analyzed with the help of the VocabProfile (Cobb, 2008b) used in the present dissertation. Academic success was defined in terms of grades received in two obligatory grammar classes. Morris and Cobb's data support the hypothesis that the vocabulary profiles of essays are good predictors of academic success of this graduate student population, when used together with other measures, such as grammar tests, interviews or academic records. What they conclude is that the lexical profiles "are simple to run, cost effective, and able to get at information that interviews and measures of declarative language knowledge do not reveal" (Morris & Cobb,

²⁰ For more on lexical issues in learners' corpus studies see Gardner, 2007; Szirmai, 2005; Granger, 2004; Horváth, 2001; Altenber & Granger, 2001.

2004: 85). Moreover, they point out that, although the correlations found between expressive vocabulary use and academic performance were moderate, they proved to be statistically significant in each case. They highlight the finding that students who had the richest lexicon, including academic vocabulary, showed the best thinking and metacognitive skills necessary for course work. The profiler proved to be a practical source of identification of at-risk students, as of the 14 drop-outs of the studied population, 11 had a high percentage points (above 88%) of the first 1,000 word families and a very low percentage point (below five) of academic vocabulary in their entrance exam texts. Their finding can be directly related to the issues under investigation and the student population used in the empirical investigation of this dissertation, therefore, will be discussed in more detail in the next chapters.

Muncie (2002) investigates the potential of the Lexical Proficiency Profile (based on a frequency count of the first 2,000 English word families) as means to assess the vocabulary growth during the writing process. The author analyzes three drafts of timed compositions of Japanese university students (n=25). Data show a significant increase in sophisticated lexical use between the first and the final draft which calls attention to the importance of encouraged vocabulary building during process writing.

Two studies carried out among secondary students address directly the question of the degree of correlation between results obtained on the Levels Tests and the lexical profile of texts written by the same students. In the first study, Laufer and Nation's (1995) data show that learners' lexical frequency profile correlates with the same learners' Productive Levels Test scores. They compare the result of EFL and ESL university students in New Zealand and in Israel. In the second investigation Laufer (1998) reveals a gain in vocabulary size during one year of L2 study. The author compares the results of the two Levels Tests and the Lexical Proficiency Profile of 16 and 17-year-old EFL high school students in Israel. Older students scored higher on both types of the Levels Tests. However, those who had better level test scores (both receptive and productive) did not use a greater variety of words in their essays (or have more advanced lexical frequency profiles). Laufer attributes these results to the probability that the subjects in the 1998 study used more high-frequency, safer and more familiar words. "The strategy of learners of using few resources (managing with as little as possible) can be referred to as a risk avoiding strategy, task simplification, or simply, taking the easy way out" (Laufer, 1998: 297). She also notes that the grading conventions in schools

in Israel emphasize correctness and rarely reward experimenting with lexically richer vocabulary. In these school instructors use conventional re-writing exercises with the result of providing pre-set format and content for essays. A similar report is given in Altenberg and Granger (2001) about learners' avoidance of using low-frequency words.

These two Laufer studies using the same research tools seem to give contradictory results. However, if we consider the explanation on the background of language teaching used in Israel it becomes clear that we cannot expect a great difference between the writing students do in two consecutive years. However, it is important to stress that Laufer (1998) and Altenberg and Granger (2001) formulated only assumptions on the vocabulary strategies students use, but unless the students are asked to report on their possible avoidance or substitution strategies, the results cannot be fully explained. Therefore, it will be a significant addition to already existing data to include a questionnaire in the data collection for the present empirical investigation that will report on the subjects' strategies of vocabulary use during essay writing.

The data available in the Hungarian context relating to vocabulary knowledge and lexical profile is limited to two pilot-size studies carried out by Lehmann (2003) and Doró (2007e). Lehmann (2003) compares English majors' (n=15) receptive vocabulary using a self-assessment 50-item test and the Lexical Frequency Profile of a take-home essay. The author does not find a clear correlation between vocabulary knowledge and lexical use in essays. This can be explained by several factors. First, the study is small-scale and exploratory in nature, which cannot account for the great variability of the test results. Second, the fact that the essays were not written under controlled circumstances significantly affect the profile they show, since we do not have any information concerning the writing circumstances (time, references, number of drafts, etc.). Although the methodology of this study is similar to that employed in the present dissertation, due the above discussed shortcoming this study will not constitute a significant basis of data comparison. Doró (2007e), in a pilot study for this dissertation, employs a subgroup (n=40) of the data gathered for this dissertation in order to answer questions about the relationship between receptive vocabulary knowledge and vocabulary use in essays. This study also investigates the effect of topic choice on vocabulary use. Data shows significant correlations between the investigated data pairs. The questions and results will be discussed in more detail in the result section of data analysis in Chapter 8.

As has been noted, within the significant interest towards L2 writing research relatively little attention is dedicated towards the text-based assessment of student writing in relationship to students' specific lexical proficiency. There is a great need to further explore some of the issues discussed above, with special attention to the Hungarian context, as this has been given even less attention.

5.4 Concluding summary

In the first part of this chapter a review of findings on text coverage, repetition and role of reading in vocabulary gain were given. Subsequent studies show that a minimal 95 to 98% coverage is needed in order to read fluently. If turned into vocabulary size, this implies a threshold level of 5,000 words and academic vocabulary in the case of academic texts, and a less demanding 3,000 words in the case of general texts. It was also pointed out that a sufficient knowledge of topic-related technical vocabulary is the base for academic studies, as they cover a relatively high percentage of the running words in university reading materials.

Extensive reading has been found to promote reading, writing, spelling, vocabulary growth, reading fluency and to improve motivation (Day & Bamford, 1998, cited in Hunt & Beglar, 2005: 39). It is to be shown whether the Hungarian student population under investigation employs the strategy of extensive reading, how much they read in connection with their studies and how much more they read for pleasure or other purposes. There is a clear gap in the literature in terms of Hungarian language majors' reading habits. There is only instructors' experience with students doing or not doing the assigned course work including reading activities. It also needs to be explored whether students meet the threshold level discussed in this chapter, as a possible lack of the minimum vocabulary size may pose a serious burden on the reading activities of these students.

The theoretical models of reading reviewed in this chapter show that vocabulary is both a basis for understanding texts and a source of new vocabulary. Based on the proportion of words known by the study population involved in this investigation, it will be discussed whether students are more likely to employ a 'top down' or a 'bottom up' strategy for reading. It will also have a direct bearing on their vocabulary development to see whether they are able to switch from limited, intensive reading of secondary schools to a multi-topic

and multi-genre extensive reading involved in implicit vocabulary building necessary for academic achievement. Alongside with the amount of reading down in the target language, a small group of students will be interviewed to see how they combined extensive reading with explicit vocabulary learning strategies and how these help their lexical development.

In the second part of this chapter the topic of vocabulary in writing was reviewed. It was pointed out that, similarly to reading, the lexicon is only one of the key factors involved. After an introduction to the views of writing as a process and writing as a product, some of the studies that employ the methodology used in the empirical investigation of this dissertation were reviewed. A clear need for the lexical analysis of texts which takes into account the writer's vocabulary knowledge was shown. University students, who encounter writing tasks of various kinds, including longer pieces of academic texts, should employ sophisticated vocabulary which reaches beyond the writing class into other courses and subsequently into real-world needs. This empirical investigation is necessary to explore whether English majors in Hungary are able to turn their vocabulary knowledge tested in various ways into use and how the task may affect this process. It should also be of interest to a wider audience how the assessment of writing as a product can be combined to the assessment of the process (the importance dedicated to vocabulary alongside with many other factors involved in text production) and how this information may help us in the evaluation of empirical data and in the methodology of teaching writing in an academic context.

PART III

EMPIRICAL INVESTIGATION

Introduction

As has been discussed in the previous chapters, vocabulary is a significant issue in second and foreign language research. However, while looking at the growing body of research on lexical proficiency and use, one may wonder what another study can add to the previously gained knowledge on learners' vocabulary. The aim of the present study is multifold, as it attempts to provide both theoretical and practical insight into measuring students' lexicon (receptive and productive vocabulary size, knowledge and use of lexical items of different frequency levels with or without contextual clues) while employing various testing methods. In designing this study, the following two major issues had to be considered.

On the one hand, as the previous chapters have shown, the multifaceted nature of vocabulary knowledge cannot be overlooked; however, at present there exists no test of lexical proficiency that could measure all the possible types of sub-knowledge or dimensions that are involved. All tests focus on one or a few aspects of the learners' lexicon, and many of these instruments have been under heavy criticism. I cannot accept strong opinions which suggest that data collecting instruments that target vocabulary as a separate constraint are of little value (e.g. Singleton, 1999). I do, however, see the common problem of using one method without closely examining what information it can provide us and how it relates to other methods of assessing learners' vocabulary. It is, therefore, of great importance that several instruments are used in this study, with the attempt to provide a clear reference as to what they are informing us about.

On the other hand, the reviewed literature has shown various, often seemingly contradictory results. This is not surprising in the light of the fact that studies employ a wide range of methodologies, including various numbers and types of subjects, instruments and data analyses, all of which have a direct bearing on final conclusions. It is, therefore, crucial to focus on a local student population if we wish to fully benefit from the practical outcomes of the present study. The present investigation aims to provide information about aspects of

Hungarian university students' vocabulary knowledge and language practice that we could, so far, only make assumptions about on the basis of the coursework they had been involved in.

On the whole, the added value of the present empirical investigation is the following: a) it employs various testing methods with clear reference as to what they aim to measure, and it assesses the relationship between test results, vocabulary use and background factors that may affect these results; b) it involves a student population whose vocabulary knowledge and use has not been thoroughly investigated; c) it analyzes test results as possible predictive measures of academic success; and d) it relates research questions and discussion to theoretical, research and practical issues in need of further evaluation both for the specific research population involved and in a wider context.

Following the previous chapters of literature review, in this third part of the dissertation I will present and discuss the design of the empirical investigation in detail. The following chapters (Chapters 6 and 7) will provide a discussion of the research questions and methodology, before turning to the results and discussion of the findings.

CHAPTER SIX

RESEARCH QUESTIONS

6.1 Introduction

This chapter discusses the seven major research questions and the rationale behind them. After listing the questions, each one is discussed in detail in order to facilitate a clear understanding of the reasons behind them and the connection between the questions. Specific questions that help to provide an answer to the questions are also treated below.

6.2 List of research questions

In the light of the issues raised in the literature review and considering the findings of a series of pilot studies discussed later in Chapter 6, the following major research questions are addressed in this dissertation:

1. What is the vocabulary knowledge of first- and third-year Hungarian university English majors as measured by receptive and productive tests?
2. What is the influence of the amount of academic experience and the amount of language practice (as defined by the amount of time spent with English inside and outside of school and language learning background) on the learners' vocabulary knowledge?
3. What is the relationship between the knowledge of receptive and productive vocabulary of students as measured by tests? In other words, to what extent can scores on a receptive test predict scores on a productive test? Also, how do sub-scores and overall test scores relate to each other?

4. How can vocabulary test results predict academic success in terms of reading ability and successful passing of courses? In other words, first and foremost, do students meet the minimum threshold level needed to read general and academic texts? And also, can we set a minimum lexical proficiency level for students without which they are likely to fail in their English language medium studies?
5. What is the relative role of vocabulary knowledge (as measured by the vocabulary tests in research question 1) in the lexical richness of L2 written production tasks? In other words, does a larger lexicon lead to more sophisticated vocabulary use during written production?
6. To what extent can differences in the vocabulary use in written production be explained by a) students' experience of producing texts in the target language, b) students' stated overall text writing strategies, including their view of the role of vocabulary in text production and c) topic choice?
7. How does the lexical knowledge of English majors, as measured by vocabulary tests, change over one academic year?

6.3 Explanation of research questions

The first major research question, namely the vocabulary knowledge, as measured by three discrete instruments of varying formats, serves a descriptive purpose. As has been pointed out in the overview of the literature, an immense number of empirical studies have been published in the last two decades that report on results in a variety of educational settings, with learners of different L1 backgrounds, second or foreign language settings (Nation, 2001; Schmitt, 2007), but we have a very limited amount of published data related to the vocabulary knowledge of Hungarian students. Three groups of students are targeted in this study, selected with the specific aim of assessing them at major stages of their English-language studies. Descriptive data analysis is carried out with a) students who have just

entered university, to see what the vocabulary knowledge that they bring with them is; furthermore, with b) students at the end of their first year; and c) students at the end of their third year of their English major, as they are two turning points in students' academic studies carried out at the study site. The descriptive data will serve as the basis for further investigation targeted by the other research questions. From a theoretical and assessment point of view the seemingly simple dichotomy of receptive–productive vocabulary discussed in Chapter 2 will be challenged with the use of one test which is called receptive and two tests which are called productive. It will be discussed how test results may reflect the understanding of what these tests are actually measuring.

The second major research question works with the data described while answering the first research question. These data elicited through tests can be fully interpreted only when looking at some factors other than simply the students' given year of studies at the university. This question is important to be raised, since students within the same years are expected to show different language learning backgrounds, amount of time dedicated to the target language, amount of time spent in an English speaking country. All these factors are expected to have an influence on vocabulary knowledge, but their degree of importance is yet to be shown. This investigation is significant, as it aims to support the importance of treating subject populations not necessarily as homogeneous groups of adult EFL or ESL learners. In many studies, indeed, the reported information on the subjects is limited to the indication of their L1, the year of English studies or the type of current education. It is expected that other factors may have direct implication on the results and their interpretation.

The third research question targets the use of parallel testing methods, which, as has been discussed, provides a fuller picture of students' lexical competence. This question explores the correlation between scores obtained on the three tests to examine how results on one test can be extrapolated to other aspects of lexical proficiency of the same subjects. A limited number of previous studies have used the VLT and the PVLТ parallelly to explore the gap between the receptive and productive vocabulary knowledge types they are measuring, but these have used a limited number of subjects and were administered in different contexts (e.g. Laufer, 1998; Laufer & Paribakht, 1998). Especially in the case of Laufer (1998), the various gaps found between the two subject groups is difficult to interpret due to the unclear nature of influencing factors such as extra curricular activities in the target language, general

English proficiency, attitude or motivation. There is a clear need for a theoretical justification of how the two parallel testing instruments work and whether the same or different correlations are seen in the case of groups of subjects who differ only in their vocabulary size. In order to explore this question, the 342 subjects will be divided into new study groups according to their receptive test results. A similar need for clarification is needed in the case of the Lex30 which has previously been reported only against the PVLТ, but not the VLT (Fitzpatrick & Meara, 2004). It needs to be shown how the Lex30 association test with its context-free testing of partial word form or meaning fits into the complex nature of receptive and productive vocabulary sub-knowledge types. Since the two instruments of controlled vocabulary knowledge include target words of various frequency levels, the question of the predictive nature of specific levels is also investigated. From a more practical point of view, correlation matrixes can give us information on how to interpret sub-scores and overall scores. Since usually only one testing method is used in academic context for practical reasons, having parallel tests under investigation is a crucial point in this study. Only after we investigate the correlations of parallel test results of a larger student population, can we extrapolate results on one test to other types of information in future testing of similar study populations, as it has been shown that results obtained on small-scale studies in a variety of contexts are difficult to summarize and generalize (Read & Chapelle, 2001). Furthermore, the correlations of sub-scores on the Levels Tests with overall test results will also be of practical value, as it will be possible to see how the administration of only certain levels of the test can be interpreted as a more general evaluation of vocabulary knowledge²¹.

The next research area, targeting English majors' academic success influenced by their lexical knowledge, is mainly diagnostic in nature. Although research has shown a strong link between vocabulary knowledge of all types, including general language proficiency and various language skills, as discussed in the previous chapters, little is known about how lexical knowledge can determine the success of activities done in the target language, more specifically academic studies carried out in the L2 (Laufer, 1992; Morris & Cobb, 2002). Moreover, there is a strong lack of published knowledge in the case of the EFL and not ESL context. This fourth major research question is also of immediate practical nature, since it

²¹ A practice, for example, in the case of the proficiency language testing of English majors at the University of Pécs, as reported by Lehmann (personal communication).

emerges from the observation of Hungarian university instructors according to which students who wish to follow academic studies as English majors apply for the university with different levels of overall language proficiency, including vocabulary knowledge. It would be desirable to know how vocabulary test results can predict academic success in terms of reading ability. Since written input and output are key factors and source of information for university students, it is crucial to investigate whether they meet the threshold level for reading set by authors discussed in the literature review (Laufer, 1997; Nation, 2001; Hirsh & Nation, 1992; Hu & Nation, 2002). A minimal reading vocabulary is defined for all three groups of subjects and it is calculated how many of them reach this requirement. On the basis of this threshold level and the amount of reading students do, it will be explored how reading, content knowledge and lexical knowledge gains are related.

Furthermore, academic success can be defined in terms of students' completion of the courses prescribed by the syllabus. The test results of the incoming first-year students involved in this study are also analyzed in terms of final grades in their first semester language courses, as these are minimum requirements for their studies. These students need to pass four of the following language classes of their choice during their first two semesters: Use of English, Reading Skills, Writing Skills, and either Listening Skills or Communication Skills. It is a tendency that many students fail these classes, and if motivational problems are not considered, it is more than probable that the lack of sufficient vocabulary and, therefore, inadequate understanding of oral and written input and the scarce quality and quantity of their language output are key factors in their failure. Those students who encounter great difficulty in general language proficiency classes are also likely to be unsuccessful in content classes. The second group of students tested at the end of their first year is investigated in terms of passing all their classes, including language and content classes. From a diagnostic and syllabus design point of view, the analysis of all three test results is expected to show the minimal lexical competence needed for these students in order to successfully complete their studies and identify a risk group who do not meet the lexical requirements and should go through an accelerated vocabulary building program at the study site.

While the first three research questions target vocabulary proficiency measured by tests and the fourth one investigates the relationship between productive test results and reading ability, the fifth area of research links receptive test results with vocabulary use in written

production. As has been discussed in Chapter 4, the three tests of vocabulary knowledge can be placed on a continuum from a controlled receptive test to a free production test, the latter one aiming at assessing productive vocabulary similarly to real language use. However, it is yet to be shown how productive test results obtained on the two productive tests relate to the actual vocabulary retrieved and employed while producing written texts. As has been noted in Chapter 5, university language majors, during their studies, are expected to produce texts of various length and genre, from brief test answers to research papers. It is of crucial importance for them to make use of their lexical knowledge, even when they do not have the immediate opportunity to consult dictionaries, written materials or look for any other kind of help. In order to analyze students' written vocabulary use, with focus on the lexicon readily available for them and retrieved for the specific task, essays written under controlled circumstance during comprehensive exams were collected. The importance of the use of texts not written at home is discussed more in detail in the methodology section. These texts are analyzed with the help of the same on-line research tool, the VocabProfile (Cobb, 2008b) used to assess Lex30, in order to gain information on the type-token ratio, lexical density and lexical frequency profile. It will be under investigation whether these three measures of lexical richness show better results for students who have a larger productive vocabulary measured by the Productive Levels Test. Yet, as has been discussed in detail in Chapter 5, text production involves many parallel features, all of which need to be paid some attention. The assessment of essays as written products will be combined with the assessment of writing as a process through questionnaire data on the importance dedicated to vocabulary during writing alongside with text organization or grammar.

The sixth research question is in close connection with the previous one, which compares the lexical richness of essays with the vocabulary proficiency of the authors. Factors are investigated which may have a direct effect on the lexical parameters of the texts. These are students' experience producing longer texts in English and some of their text writing strategies such as the avoidance of vocabulary items not fully known or language used for essay planning (information gained as part of a questionnaire)²². As another sub-question, it is also explored how their view of the importance of lexical choice (compared to grammar,

²² Váradi's study (1980b) on message adjustment confirm the theoretical presupposition that L2 learners tend to tailor their message to the linguistic resources available to them and adjust their ends according to their means or adopt the strategy of risk avoidance.

organization and spelling) affects their use of less frequent words in their essays. Indeed, it needs to be analyzed whether those who claim to pay the most attention to vocabulary, do use a wider range of lexical items while writing. As a final area within this research question, the relationship between topic choice and vocabulary is investigated from two different points of view. First, it is explored whether lexical proficiency plays a significant role in topic choice made by the subjects. This investigation is made possible thanks to the fact that at the exam when essays were gathered each group of students was given a choice of essay topics. It is investigated whether students with higher and lower lexical proficiency show a preference for any of the topics offered to them. Second, the effect of topic is investigated from the opposite direction by exploring whether certain topics or genre choice pre-determine the lexical richness of the essays. Answers to this last area of research will be of utmost importance for essay writing instruction and exam task design.

The final research question of longitudinal nature emerges from the need to gain information on how the lexical proficiency measured by tests change with time. Since the main research aim of the dissertation is that of assessing the lexicon of university students at important stages of their studies in order to see what their knowledge is, what influences this lexical proficiency, what this can inform us about their reading and writing ability, the research design was set up for a mainly cross-sectional study. Yet, the question may emerge whether the difference between the results of one major group and another can be explained by a continuous gain in vocabulary with time or rather with a selection of more proficient students in the upper years. It is of both theoretical and practical interest to see how real-time changes in vocabulary happen compared to the apparent-time changes investigated in the previous research questions. Keeping in mind the fact that overall group results often mask individual differences and development, a set of 15 first-year students were selected from the first group to track the changes in their lexicon across the period of one academic year. Following the initial testing in the first weeks of their university studies, they retook all three tests of vocabulary knowledge at the end of the academic year. Pairs of all test results for each subject are analyzed and explored with the help of information elicited with a background questionnaire and a structured interview in order to see what changes happen, what the direction of the change is, and how the degree of change alters with test formats and individuals.

6.4 Concluding summary

The research questions outlined above enable us to explore not only the vocabulary size of a given student population, but to investigate vocabulary knowledge from different angles. Multiple testing methods help us to get a broader picture of students' lexical proficiency and use. It will be possible to correlate various test data with the reading threshold vocabulary set by Nation and colleagues (Na & Nation, 1985; Hirsh & Nation, 1992; Hu & Nation, 2000) and the lexical richness of text produced by students. By doing so, we would be able to identify risk groups of students who do not meet the minimum requirement to successful academic studies. Although lexical proficiency is not the sole contributor to academic success for language majors, it is unquestionably a key component. Furthermore, the background information collected from students intends to facilitate our understanding of the reasons behind variability between students' performance. With these in mind, I now turn to the discussion of the research methodology.

CHAPTER SEVEN

METHODOLOGY

7.1 Introduction

After a detailed introduction into the research questions in the previous chapter, the following discussion intends to explain the rationale behind the choice of participants, the relevance of the research instruments, the method of data handling and analysis. This chapter also includes a section that explores the observations made during the several stages of piloting that further explain certain choices made for the final research design.

7.2 Participants

All participants in this empirical study are students at the University of Szeged, majoring in English and/or American Studies. For the sake of this research the few participants that are American Studies majors in the two more advanced groups will also be referred to as English majors as all their language seminars and the majority of their content classes overlap to the extent that they can be treated as one homogeneous group. The student population consists of the following three major groups: Group 1, the incoming first-year, Group 2, the end-of-the-first-year and Group 3, the end-of-the-third-year, with a total number of 342 subjects. Only eight other test-takers were excluded from data analysis due to incomplete results.

7.2.1 Major groups of participants

In the following sections I shall review the characteristics of each group. A summary of the information related to the major participant groups (year of studies, number of subjects included, time of data collection) is reported in Table 7.1.

Table 7.1 Major groups of participants

Group	Year	Number of students	<i>Time of data collection</i>
1	Incoming first year	148	September 2006
2	End of the first year	101	April–May 2006
3	End of the third year	93	April–May 2006 and April–May 2007
Total		342	

7.2.1.1 Group 1 – Incoming first year

Group 1, the largest one involved in this study, includes 148 students entering the university as English majors in September 2006. Their English proficiency ranged from lower-intermediate to advanced, as shown by the language exam certificates they reported on in the short questionnaire administered during data collection (for discussion see Section 7.2.3). They had an average of nine years of previous English studies, the lowest number of years being two and the highest number being twenty. 39% of the group (n=57) reported a stay in an English-speaking country, 26 of them indicating a longer than one month stay. These data support the general observation of instructors that students within the same year have a great variability in their language learning background and proficiency.

7.2.1.2 Group 1f – First year follow-up

This sub-group was selected from Group 1, the incoming first-year students, with the aim of assessing the change in their test results over one academic year. All 15 students were first tested in September 2006 and were retested in April and May 2007. At the time of test retake, they were enrolled in a Communication Skills seminar. Participation selection proved to provide a heterogeneous group including students with both low and high test scores. A detailed discussion of each student's language learning background and target language use will be provided when answering research question number 7, as these aspects are expected to have a direct bearing on the change in the test results.

7.2.1.3 Group 2 – End of the first year

101 students at the end of their first year were included in this group. They participated in a testing session in April and May 2006 on a voluntary basis. Students had an average of ten years of previous English studies (just one year more than Group 1, which means a similar background if we consider that they are a year ahead in their studies). 77% of them reported no previous stay in an English-speaking country, and only 11% indicated a longer than one month stay.

7.2.1.4 Group 3 – End of the third year

93 subjects at the end of their third year were involved in this group. Due to the data collection methodology explained in the next section, Group 3 consists of two subgroups, each containing approximately the same number of students (45 and 48, respectively). Group 3a was tested in April and May 2006 alongside with Group 2, and Group 3b was tested a year later. As one main group they reported on an average of twelve years of English studies (similarly to the previous two groups, with the addition of the two or three extra years at the university). They proved to be the ones with the most experience in an English-speaking country, which means that 21 of them had had a one to five month stay, while 12 of them a longer than six month stay in an English native speaking environment. There is no statistically significant difference between the two sub-groups in terms of years of English studies or time spent in an English-speaking country.

7.2.2 Rational behind participant selection

It needs to be clarified what led to the inclusion of these particular students in the data collection, as the decisions made had a direct influence on the entire data collection process. After a year of ongoing piloting with students in secondary and tertiary education (discussed below in Section 6.6), the final data collection was planned to be started in April 2006 with first- and third-year students being at two key stages at their university studies. Two inclusion

options were considered. The first option, namely testing during class time, was excluded due to expected problems with arrangement, different class sizes, attendance in multiple seminars needed for detailed data collection, and the need for the presence of the researcher to ensure a very similar procedure for all groups, the crucial part of oral instruction and the inspection of the entire testing session in order to rule out incomplete tests or any other problems as much as possible. Therefore, the second option of independent testing sessions proved to be more plausible. Voluntary testing sessions were advertised among students offering a number of times they could choose from. First-year students at the end of their first academic year were expected to be relatively easy to involve in a study giving them feedback on their language knowledge; first, because they were used to being continuously tested throughout their first two semesters, and second, because they were preparing for the end-of-the-year proficiency examination. This proved to be the case, so two consecutive third-year groups were offered the same testing session to balance group numbers. It indeed turned out to be a positive aspect of data collection to have two subgroups of third-year students since it provided us with the opportunity to compare the results of these two subgroups. Incoming first-year students, on the other hand, could be tested in a much larger percentage, as testing for them was organized parallelly to a mandatory proficiency placement test, during the second week of their studies.

As the present investigation is cross-sectional rather than longitudinal in nature, the first three major participant groups are independent from each other. It needs to be clearly stated that students who participated in the study in Group 1 (incoming-first-year), for example, are not retested later as Group 2 (end-of-the-first-year). Also, an attempt was not made to perfectly balance the number of participants in each group as the analysis of data on the basis of the comparison between the three groups will only be treated in some of the research questions; moreover, several other groupings will be made on the basis of other factors, such as amount of language practice or test results in order to answer specific research questions.

For the follow-up study involving the last research question, a group of 15 first-year students were selected from Group 1, which is referred to as Group 1f. In order to insure anonymity of the subjects involved, no names of individuals will be referred to in this dissertation, instead, numbers will be used to identify subjects where necessary.

7.3 Research instruments

In order to explore the research questions discussed in the previous chapter, the following broad categories of data were collected:

- Three tests of vocabulary knowledge from all three groups
- Written product: an essay written during comprehensive language exams from Groups 2 and 3
- Questionnaire: background information about the subjects with particular reference to their language-learning experience and practice with reading and writing in the target language. A short questionnaire was used with Group 1 and a more detailed questionnaire with Groups 2 and 3 (for discussion on the rationale see Section 7.2.3)
- Course grades from Groups 1 and 2
- A structured interview for the case study, involving Group 1f.

The following sections will briefly discuss all the above mentioned categories of research instruments.

7.3.1 The three tests of vocabulary knowledge

Of the tests discussed in Chapter 4, the Vocabulary Levels Test and the Productive Levels Test were primarily selected for empirical data collection of this dissertation for the following reasons: they had been successfully used for years in vocabulary research, they are easy to obtain, and can be administered together using no more than one hour. Furthermore, they were designed by the same researchers and test the same vocabulary frequency bands, therefore data obtained on the two tests can be directly compared and correlated. The Lex30 test was added to these two tests in order to test other types of vocabulary knowledge. Its different format and knowledge types it assesses were expected to give a different insight into the receptive–productive vocabulary dichotomy discussed in Chapter 2. The advantages of adding the Lex30 test to the research instruments are the following: it is easy to administer and time efficient, it appears to be a game-like task rather than a test (where students are less conscious about being tested on their vocabulary knowledge), it provides some information

about the depth of vocabulary knowledge and the mental lexicon, and the rich data can be quickly analyzed using a computer text tool.

A comprehensive summary of the testing instruments in terms of format, different knowledge frameworks and major underlying mental processes was discussed in Chapters 3 and 4. It is important to review them here, in order to provide a clear reference. Table 7.2 below shows that the three testing instruments are different in test format (multiple-choice, sentence completion and association), which implies that they involve different underlying mental processes during test taking (recognition, aided recall and unaided recall). The table also summarizes the word knowledge frameworks or dimensions involved in the tests in light of three major frameworks. The first one is Read's (2000) vocabulary assessment dimensions (see in detail Section 4.1). While all tests are discrete, as they measure vocabulary as an independent construct, they differ in terms of selective–comprehensive nature and context dependency. Similarly, in light of Laufer's (1998) categorization, the tests can be called passive, controlled active and free active, referring to the receptive and productive nature of the vocabulary they test. In terms of Nation's (2001) vocabulary knowledge categories, it is shown in the table that some of the assessed knowledge types overlap in the tests, while certain others are tested only by one of the instruments. The heterogeneity of the testing instruments chosen for the present empirical investigation is a major strength of this study as they enable us to see vocabulary in a wider context and not restricted to one instrument which is the case in many studies (For the tests refer to Appendices A, B and C)

Table 7.2 Comprehensive summary of the testing instruments: Vocabulary Levels Test (VLT), Productive Vocabulary Levels Test (PVL) and Lex30

	VLT	PVLT	Lex30
Test format	Multiple choice	Reconstruction of partially deleted words in sentences	Association
Major underlying mental process	Recognition	Aided recall	Unaided recall
Read's (2000) test dimensions	Discrete Selective Context-independent	Discrete Selective Context-dependent	Discrete Comprehensive Context-independent
Laufer's (1998) word knowledge	Passive	Controlled active	Free active
Nation's (2001) word knowledge	Receptive written form Receptive meaning	Productive written form Receptive grammatical form and position Receptive and productive meaning	Productive written form Productive meaning concept Productive meaning association

7.3.2 Written production tasks

Written production was elicited from Groups 2 and 3 during a formal exam situation by giving students topics for expository or argumentative essays. This essay writing was a regular part of students' Academic English proficiency exam at the end of their first and third year, therefore, no additional data elicitation sessions of this type had to be arranged. The students were given the opportunity to choose between topics and were expected to show their best as in any formal exam situation. Also, the same amount of time was given to each student for drafting and revision within the same group, 75 minutes for a 220–250 word essay for Group 2 and 90 minutes for a 300–350 word essay for Group 3. No supplementary material was allowed to be used. Group 2 was offered four prompts, two expositions and two argumentations, each topic consisting of an approximately 2–3 sentence prompt. Group 3a was provided a choice of three prompts, all argumentative essays, with two argumentative sides to choose from. This meant six choices altogether. Group 3b had a choice of two prompts, as they were found to offer a selection of four choices. For Group 3 the tasks consisted of an approximately 3–4 sentence prompt on the topic of discussion and a one-sentence instruction of the argumentative sides to take. For the prompts used for data elicitation, refer to the Appendices.

Although using exam papers for written production may seem to present problems of representativeness, this type of data selection proved to have the following advantages: controlled data collection circumstances (same amount of time for planning and revision, no reference materials), academic genre (including expositions and argumentations) and similar topics. Also, an exam situation pushes students to show their best, unlike essays written in less controlled circumstances. Piloting among university students showed that the amount of time they dedicate to writing homework assignments and the outside help they use (dictionaries, native speakers, sample essays, and the Internet) greatly vary, which makes these types of essays a less valid indicator of the students' actively available vocabulary for written academic work (for more details on piloting refer to Section 6.6).

7.3.3 Questionnaire

The use of a questionnaire is of utmost importance to provide detailed information about the student groups. Simply knowing that subjects are, for instance, first-year students is not sufficient. If we had no information on their language learning background and experience with reading and producing texts, the data collected would be difficult to interpret. In order to obtain information on the subjects' language background, the amount of time spent with the target language and their reading and writing habits, a questionnaire was developed using Dörnyei (2003) and drawing on already existing questionnaires used by myself in previous studies (e.g. Doró, 2006) and during the piloting phase, involving university students enrolled in essay writing classes. The questionnaire used for piloting with high school students was modified in order to fit the written academic work university students are involved in. The questionnaire used with Groups 2 and 3 has the aim to report on their previous English studies (number of years English has been studied, number of months spent in an English-speaking country), intensity of current English studies (amount of time spent with English, amount of reading and writing done in English), and the way of dealing with vocabulary during essay writing (attention dedicated to vocabulary among other factors involved, reference materials used, compensation strategies of vocabulary use when no outside help is given).

Once the questionnaire for this research project was designed, its validity was checked to ensure that the questions would be understood and interpreted the same way as intended. University students were asked to fill out the questionnaire using the think aloud method (see, for example, Kormos *et al.*, 2002). Students reported no problems interpreting and answering the questions. The majority of the questions were changed since they were administered during piloting. The questions were given in Hungarian to facilitate the reading and filling out of the questionnaire and to avoid problems due to misunderstanding. The questionnaire and its English translation are seen in Appendix D.

The following paragraph intends to review each of the eleven questions used to elicit background data from Groups 1 and 2. Question 1 asks about the number of years English has been studied by them. Question 2, referring to the number of months spent in an English-speaking country, is included to identify those who have spent a significant amount of time in a native environment, which is expected to have a direct impact on students' vocabulary

growth. Both questions 3 and 4 ask students about the amount of time they spend with English per day, but probably there is a difference between weekdays when they attend classes and weekends when only individual work is done; therefore, the variable is split into two questions. As research has shown that a significant amount of vocabulary learning happens through extensive reading (e.g. Waring & Nation, 2004; Zahar *et al.*, 2001) and that reading is also a necessary basis for writing (Tsang, 1996), three questions on the amount of reading done by the students are included. Question 5A targets the total number of pages read in English per week, while in question 5B students are asked to indicate how many of the above number of pages are related directly to their academic studies. Question 5C is interested in the degree of difficulty students face when reading books or articles as part of class work. The four-point scale list ranges from the answer “I almost never have problems” to “the majority of the selected readings are too difficult for me; therefore, I do not read them”. As it is expected that the amount of practice of texts production in the target language may have an impact on the writing quality of students, this variable is also targeted in the questionnaire. Parallely to question 6A, which asks about the number of times students produce longer texts in English, the same question is formulated for the Hungarian context (question 6B), as it is expected that students produce a limited number of texts even in their mother tongue. Questions 7–11 ask students to report on the strategies they use during written production. In question 7 students need to indicate the order of importance between some elements involved in text production, such as text organization, sophisticated vocabulary, grammatical accuracy and spelling. While in question 8 students can indicate the type of help they refer to during essay writing, questions 9–11 require subjects to report on the compensation strategies they use when they cannot use any outside help during writing, e.g. in an exam situation. The options are designed in a way that students are forced to choose between the proposed two or three options and cannot avoid the answer. The last two questions ask subjects to indicate the strategy they use when they cannot find a word/expression that would best fit the context or are not sure about its spelling. Answers to these questions may help interpret the reason for using high frequency words in a written production task even when students have a large size of vocabulary.

This longer version of the questionnaire referring to academic work and the amount and quality of English-related activities was administered to Groups 2 and 3. Group 1, the

incoming first-year students, were not asked to fill out this questionnaire, as many of the questions would have been irrelevant to them, as most of them come directly from a secondary school with limited experience of extended reading and writing tasks in English (Doró, 2006). However, they were asked to indicate the number of years of previous English studies and the amount of time spent in an English-speaking country. They were also asked whether they had any (intermediate or advanced level) language exam certificate in English and what kind of extra-curricular activities done in English they were engaged in. The first two questions are parallel with questions 1 and 2 of the questionnaire used with academically more advanced students, therefore, allow for direct comparison. The last questions provide some general picture about their activities carried out in English (see Appendix D).

7.3.4 Course grades

As vocabulary is expected to play a significant role in academic achievement, language seminar grades for the first semester of incoming first-year students and a general fulfillment of all the first-year syllabus by the end-of-the-first-year students will be investigated and correlated with their assessed lexical knowledge. The completion of a number of language seminars is a requirement for all first-year English majors at the study site. In case of failing grades, students must retake the same seminar in the second semester. An insufficient number of successful seminar grades means an incomplete academic year and students are not allowed to take the comprehensive language exam at the end of the term.

7.3.5 Structured interview

A brief face-to-face interview followed the retesting of Group 1f. During the interview students were asked to further report on some of the questionnaire data on language learning background and the individual interview sessions also enabled me to discuss target language use and vocabulary learning strategies with students. Subjects were also interviewed on their test taking experience, whether they found the tests difficult or unclear and how much gain in receptive and productive vocabulary size as measured by the Levels Test they predicted before discussing their scores and specific answers in detail.

7.4 Procedure

All subjects participated in a similar procedure under the guidance of the researcher. This was done in order to ensure the equality and comparability of the data collected at various sessions, which meant that I was present at all of them and also did the evaluation and data handling myself. Testing sessions were organized in the spring and fall of 2006 and in the spring of 2007, as shown in Table 7.1. All instruments were administered under controlled circumstances. Students were seated far enough from each other in order not to be able to influence the results of other students. The same test battery was assigned to all students at various testing sessions. Timing of the test battery was piloted in order to assure for test completion, minimize incomplete results and balance for fatigue and attention loss. The three tests were administered starting from the most demanding one (as shown by piloting), the PVLТ and finishing with the least demanding and controlled one, the association task. Subjects were first asked to do the PVLТ, then the VLT for which were given a maximum of 60 minutes. These two tests were photocopied on A/4 size paper each, and stapled together. When these test papers were collected, the association test was handed out together with the questionnaire in Hungarian. Students were required to work on the first two tests for at least 50 minutes which restricted the number of incomplete or carelessly filled out tests. The association task was introduced on a separate sheet of paper. No time limit was given for this task, but approximately 15-20 minutes were used by the test-takers. Students were instructed to give responses in English, possibly excluding proper names, multi-word phrases and abbreviations.

In order to minimize bias due to fatigue and to avoid response order contamination, test items from the different bands were systematically mixed for the VLT and the PVLТ, as has been reported in other studies using these tests (Xing & Fulcher, 2007; Tschirner, 2004). This proved to be useful, as students could not see and were not influenced by the growing difficulty of items of the original tests; moreover, they could not pinpoint items that should be easy or difficult because of their place in the test.

7.5 Data handling and scoring

The scoring method for quantifying each variable was the following:

a) *Vocabulary Levels Test*. Scores were given for each correct answer suggested by the key. One point was assigned for each correct answer, leading to a maximum of 30 points at all six frequency bands.

b) *Productive Vocabulary Levels Test*. Items were considered correct and awarded one point if students gave the word and part of speech provided by the key or another word that fit the context, even if with small spelling mistakes. This method balanced for test sensitivity and allowed for partial knowledge, as in the case of the association task and the essays. For evaluating alternative responses, native speaker university instructors were involved. The following sentences show some examples that were accepted reflecting the partial knowledge required and some that were not (the first option is the one given by the key, the second word is an alternative solution written by students; the word in brackets in sample sentence c) was not accepted, the other solutions were awarded points):

a) *There are several misprints on each page of this text/test;*

b) *Pupils/pupil must hand in their papers by the end of the week;*

c) *Many people in England mow the lawn/(layer) of their houses on Sunday morning.*

As it is seen from these examples, singular or plural forms of verbs or nouns were both accepted (b), but the use of words which led to strange or unacceptable sentence meaning were not given points (c). Many similar examples to the last one seem to indicate that some students had no clear understanding of the context provided by the sentence or of the sentence they formed with the help of the word they inserted.

c) *Association task*. The words written in the association task were put on text files for each student, forming small text-like corpora. While doing this, spelling mistakes that did not distort the understanding of the responses were corrected. Each student's responses were processed using the online text analyzer tool called VocabProfile (for detail see Chapter 4), which breaks down texts into frequency bands. Following the test designers' scoring method (Meara & Fitzpatrick, 2000), 0 points were given for the most frequent 1000 words, but also

for missing, illegible or non-existing English words, proper names and numbers, as these groups of responses would have distorted the results by being classified as rare words. All other responses scored one point, up to a maximum of 90 points for the 30 times 3 responses.

d) *Essays*. The hand-written essays were typed into computer files. Similarly to the association task, spelling mistakes were corrected in order not to be classified as low-frequency words by the text analyzer. Grammatical or stylistic problems and punctuation were not modified. Unlike in Laufer and Nation (1995), lexical errors were not excluded from the texts for the following reasons: the problematic nature of what should be classified as error and how much of the text around it should be deleted, also, the texts themselves were too short to allow for extensive deletion. On the other hand, non-existing English words (e.g. *embetter* instead of *improve*) and proper names were excluded from the analysis. As type/token ratio and lexical density measures are sensitive to text length, the essays were standardized in the following way: the first 200 words (excluding the group of words discussed above) of each essay written by students in Group 2, and the first 300 words of each essay produced by students in Group 3 were taken as the basis for investigation. The standardized texts were analyzed with the help of the VocabProfile in order to calculate the type/token ratio, the lexical density and the vocabulary frequency profile (VFP) of the texts. Although the VocabProfile gives information only on the item level, advantages of using it for text analysis are the following: the original software tool (RANGE) was designed by Nation who is also the author of the Levels Tests, therefore, the basic frequency lists are the same for all instruments. The online text tool is available for free and is easy to use. The texts do not necessarily have to be converted into text files nor need to be lemmatized. For this study, instead of using a four-percentage profile (first 1,000; second 1,000; academic; and off-list words), a condensed profile is calculated, as discussed in Laufer (1992, 1998). Following her method, which she finds easier to use for studies involving large number of texts, the percentage figure used here shows the ratio of the first 2,000 most frequent word families, as calculated by the VocabProfile. Usually a very high proportion (around 90%) of the texts in this study fall into this category, while there is a restricted number of academic and low-frequency words. Due to the careful data pre-processing, including the standardization of spelling and the exclusion of non-English words, the off-list category includes only low-

frequency words, and no problematic items or names that would be categorized as non-frequent if raw texts were analyzed.

For statistical purposes data were analyzed using SPSS 11.0. For all data analyses, statistical significance was set at $p < 0.05$, always using two-tailed tests. Smaller p-values are indicated separately.

7.6 Piloting of the research methodology

In order to test the feasibility of the data collection methodology among Hungarian students and also to test the possible methodology for data analysis, pilot studies were carried out. The pilot data collection and data analysis provided valuable information concerning what steps needed to be taken during the final project design (e.g. the systematic mixture of the levels in the VLT and PVLТ, and clearer instructions for Lex30).

All measurement instruments and procedures were piloted with secondary school students in order to see what the skills and knowledge students are likely to bring with them to the university. The pilot data collection suggested that the instruments work well with this age group and school setting. No major problems were reported by the participating instructors. The following section aims at summarizing the data collection and the answers to some of the research questions addressed during the piloting phase. Due to the different proficiency level and academic background, not all main research questions will be dealt with in detail. Findings related to the lexical analysis of text produced by secondary school students have been reported in Doró (2006).

In order to test how the VocabProfile as a measurement instrument works with essays written by secondary school and university students, production tasks in different genres were collected from these two groups. Secondary school students were given a free writing production task with picture prompts (B. Fejes, 1981; Doró, 2001). The same task was repeated as homework assignment to trace the supposed quality improvement between a controlled task when no outside help was provided and a free homework task when room for planning and using reference materials was given. The texts written by a group of first-year students were put on computer in text format and analyzed by the VocabProfile program. The data collection method generated good essays from students, although the narrative genre did

not ask for the use of a high proportion of academic vocabulary. Surprising results were found when comparing the VocabProfile output for the two essays written under different circumstances: half of the essays showed a good vocabulary frequency profile, half of the essays showed unsophisticated vocabulary use. This difference was due to the way students wrote their essays at home: those who improved spent a significant amount of time (more than two hours) planning, writing, revising their written production, and using dictionaries. The other group did not dedicate much time and attention to the task and did free writing similarly to the classroom task. This suggests that it is difficult to predict whether papers written at home reflect the quality of writing done under controlled circumstances in terms of length and vocabulary use.

I also used the profiler to trace the differences between different genres university students write in. As it could be expected, the narratives included a much higher proportion of the first 500 and 1,000 band and very small proportion of academic words. Contrary to this tendency, argumentative essays written by third year students elicited a small proportion of the very high-frequency words and a much greater number of academic and off-list words. The VocabProfile also showed differences between two essays that were judged by the instructor as strong or weak in their vocabulary usage.

Thus, the pilot data suggested that argumentative and expository essays rather than narratives should be used for final data collection, as they force students to use a higher proportion of academic vocabulary. Also, it proved to be of utmost importance that the written production tasks were completed under controlled circumstances (with the same amount of time provided, no reference materials used, same genre, similar topics) to avoid the difference that was seen in the vocabulary used during the repeated, homework-type assignment of the first-year students as mentioned above. Therefore, the exam essays of first and third-year comprehensive language exams which were already part of students' regular study tasks, proved to provide an opportunity for data collection.

A trial version of the questionnaire to be implemented in data collection was used with first, second and third-year students as part of their writing class evaluation. The answers suggested that students' attention among the four aspects of written production (vocabulary, organization, grammar and spelling) is divided in slightly different ways during writing than in the case of secondary students. More attention was reported to vocabulary during essay

writing by university English majors than by high school students. Surprisingly, however, many university students reported that they engage in written production tasks only once a week, for classroom assignments. This provides only slightly more experience than that reported by high school students. These findings strengthened the need to include questionnaire data in the empirical investigation carried out during the final data collection sessions. The pilot phase inevitably raised questions like the following: how much time do English majors spend with English and how much do they read? As we have seen, input has a great influence on vocabulary growth, but unless learners have the threshold level of 3,000 word families to read successfully, they probably do not read enough, therefore, they do not gain valuable experience in how an academic text is written and what the kind of vocabulary is that would be desirable to incorporate into their own written tasks.

7.7 Concluding summary

This chapter has introduced the research methodology employed in the empirical investigation outlined in Chapter 6. It has been noted that participant groups and research instruments were selected in order to enable us to explore the research areas in as much detail as possible. This means that English majors' lexical knowledge and use are investigated at key points of their studies and from a variety of data sources. Another key aspect highlighted in this chapter is the fact the research instruments and data handling methods were piloted with university and secondary school students in order to rule out all possible problems that could emerge during the final empirical investigation. The pilot phases proved to provide valuable information that facilitated the research discussed in this dissertation.

A major strength of the above proposed methodology is the rich data it is able to elicit. Students' vocabulary knowledge is targeted from various angles, not limited to one or two aspects. Most of the elicited information can be translated into numerical data which enables in-depth, quantitative analysis. However, qualitative data are also explored to supplement and illustrate the major finding of the empirical investigation.

PART IV

RESULTS AND DISCUSSION

Introduction

The next four chapters report on the findings of the empirical investigation carried out among the Hungarian university student population described in Chapter 7. In Part III the research questions raised and methodology used in this research were introduced. In this fourth part of the dissertation, the research questions will be treated in order, one after the other in order, to facilitate the understanding of the specific subject groups involved, the methodology used, and the research areas investigated. For each of them aims, subjects, procedure, results and discussion will be elaborated, with detailed discussion sections at the end of the data analysis of each major research area. As has been pointed out in the previous two chapters, Studies 1 through 6 investigate the vocabulary knowledge and use of various student groups, therefore, the comparison between these groups in Chapters 8 and 9 is apparent-time in nature. In contrast, Study 7 is treating a small group of students with the aim of following the change in their vocabulary test results. This last study, discussed in Chapter 10, aims to explore real-time changes, and involves some qualitative analysis alongside the quantitative statistics employed for data analysis. Finally, Chapter 11 provides an overall discussion of the results with some specific reference to findings that emerge during the analysis of the empirical data.

CHAPTER EIGHT

VOCABULARY KNOWLEDGE AS ASSESSED BY TESTS

8.1 Introduction

As has been discussed in Chapter 2, vocabulary knowledge is a complex notion. It, however, may seem an easy task to distinguish between two main types of this knowledge and call them *receptive/passive* and *productive/active*. Nevertheless, as Waring (1997) and Melka (1997) rightly pointed out a decade ago, these notions should not be treated and accepted as given. Notwithstanding the large body of literature concerning learners' lexicon, we still know relatively little about how vocabulary is learned, how it is stored in the mental lexicon and how it is retrieved for use (Webb, 2007). A number of frameworks for vocabulary knowledge have been reviewed in Chapter 2, of which Nation's (2001) seems to be the one that recognizes the most the complex sub-categories involved in receptive and productive knowledge, including various aspects of meaning, form and use. As has then been noted in Chapter 3, to the complexity of the description of this knowledge, the problematic nature of assessing knowledge types or categories is added. Measurement instruments have been designed to assess lexical competence drawing on existing test formats and often claiming to test either *the* receptive or productive vocabulary of learners. What comes from this practice, as stressed by Waring (1997), is the restricted understanding and use of the terms *receptive* and *productive* without clearly referring to the underlying mental processes involved in the testing or the sub-categories of the types of knowledge assessed. As a reaction to this problem, researchers have tried to come up with more innovative vocabulary testing methods to target other than the subcategories already in the center of attention, such as form, primary meaning, translation equivalents, or to assess a variety of knowledge types at the same time (e.g. vocabulary knowledge scale). Others have seriously questioned the necessity of testing vocabulary as a separate constraint (see, e.g. Read & Chapelle, 2001; Bachman, 2000; Singleton 1999).

The present study is in accordance with the view that vocabulary assessment as a separate component of language proficiency is crucial and helpful for both research and

pedagogical purposes, if done with care. This implies clear reference to tested knowledge types, underlying mental processes of the tests and clear understanding of the study population's language learning background. The investigation involving the first three research questions aims to add to the existing knowledge on the relationship between various types of vocabulary knowledge by employing more than one assessment instrument. This enables us to reevaluate the seemingly clear relationship between receptive and productive vocabulary. Widely used tests of vocabulary size that assess partial knowledge of meaning and form (both receptive and productive) are treated together with a more innovative testing instrument that draws on the associative links between words in the mental lexicon, a crucial aspect of word knowledge employed during retrieval and use. First, descriptive statistics are presented for the three instruments; second, the role of possible influencing factors on the test scores is investigated; and third, correlations between the test and subtest results are explored.

8.2 Study 1 – Descriptive vocabulary test data

8.2.1 Aim

This section investigates the vocabulary knowledge of first and third-year Hungarian university English majors as measured by receptive and productive tests. This investigation is descriptive in nature and employs three data collection instruments different in format, type of vocabulary knowledge tested, sensitivity and underlying mental processes (as discussed in Sections 4.2 and 7.2). It is relevant to get an overall picture of the entire study group before we get to the discussion of the test results of different groups in Studies 2 and 3. Data gained in this study will serve as the basis for answering subsequent research questions.

The following major research question is addressed in this section:

1 What is the vocabulary knowledge of first- and third-year Hungarian university English majors as measured by receptive and productive tests?

While answering this broad research question, the following sub-questions will be explored:

- 1a Is there a decrease in scores with the growing difficulty of bands containing less frequent words?
- 1b Is there a growing gap between results in subsequent frequency levels (from 2,000 to 10,000)?
- 1c What is the degree of variability in test results as shown by standard deviation figures?
- 1d Does variability change in the upper frequency levels in the case of the two Levels Tests?

8.2.2 Procedure

As the first research question has the aim of giving an overall descriptive analysis of the data for the given university student population, all three large groups of students described in Chapter 7 were involved. This means a total of 342 English majors. No follow-

up scores were included in this number, as the longitudinal scores are explored in a separate study in Chapter 10. Previous English language learning experience is reported in Chapter 7 for the three separate groups. Here subjects are treated as one overall group of EFL students, with an expected minimum of B2 language proficiency as defined by the Common European Framework of References. This is demonstrated by the proficiency level of the language exams on the basis of which they were admitted to the university (for a discussion, see Chapter 7).

Descriptive statistics were calculated for all test data. Besides overall test results, data is reported for all five frequency levels separately in the case of the VLT and PVLТ.

8.2.3 Results

8.2.3.1 Vocabulary Levels Test

The descriptive statistics (minimum and maximum scores, means and standard deviation) for the VLT for all five frequency bands and for overall scores are reported in Table 8.1. As could be expected, results show great variability among subjects, which are expressed by the high standard deviation figures. Since the test is made up of five frequency bands, it was possible to calculate scores for each of these levels. The standard deviation is the lowest in the case of the first frequency band (8.77), showing the most homogeneity of results on this 2,000 frequency level. This relative homogeneity of results sharply changes at levels containing less frequent words, being between 15 and 20. Maximum scores were obtained on almost all frequency bands, except for the 10,000 level. The growing difficulty of test items from level to level is shown by minimum scores, which sharply decrease from level 2,000 to level 10,000 (33% and 3%, respectively). A sharp drop is seen in the minimum scores at the 5,000 level, with some students scoring only one right answer out of the 30 items. The academic vocabulary level has results between 10 and 100 percentage points.

Table 8.1 Vocabulary Levels Test scores for all subjects

Frequency band	Minimum	Maximum	Mean	Std. Deviation
VLT 2,000	33	100	93.50	8.77
VLT 3,000	27	100	85.29	15.23
VLT 5,000	3	100	72.97	19.50
VLT 10,000	3	93	50.54	18.67
VLT academic	10	100	84.32	15.26
VLT overall	3	100	77.27	13.97

All values are expressed in percentage points, n=342

When comparing mean scores, it is immediately noticeable that there is a continuous decrease in the results from band to band, the 2,000 level showing a very high 93.50% score, while the 10,000 level receiving only 50.54%. A growing gap is seen in the results between the various levels: 8.5 percentage points between the 2,000 and 3,000 levels, 21 percentage points between the 3,000 and 5,000 levels and 12.5 percentage points between the two most demanding levels. The academic vocabulary level received a percentage score of correct answers similar to that obtained at the 3,000 level, showing also an almost identical standard deviation figure (15.26 and 15.23, respectively).

8.2.3.2 Productive Vocabulary Levels Test

Table 8.2 presents the descriptive statistics (minimum and maximum scores, means and standard variation) for the PVLТ for all five frequency bands and for overall scores. The scores on each frequency level show great variability among subjects. The highest standard deviation figures are indicated on the 3,000 and the academic word levels (21.74 and 22.53, respectively), while the standard deviation figure is the lowest in the case of the 2,000 level (15.72), similarly to the order found in the case of the VLT scores. Maximum scores were obtained on all levels, except for the 3,000 level, which meant at least one incorrect answer for all subjects on this level. Minimum scores report 0 correct answers for all bands, which are only partially due to incomplete testing, as students who did not fill out the test or left most answers blank were excluded from evaluation. In the case of this test, the academic

word level mean values show the closest resemblance with the scores obtained on the 3,000 level (53.26 and 55.07, respectively).

Table 8.2 Productive Vocabulary Levels Test scores for all subjects

Frequency band	Minimum	Maximum	Mean	Std. Deviation
PVLT 2,000	0	100	82.77	15.72
PVLT 3,000	0	94	55.07	21.74
PVLT 5,000	0	100	37.35	18.84
PVLT 10,000	0	100	31.15	19.44
PVLT academic	0	100	53.26	22.53
PVLT overall	0	100	51.95	17.70

All values are expressed in percentage points, n=342

When analyzing mean scores, a continuous decrease is observed from band to band as frequency levels grow. While the VLT mean scores revealed a growing gap from level to level, the gap between the frequency bands in the case of the PVLT gets smaller and smaller. This means a 28 percentage point difference between the 2,000 and 3,000 levels, an 18 percentage point between the 3,000 and the 5,000 levels and only a 6 percentage point difference between the mean scores between the two most demanding levels.

8.2.3.3 Lex30 test

Descriptive statistics for the Lex30 productive test is reported in Table 8.3. Results show great variability within the subject population. Scores range between 13% and 76%, which refer to the proportion of less frequent words of the possible 90 responses (words defined as not included in the first most frequent 1,000 word families in English, as calculated by the VocabProfiler). There was a high rate of no responses in the case of the prompt words *pot* and *substance*, and after the testing sessions many students reported the lack of knowledge of the meaning of these words, especially in the case of first-year students.

Table 8.3 Lex30 test scores for all subjects

Test	Minimum	Maximum	Mean	Std. Deviation
Lex30	13	76	49.82	10.27

All values are expressed in percentage points, n=342

8.2.4 Discussion of the results in Study 1

The study reported in this section had a mainly descriptive aim of presenting results of three tests of vocabulary knowledge of a given student population. There was a great variability in all overall test scores. This calls attention to the fact that a subject population like this one, seen in many EFL/ESL studies, should not be treated as one homogeneous EFL advanced population. Overall group results can inform us about general tendencies, but inevitably raise questions of how results change on the basis of certain features. Therefore, research question number one calls attention to the need for careful data treatment (which should be a key aspect in data collection and evaluation) without which theoretical justifications of data interpretation would also be seriously limited. However, overall descriptive data treatment was necessary as a starting point and as a basis for further analysis. Breaking the overall subject group into various sub-groups is likely to give us better results and more room for understanding relationships between various scores. For that reason, in the following studies various groupings will be employed when treating results.

When analyzing the variability of results for each frequency level of the two Levels Tests, standard deviation scores showed the most homogeneity within the group at the lowest frequency band. As the levels got more demanding, the variability of results changed. However, the largest variability was not seen in the case of the highest frequency level, but at the 5,000 level for the VLT and the 3,000 level for the PVL. This suggests that all students in their tertiary education carried out in English can be expected to perform well at the first band, but most of them have difficulty at the highest frequency band. What happens at the other levels seems to suggest that while the most proficient students score well at these 3,000 and 5,000 levels, less proficient students already show difficulties and perform less successfully. This discrepancy results in the heterogeneity of results. This question will be further analyzed in the next section.

While targeting the gradual decreasing nature of the sub-scores of the two Levels Tests was also only partly confirmed, overall scores did show a decreasing nature, as could be expected, but the maximum score at the 3,000 level of the PVL. This indicated that no student could score the maximum points, while on the subsequent levels maximum points were obtained at least by a few students. This seems to support the findings of previous research

discussed in Section 4.4 (Honeyfield, 1977; Goodfellow, Jones & Lamy, 2002; Schmitt & Dunham, 1999; McCrostie, 2007;) which suggest that corpus-based frequency data do not necessary overlap with the frequency or difficulty of specific lexical items, neither its frequency judged by learners or instructors.

The growing nature of gap between more advanced frequency bands was confirmed only in the case of the VLT, although the differences in mean scores were small. However, the opposite tendency was revealed for the PVLTL, namely, the discrepancy in the mean scores between the frequency levels shows a sharp drop. This suggests that on the productive version of the Levels Test the first frequency band is at a difficulty level that is overcome by most students, while the subsequent levels already starting with the 3,000 band are very demanding. The last two levels show similar mean scores which support this overt difficulty of the higher bands.

Results concerning the different nature of demand posed on the learners are in line with previous studies using the two Levels Tests (e.g. Waring, 1997a; Laufer, 1998; Laufer & Paribakht, 1998) or studies targeting the difference between receptive and productive vocabulary employing frequency bands but also using other testing methods (e.g. Laufer & Goldstein, 2004). The receptive recognition test was less demanding than the parallel productive recall test, although both allowed for partial knowledge.

What makes the interpretation of the *receptive-productive* results more challenging in our case is the scores obtained on the Lex30 test. The overall descriptive analysis presented in Study 1 does not give a final answer as to how this different test fits into the *receptive-productive* dichotomy discussed in Chapter 2. For this reason, Study 3 will look at the relationship between the results gained on all three tests in more detail.

8.3 Study 2 – Factors influencing test results

8.3.1 Aim

Evidence from a review of the empirical research and the discussion of research question one strongly suggest caution in adopting a “one size fits all” description of student populations. As discussed earlier, students classified as L2 learners, even the specific population of English majors, are not a homogeneous group. They differ in many respects, including the actual level of English proficiency, motivation, or language learning background. The study here accommodates for some of the factors expected to play a crucial role in vocabulary proficiency.

Study 2 investigates how certain factors may influence the test results discussed in the previous section. These factors include: the number of years English had been studied by the subjects with reference to the time spent in the English language medium education at the university, the latter one being the basis for selecting students into the three major study groups.

The second broad research question will be explored in this section:

- 2 What is the influence of the amount of academic experience and the amount of language practice (as defined by the amount of time spent with English inside and outside of school and language learning background) on the learners’ vocabulary knowledge?

This main research question will be interpreted on the basis of the following specific questions, treated on the basis of data obtained with the help of a written questionnaire (discussed in Section 7.2.3):

- 2a What is the role of the year of enrollment in students’ vocabulary test scores? In other words, what is the difference in the results of incoming first-year students, those at the end of their first academic year and those at the end of their third year?

- 2b Is there a significant difference in the test scores of the two parallel sub-groups of third-year students? If yes, what could explain it?
- 2c How does the number of years English has been studied influence vocabulary knowledge?
- 2d How does the amount of time spent in an English-speaking country influence test results?
- 2e Do students who spend more time with activities in the target language (including study-related activities) show higher scores on the tests?
- 2f How does the amount of reading done in the target language influence test scores? In other words, do students who report on reading more pages per week (including overall reading and reading for academic purposes) have better test scores?

8.3.2 Procedure

Subjects in this study were the same as in Study 1, which means all 342 students (for more detail see Sections 8.2.1 and 7.1) were included for investigation. While exploring some factors they were treated as one large group, in other cases they are referred to as Groups 1, 2 and 3, indicating incoming first-year, end-of-first-year and end-of-third-year students. For some of the investigation Group 1 was not included, since at the time of data collection they were at the very beginning of their studies; therefore, they were not asked to report on English study-related activities like those of the two upper groups. Furthermore, the study population was also divided into several new groups on the basis of the influencing factors discussed above. The number of participants and reasons behind the identification will be clearly reported for each newly formed group.

The number of years of English studies and the influence of staying in an English-speaking environment were factors analyzed for all three groups of students. Information about the time spent with English per week was drawn from the questionnaire used with Groups 2 and 3, not including Group 1, as explained above. Similarly, the number of pages read in English each week and the number of these pages directly related to their studies were also taken into consideration. These influencing factors were compared to the results on all three test types.

8.3.3 Results

8.3.3.1 Test results for the three subject groups

As a first step, the three study groups were treated separately in order to see whether significant differences could be identified in their test results. Table 8.4 reports the descriptive statistics (mean and standard deviation) for all the frequency levels and the overall score obtained on the VLT. The scores show growing tendency at all levels and in the overall results. What is more, standard deviation figures decrease from group to group in each column of the table. This may mean that the students who enter the university catch up with their more proficient classmates, as they advance in their studies or that the composition of the upper groups change with weak students dropping out²³. These possible reasons will be explored with research question number 4 which studies the drop-out rate of first-year students, and with research question number 7 which studies the change in students' vocabulary knowledge over one academic year.

Table 8.4 Vocabulary Levels Test results for the three main subject groups

Group		VLT 2,000	VLT 3,000	VLT 5,000	VLT 10,000	VLT academic	VLT overall
Group 1 n=148	Mean	90.43	78.78	63.85	42.50	77.50	70.62
	SD	10.69	16.60	20.03	17.43	15.86	14.52
Group 2 n=101	Mean	93.70	85.35	73.08	51.32	83.90	77.33
	SD	7.29	13.77	17.37	17.15	14.66	12.21
Group 3 n=93	Mean	98.15	95.59	87.37	62.49	95.62	87.81
	SD	2.73	6.20	9.84	15.42	4.78	6.45
Total n=342	Mean	93.50	85.29	72.97	50.54	84.32	77.27
	SD	8.77	15.23	19.50	18.67	15.26	13.97

Scores are expressed in percentage points

²³ Unpublished VLT data of incoming first years suggest that upper groups involved in this study entered the university with similar overall group scores of receptive vocabulary size (Peckham, personal communication).

In order to test whether the differences in the overall scores are statistically significant, one-way analysis of variance (ANOVA) was performed. Results show a significant difference between the three groups for overall scores ($F= 57.50$, $DF= 2$, $p<0.01$) and similar significant scores for the separate frequency bands. In order to reveal whether there is a significant difference for all pairs of groups (Groups 1 and 2; Groups 2 and 3; and Groups 1 and 3), Tukey's post hoc test was performed. Differences between all pairs of scores for the three groups were found to be significant at all band levels.

A similar set of analysis was performed for the productive version of the Levels Test. Descriptive statistics (mean and standard deviation) for the three study groups are reported in Table 8.5. Similarly to the VLT, an increasing nature of all types of scores is seen from group to group. A decrease in standard deviation scores, however, is not as clear as in the case of the VLT. For example, at the 2,000 and academic levels, the largest standard deviation figures are reported for Group 2, students at the end of their first year of studies.

Table 8.5 Productive Vocabulary Levels Test results for the three main subject groups

Group		PVLT 2,000	PVLT 3,000	PVLT 5,000	PVLT 10,000	PVLT academic	PVLT overall
Group 1 n=148	Mean	76.46	45.53	30.41	23.91	40.13	43.36
	SD	15.96	19.78	19.36	17.90	18.46	16.53
Group 2 n=101	Mean	82.19	52.40	35.24	28.07	52.23	50.09
	SD	16.31	20.50	15.71	16.90	19.65	15.48
Group 3 n=93	Mean	93.43	73.14	50.71	46.02	75.27	67.63
	SD	6.35	13.65	13.65	16.15	12.22	9.62
Total n=342	Mean	82.77	55.07	37.35	31.15	53.26	51.95
	SD	15.72	21.74	18.84	19.44	22.53	17.70

Scores are expressed in percentage points

The statistical differences at all band levels were tested using ANOVA. Results of the test confirmed that there was a significant difference in the mean scores across the three groups (for overall scores $F=79.67$, $DF=2$, $p<0.01$). Difference between pairs of groups was

tested using Tukey’s post hoc test. Significant differences were found between each of the three possible combinations of group pairs ($p < 0.05$), except between Groups 1 and 2 on the 5,000 and 10,000 levels ($p = 0.07$ and 0.146 , respectively).

Comparison of the mean scores obtained on the Lex30 was also made between the three study groups. Results are reported in Table 8.6. Similarly to the other two tests, a growing tendency of the scores is seen for the PVLТ; however, the increase between the first two groups is very small. Standard deviation figures also seem considerably consistent across the groups. In order to test the statistical significance of the mean scores, an ANOVA was performed, which showed a significant difference between the three groups. Tukey’s post hoc test was also performed in this case to see whether this significant difference appears in all possible combinations of groups. The post hoc test confirmed that there is a statistically significant difference between Groups 1 and 3 ($p = 0.015$), but not between Groups 1 and 2 ($p = 0.776$) or between Groups 2 and 3 ($p = 0.123$).

Table 8.6 Lex30 results for the three main subject groups

Group		Lex30
Group 1 n=148	Mean SD	48.53 10.88
Group 2 n=101	Mean SD	49.43 9.68
Group 3 n=93	Mean SD	52.31 9.57
Total n=342	Mean SD	49.82 10.26

Scores are expressed in percentage points

It is expected that the differences in the results across the three study groups are a result of the combination of factors, and not simply due to increasing number of years of studying English, including the years of enrollment in tertiary education. These and other possible influencing factors reported with the help of questionnaires were analyzed for all three groups as reported in Table 8.7. Overall, it can be seen that the analyzed values increase from group to group, except for the time spent with activities in English on weekdays which was found to be higher for the first-year group than the third-year group (5.15 vs. 4.78 hours),

which is, though, counterbalanced by the time spent with English at weekends. A one-way analysis of variance showed a significant overall difference between factors analyzed in Table 8.7, except for the amount of time spent with English-related activities during the week and at weekends. Moreover, the increase in the number of years of previous English studies is only the reflection of the extra one to three years spent at the university, which suggests that all three groups entered the university with a similar number of years of English studies as a group. When comparing the two advanced groups in terms of reading habits, significant differences are seen both in the case of general and academic-related reading. These results are group means, and do not account for individual differences, therefore, results will need to be discussed in more detail in the next sub-sections when exploring the various influencing factors one by one.

Table 8.7 Mean values for influencing factors for the three main subject groups

Group	Years of English	Time spent in an English-speaking country*	Hours of English per day on a weekday	Hours of English per day at a weekend	Number of pages read per week	Number of pages read for university studies per week
Group 1 n=148	9.03 SD=3.20	1.24 SD=0.24	N/A	N/A	N/A	N/A
Group 2 n=101	9.51 SD=3.32	1.30 SD=0.22	5.15 SD=2.89	6.47 SD=4.91	35.83 SD=82.58	18.72 SD= 25.62
Group 3 n=93	12.02 SD= 3.37	1.48 SD=0.34	4.78 SD=2.52	6.60 SD=5.03	59.18 SD=82.92	38.45 SD=45.58
Total n=342	9.99 SD=3.51	1.32 SD=0.27	4.97 SD=2.72	6.53 SD=4.95	47.03 SD= 83.35	28.18 SD=37.79

* 1→0–3 weeks, 2→1–5 months, 3→6 months or more; N/A means no data were collected regarding these factors from Group 1 (for detailed explanation, see Chapter 7).

8.3.3.2 Results for the two subgroups in Group 3

Since data collection was carried out using two sub-groups of third-year students of similar size (Group 3a consisting of 48 and Group 3b consisting of 45 students), tested in two different academic years but using the exact same procedure, it was possible to directly compare the two groups keeping the influence of the number of years of academic experience

controlled in the target language. As a first step, the total results for the three testing instruments were analyzed to see whether the two groups showed similar scores. Table 8.8 presents the results which suggest that the two subgroups, indeed, do not have the same test results. This is statistically confirmed by performing an independent samples t-test, which shows $p < 0.05$ significance values for VLT and Lex30 tests. For the PVLt, although scores are higher for the second sub-group, no statistical difference could be shown ($p = 0.086$).

Table 8.8 Test of vocabulary knowledge results for the two subgroups of Group 3

Group		VLT overall	PVLt overall	Lex30
3a n=48	Mean SD	86.00 7.17	65.87 10.41	50.18 10.21
3b n=45	Mean SD	89.50 5.22	69.29 8.60	54.19 8.80
Total n=93	Mean SD	87.81 6.45	67.63 9.62	52.30 9.64

All values are expressed in percentage points

Since Groups 3a and 3b were selected and tested using identical measures and being at the same stage of their studies, reasons behind the difference were investigated by looking again at factors such as number of years of English and time spent abroad in an English-speaking country, time spent with English and, finally, the amount of pages read, as these are expected to be strong influencing factors on the vocabulary size and nature of the subjects. Table 8.9 reports on these factors for the two sub-groups. Results in the table show higher values for the second group in terms of the years of English studies (with the second group showing one more year in the mean number), the time spent with English-related activities, including both academic studies and leisure-time activities, and the number of pages read in English each week. Standard deviation figures also show greater variability for the last three factors, the second group reporting higher maximum amounts of reading and time spent with English at weekends. The factors of overall reading and time spent with the target language seem to suggest that students in the second group are more motivated users of the target language. In other words, members of the second group use English more for activities that are not directly required for their studies and not prescribed as assignments. However, these differences in these influencing factors between the two sub-groups were found not to be statistically

significant when using independent-samples t-tests. It is more likely that the different factors, when combined, lead to the statistically significant difference in the overall vocabulary test results.

Since these factors are expected to have direct influence on the vocabulary knowledge of our student population, independently or as co-occurring factors, they are discussed in more detail in the next sections involving the entire subject group.

Table 8.9 Mean values for influencing factors for Groups 3a and 3b

Group	Years of English	Time spent in an English-speaking country*	Hours of English per day on a weekday	Hours of English per day at a weekend	Number of pages read per week	Number of pages read for university studies per week
Group 3a n=45	11.49 SD=3.48	1.16 SD=0.37	4.47 SD=2.68	5.82 SD=3.74	53.91 SD=60.01	37.69 SD=36.31
Group 3b n=48	12.52 SD=3.22	1.10 SD=0.31	5.08 SD=2.35	7.33 SD=5.94	64.13 SD=100.18	39.17 SD=53.21
Total n=93	12.02 SD=3.37	1.13 SD=0.34	4.78 SD= 2.52	6.60 SD= 5.03	59.18 SD=82.92	38.45 SD=45.58

* 1→0–3 weeks, 2→1–5 months, 3→6 months or more

8.3.3.3 Years of English as an influencing factor of vocabulary knowledge

As a first step, Pearson's correlation was run between the overall scores obtained on the three vocabulary tests and the factor of years of English studies. Weak, but significant correlation was found between these parameters, for VLT $r=0.35$ ($p<0.01$), for the PVLT $r=0.38$ ($p<0.01$), and for the Lex30 $r=0.22$ ($p<0.01$). However, students enrolled in the study show great variability in the number of years they have been studying English, a variability that cannot be clearly deduced from the mean numbers reported in Table 8.7. Years of previous English studies range between 1 and 20 years, both, of course, are extremes since the one year means that students could enter the university with an accelerated one year of preparation, while the 20 years suggest that these students had been exposed to English from birth or starting from the very first years of their life.

In order to control for the great variability of years and to show whether tendencies of stronger relationships between test results and years of study could be found, the entire study population was divided into three subsequent groups according to the years of English studies: Group A had 1–5 years of studies (n=46), Group B had 6–10 years of studies (n=153) and Group C had more than 10 years of studies (n=143). Note that these groups do not overlap with the original Groups 1, 2 and 3. Attempt was not made to have equal number of students in each new group, rather, gaps in the results of the influencing factors were used to find the best cutting points for forming groups. Overall scores were calculated for the three tests of vocabulary knowledge using the newly formed groups. Results are reported in Table 8.10 and indicate clearly visible differences in scores between the groups. This is the case for all three tests, for the VLT results are 70.78, 74.61 and 82.22 for the three groups, for the PVLТ mean scores change from 42.82 to 48.34 and then to 58.75. A similar increase is seen in the case of the Lex30, with mean scores ranging between 46.28 and 51.74.

Table 8.10 The influencing factor of years of English studies

Group		VLT overall	PVLТ overall	Lex30
Group A (1–5 years of studies) (n=46)	Mean	70.78	42.82	46.28
	SD	13.94	16.39	11.72
Group B (6–10 years of studies) (n=153)	Mean	74.61	48.34	49.02
	SD	15.06	17.35	10.53
Group C (min. 11 years of studies) (n=143)	Mean	82.22	58.75	51.74
	SD	10.90	16.03	9.12

All values are expressed in percentage points

The statistical significance across the groups for overall scores on the three tests and the sub-scores on the VLT and PVLТ were tested using ANOVA, which supported the statistically significant nature of the differences ($p < 0.01$ for all score). In order to locate which of the possible three pairs of groups could explain the statistical significance, Tukey’s post hoc test was performed. The post hoc test located a significant difference in all three overall test results for group pair A–C, having a minimum of 6 years of difference in the number of years they had been studying English. Moreover, a similar significant difference

was found for group pair B–C (students with less and more than 10 years of English studies) for the two Levels Tests, but not for the Lex30. However, no statistically significant difference was found between Groups A and B in any of the tests or sub-tests, although scores are higher for Group B than Group A.

These findings seem to suggest that a few years of difference is not necessarily significant in the number of target-language studies, however, results do not indicate that scores follow a certain course of increase from year to year. Indeed, the highest and the weakest test results are not obtained by students with the least or most number of years of English exposure.

8.3.3.4 Time spent in an English-speaking country as an influencing factor of vocabulary knowledge

When treating the entire study population as one group, and correlating the overall results on the three tests with the time spent in an English-speaking country, we find that a significant correlation is found only in the case of the Lex30, and this correlation is very weak ($r=0.111$, $p<0.01$). Similar to the grouping done for the previous factor, the main study population was divided into three new groups on the basis of their experience with staying in an English-speaking country. Again, the aim was to find important cut-off points in the data and not to find groups with balanced number of members. Many of the students who participated in this study had had very little or no experience staying in an English-speaking country ($n=258$), some reported a stay between one and five months ($n=58$), and less than 10% had had a half-year period or longer stay ($n=28$). The investigation aimed to reveal possible differences in test scores across these three groups (called Groups D, E and F). Table 8.11 reports the overall results obtained on the three tests of vocabulary knowledge for these three newly formed groups. Although results visibly change from group to group, these changes could not be confirmed with ANOVA, differences were slightly above the 5% significance level. When Groups E and F were combined and tested against Group D (resulting in a group with more than one and another with less than one month experience), differences were still found to be statistically not significant. This may be partly due to the

uneven number of subjects in the groups or to the high standard deviation figures, showing a great overlap of results between groups).

It can be concluded that those who had spent a period between one and five months in an English-speaking country, as a group produced scores slightly higher than those with little or no such experience, and those who could spend at least half a year in such an environment scored well above the mean values. However, these differences were not statistically supported. Again, it can be expected that a combination of factors leads to significant increase in vocabulary knowledge, as will be explored in Section 8.2.3.6).

Table 8.11 The influencing factor of time spent in an English-speaking country

Group		VLT overall	PVLT overall	Lex30
Group D (less than a month) (n=258)	Mean	76.77	51.13	49.26
	SD	13.90	17.09	10.70
Group E (1–5 months) (n=58)	Mean	77.36	52.83	50.80
	SD	14.33	20.55	8.57
Group F (6 months or more) (n=28)	Mean	82.10	58.11	53.29
	SD	13.52	16.02	8.98

All values are expressed in percentage points

8.3.3.5 Time spent with English as an influencing factor of vocabulary knowledge

Students in Groups 2 and 3 (n=194) were asked to report on the number of hours they spend with any activity carried out in English, not limited to their university studies. This possible influencing factor was divided into two parts: the time spent with English on weekdays and that spent during the weekends, as these were expected to be different. As a first step, Pearson correlation was carried out between the three test scores and the two influencing factors of time. A weak significant correlation was found only for the PVLT in terms of the time spent with English on weekdays ($r=-0.150$, $p<0.01$). This weak negative correlation, however, is difficult to interpret as it would suggest that less time spent with the target language leads to higher test scores. It needs to be noted that these relationships between influencing factors and test scores are based on reported overall time in a written

questionnaire. Students may under- or overestimate the average time they spend with English. Underestimation may be especially true for students who report one or two hours, since they are likely to spend more time with English simply counting the time they spend in class. Here again, students reported on a variety of figures, between one and sixteen hours daily on weekdays and between zero and 30 hours for an average weekend. If weekday figures are possibly underestimated in some cases, the zero or close to zero figures at weekends call for concern, as they indicate that some students are not involved in target language activities during this period of the week. The large differences in the reported number or hours called for the need to set up groups, similarly to what was done in the case of the other factors, taking into consideration the range of the data, to see whether there could be larger tendencies of influence on the vocabulary test results in terms of time dedicated for English.

For the influence of time spent on average with English on weekdays, three groups were identified: Group G contains those who reported a time between one and four hours (n=92), Group H is made up of students who dedicate between five and eight hours to English (n=88) and Group I consists of students with more than eight hours spent with English on weekdays (n=14). This last figure of eight hours may seem to be an overestimation, but can easily be understood if we count the classes they have, the assignments they do, the music they listen to and films they watch on a daily basis as English-related activities. Statistically significant differences were not found between these groups. For mean scores and standard deviation figures for these groups refer to Table 8.12.

Table 8.12 The influencing factor of time spent with English on weekdays

Groups		VLT overall	PVLT overall	Lex30
Group G (1–4 hours per day) (n=92)	Mean	83.41	60.60	51.76
	SD	9.62	14.19	8.86
Group H (5–8 hours per day) (n=88)	Mean	81.86	57.31	49.72
	SD	12.49	16.88	10.57
Total (n=194)	Mean	82.35	58.50	50.81
	SD	11.17	15.67	9.71

All values are expressed in percentage points

Similarly, the study population was divided into groups on the basis of the average amount of time they reported on for a weekend (for two days, Saturday and Sunday together). Cut-off points were identified on the basis of the range of values reported on by the subjects. What is surprising, some students claimed to spend not even an hour at an average weekend with any activity carried out in English. Group J consists of students who reported between zero and four hours of activities carried out in English (n=77), subjects in Group K dedicate between five and ten hours to English (n=94) and students in Group L more than ten hours (n=23). Table 8.13 shows results that indicate little influence on the time spent at a weekend with English, and, as a result, a clear and statistically significant influence of this factor was not found for any of the test scores.

Table 8.13 The influencing factor of time spent with English at weekends

Groups		VLT overall	PVLT overall	Lex30
Group J (0–4 hours per weekend) (n=77)	Mean SD	82.22 11.58	59.99 14.81	50.54 10.52
Group K (5–10 hours per weekend) (n=94)	Mean SD	82.29 10.79	57.17 15.76	51.54 8.54
Group L (min. 11 hours per weekend) (n=23)	Mean SD	83.04 11.76	58.96 18.17	48.64 11.40
Total (n=194)	Mean SD	82.35 11.17	58.50 15.67	50.81 9.71

All values are expressed in percentage points

8.3.3.6 Amount of reading as an influencing factor of vocabulary knowledge

As a last possible influencing factor of the subjects' vocabulary knowledge, the amount of reading done by the students was also brought under investigation. Groups 2 and 3 were asked to report on the number of pages read on average each week and then indicate the number of pages out of this previous figure related to their university studies. Since the figures varied greatly, ranging between 1 and 300 for all readings and between 0 and 250 for university studies, four groups were identified for both variables. Group M indicate those who

read between 1 and 10 pages (n= 57), Group N who read between 11 and 20 pages (n=43), Group O who read between 21 and 50 pages (n=60), and Group P who read more than 50 pages per week in English (n=32). Mean scores for the vocabulary tests are presented in Table 8.14 for these newly identified groups. Results for the VLT show a visible difference between those students who read more and those who read less than 20 pages per week. A more gradual change is suggested in the case of the PVLТ and the Lex30. To confirm the statistically significant nature of the differences across the four groups, ANOVA was conducted, which showed a statistically significant difference for VLT and PVLТ overall scores (p=0.013 and p=0.036, respectively), but not for the Lex30 results.

Table 8.14 The influencing factor of all readings done in English

Group		VLT overall	PVLТ overall	Lex30
Group M (1–10 pages per week) (n=57)	Mean SD	79.89 11.42	55.21 15.98	49.71 10.76
Group N (11–20 pages per week) (n=43)	Mean SD	79.63 13.52	56.33 15.25	50.10 9.58
Group O (21–50 pages per week) (n=60)	Mean SD	84.50 9.47	59.70 15.64	51.36 10.21
Group P (more than 50 pages) (n=32)	Mean SD	85.84 8.51	64.53 13.98	52.09 6.68

All values are expressed in percentage points

A similar analysis was performed for the amount of reading in connection with the subjects' university studies. Four new groups were identified for this purpose. Students in Group R reported to read between zero and ten pages (n=80), in Group S between eleven and twenty pages (n=46), in Group T between twenty-one and fifty pages (n=46) and in Group V more than 50 pages per week (n=21). Even though these are all students at the end of their first or third year of studies, the largest of the four groups includes those who claim to read no more than ten pages per week for their university studies. Table 8.15 reports on the vocabulary test scores for these four groups. Results do not show a clear relationship between

reading done for academic studies and the test scores. This is confirmed by ANOVA, which located a statistically significant difference only for VLT overall results.

Group 8.15 The influencing factor of reading done in English in connection with university studies

Group		VLT overall	PVLT overall	Lex30
Group R (0–10 pages per week) (n=80)	Mean SD	80.87 11.50	56.21 15.82	45.45 9.36
Group S (11–20 pages per week) (n=46)	Mean SD	82.02 12.46	59.09 14.31	46.09 8.82
Group T (21–50 pages per week) (n=46)	Mean SD	83.13 10.44	58.91 17.47	44.85 8.85
Group V (more than 50 pages) (n=21)	Mean SD	82.33 11.19	58.43 15.68	45.69 8.76

All values are expressed in percentage points

8.3.3.7 *The role of multiple factors*

The previous subsections investigated the role of factors that were expected to have an influence on vocabulary knowledge. Since the factors by themselves were found to be statistically significant contributors of test scores only in a few cases, it is expected that factors together have a stronger influence on the vocabulary knowledge of students than individual variables. In order to determine whether there were overall differences among the test results in terms of multiple factors, a multivariate analysis of variance (MANOVA) was performed using the factors as dependent variables and the three overall test scores as independent variables. The MANOVA revealed a statistically significant difference for some groups of factors for the VLT results, less on the PVLT results and none for the Lex30 results. The combinations of factors which had a statistically significant influence on overall test scores are reported in Table 8.16. Results show that receptive vocabulary size is the variable that is most likely to be influenced not only by single factors, but also by combinations of factors. Most of these factors are in pairs, but two groups of three factors were also identified

by the MANOVA. These are, except for one case, a combination of years of English and another factor.

Table 8.16 Statistically significant combinations of influencing factors on vocabulary test results

<i>VLT results:</i>
<i>Group + years of English (F= 4.066, p=0.047)</i>
<i>Years of English + time spent with English at weekends (F=3.730, p=0.015)</i>
<i>Years of English + reading for university studies (F=9.769, p<0.001)</i>
<i>Time spent with English at weekends + reading for university studies (F=3.094, p= 0.032)</i>
<i>Years of English + time spent with English at weekends + overall reading (F=8.567, p=0.005)</i>
<i>Years of English + overall reading + reading for university studies (F=4.608, p=0.035)</i>
<i>PVLT results:</i>
<i>Years of English + reading for university studies (F=3.036, p=0.034)</i>

8.3.4 Discussion of the results in Study 2

As was expected, results confirmed that there was a high and statistically significant difference in test scores across the three original study groups. A gradual increase in the scores was seen from group to group; however, as was pointed out, group means could hide individual differences, as a certain year in the course of their studies could well result in different number of years of previous English studies, more experience with academic language use, more demanding course assignments and different amount of time dedicated to activities carried out in English. For this reason possible influencing factors were identified and analyzed for the study population as a whole (n=342). It needs to be noted again that these differences are apparent-time in nature, and could hide the influence of participant selection, drop-out rate or individual differences. Real-time changes in students' lexicon will be discussed in Chapter 10.

When new groups within the participants were identified on the basis of the years of English they had had, their experience in an English-speaking country, amount of time dedicated to English and amount of reading done in the target language, influencing factors

seemed to play the greatest role in the receptive vocabulary size of learners. All factors had a positive influence on this type of lexical knowledge, although it needs to be stressed that statistical difference, in many cases, was found only between non-neighboring groups. This implies that a small difference in the amount of time dedicated to English and the reading done has little influence per se on the size of the learners' recognition vocabulary.

On the whole, a clear and statistically strong relationship was not found in the case of all expected influencing factors, although in most cases a visible improvement was seen in the scores from group to group. The statistically significant role of the influencing factors was difficult to identify, even between two far-end groups. However, when treating influencing factors together, statistically significant differences could be identified.

On a more pedagogical note, results also indicate that it is not only the amount of target language learning that counts, but also the quality of learning, which is transferable into, for instance, amount of reading in the case of subjects enrolled in target-language medium education. This was confirmed when taking a closer look at the two parallel groups of third-year students who produced significantly different test results. The one-year difference in their overall English studies could not be the key reason behind this advantage, as other results showed that a few extra years of target language studies are needed to show overall improvement in the lexicon. Indeed, results pointed to the extra time spent with English and the extra pages of academic texts read as major influencing factors when combined with time spent with English on a daily basis or the number of years of previous English studies. The non-significant difference in mean test scores between students who spend many hours and those dedicating very limited time to English calls into question the quality of language practice done by some members of the study population. It was disappointing to see that some students, already at the end of their first or third year of English studies, close to an exam period, reported on no time spent with English during entire weekends, and only a few pages read in English each week. This is hardly possible to be regarded as extensive reading necessary to learn new vocabulary items or gain subject-matter knowledge as independent learners, as discussed in Chapter 5.

The limited time spent with the target language by many of the students implies that they have serious motivational problems and have very limited study or personal goals related to English. We have only some data on what these goals of English majors studying in

Hungary could be, based on the data published in Kormos *et al.* (2002) and on the information reported by the incoming first-year students involved in this investigation. Data on the question related to what these first-year students were using English for revealed that some students had no clear ideas or activities to list. Answers like “at this point I am using it for nothing” were provided by more than a handful of participants. I am to believe that students filled out the questionnaires to the best of their knowledge, and were reporting actual activities (as was revealed during the piloting of the questionnaire reported in Chapter 7). Some over- or underestimation could be possible in the case of questions related to overall amounts of time spent with the target language or the amount of reading done, but in general it can be concluded that overall tendencies clearly show that not enough time is dedicated to English which would be necessary for the improvement in language proficiency and, more specifically, in vocabulary knowledge.

Moreover, a limited amount of language practice makes implicit vocabulary learning very slow and of little effectiveness. The vocabulary requirement of academic studies is hardly met through solely implicit strategies. Explicit learning should be combined with implicit activities, which also would require time and motivation. If both strategies are of limited use, then very little vocabulary gain can be expected from students. The role of reading and writing and vocabulary knowledge in academic achievement will be further discussed in Chapters 9 and 10.

8.4 Study 3 – Relationship between test results

8.4.1 Aim

This study directly follows up on the descriptive data presented in Study 1 and targets the relationship between the proficiency in receptive and productive vocabulary of students, as measured by the three tests. This study addresses the following main research questions:

- 3 What is the relationship between the knowledge of receptive and productive vocabulary of students as measured by tests? In other words, to what extent can scores on a receptive test predict scores on a productive test? Also, how do sub-scores and overall test scores relate to each other?

This study elaborates on the relationship between scores obtained on different tests and between sub-scores of the same test. It takes one step further on already existing data with other, non-Hungarian subject populations by exploring the relationship between the results of two vocabulary size tests with the results of the productive association test. The study explores the relationship between scores obtained on one receptive test and two productive tests, to balance the need for a better understanding of receptive and productive vocabulary discussed in Chapter 4. Since the Levels Tests are composed of five frequency bands, it will be possible to explore the relationship of parallel frequency bands of these two tests and the overall scores of the tests to the specific frequency scores. This will be of importance for real-life assessment circumstances to know how certain sub-scores relate to overall results. With the information we gain, it will be possible to see how a small sub-set of tests can inform us about overall vocabulary knowledge or help us chose the most appropriate level for assessment purposes in case the full test battery cannot be used.

The following specific research questions will be explored in this section:

- 3a Are there statistically significant and strong correlations between the overall scores obtained on the three vocabulary tests?
- 3b What is the degree of correlation between the overall Levels Tests scores and sub-scores of the two Levels Tests?
- 3c Is there a degree of correlation between the parallel sub-scores of the two Levels Tests?
- 3d Do correlation figures differ for more and less proficient students?

8.4.2 Procedure

Subjects involved in this study were the same as in Study 1, which means all 342 students (for more detail, see Section 8.2.2). They were treated both as a homogeneous group and as three separate groups, according to their year of enrollment. Correlation matrixes are drawn for overall scores obtained on the three tests and also for each frequency band of the VLT and the PVLТ.

8.4.3 Results

In order to answer research question 3a, the overall test results of all 342 subjects presented in Study 1 deserved to be further examined. The bivariate relationship between the variables was analyzed by using Pearson's product-moment correlations, shown in the correlation matrix in Table 8.17. The values revealed a highly strong and significant correlation between the overall results of the two Levels Tests ($r=0.88$, $p<0.01$) and a moderate correlation ($r=0.43$ and $r=0.41$, respectively, $p<0.01$) between the two Levels Tests results and the scores on the association task.

Table 8.17 Correlation matrix for the overall scores on the three tests for the entire study population

	VLT overall score	PVLT overall score	Lex30
VLT overall score	1	0.88*	0.43*
PVLT overall score	0.88*	1	0.41*
Lex30	0.43*	0.41*	1

All values are significant at $p < 0.01$, two-tailed.

As a second step, in order to test the bivariate relationship between the various frequency levels and overall scores on the two Levels Tests, sub-scores were analyzed using Pearson's correlations. This step shows which level has the strongest relationship with overall scores which is important in case we want to use not all, but one or two bands, as for example the 3,000, 5,000 or academic levels of the VLT that predicts reading success discussed in Chapter 9. Results for the VLT are shown in Table 8.18. The correlation values between the variables revealed strong and significant relationships across the frequency levels. The strongest correlations were reported in the case of the overall scores and the 3,000, 5,000 and academic levels (0.92, 0.93 and 0.92, respectively, $p < 0.01$). The weakest, but still moderate and statistically significant relationship is seen between the 2,000 and 10,000 levels (0.60, $p < 0.01$).

Table 8.18 Correlation matrix for all frequency levels on the Vocabulary Levels Test

	VLT overall	VLT 2,000	VLT 3,000	VLT 5,000	VLT academic	VLT 10,000
VLT overall	1	0.81	0.92	0.93	0.92	0.87
VLT 2,000	0.81	1	0.75	0.71	0.77	0.60
VLT 3,000	0.92	0.75	1	0.82	0.86	0.73
VLT 5,000	0.93	0.71	0.82	1	0.81	0.77
VLT academic	0.92	0.77	0.86	0.81	1	0.73
VLT 10,000	0.87	0.60	0.73	0.77	0.73	1

All values are significant at $p < 0.01$, two-tailed.

The same set of analyses was run for the sub-scores of the PVLT (see Table 8.19). The correlation values proved to be smaller than in the case of the VLT, but significant in all cases ($p < 0.01$). The strongest correlation was identified between the 3,000 and academic levels, between the 2,000 and 5,000 levels and between the 2,000 and 3,000 levels ($r = 0.80$, 0.80 and

0.79, respectively, $p < 0.01$). The weakest link was found between the 5,000 and 10,000 levels ($r = 0.32$, $p < 0.01$). With the overall score the 5,000 level showed the strongest relationship ($r = 0.74$, $p < 0.01$).

Table 8.19 Correlation matrix for all frequency levels on the Productive Vocabulary Levels Test

	PVLT overall	PVLT 2,000	PVLT 3,000	PVLT 5,000	PVLT academic	PVLT 10,000
PVLT overall	1	0.78	0.70	0.74	0.67	0.38
PVLT 2,000	0.78	1	0.79	0.80	0.77	0.39
PVLT 3,000	0.70	0.79	1	0.77	0.80	0.41
PVLT 5,000	0.74	0.80	0.77	1	0.75	0.32
PVLT academic	0.70	0.77	0.80	0.75	1	0.37
PVLT 10,000	0.38	0.39	0.41	0.32	0.37	1

All values are significant at $p < 0.01$, two-tailed.

In order to answer the third specific question (3c) related to the relationship between the parallel scores on the two Levels Tests, Pearson's correlations were tested using the sub-scores of the two tests. Results are reported in Table 8.20. Significant correlations were found between the parallel scores, r values ranging between 0.69 on the 2,000 level and 0.78 on the 3,000 level ($p < 0.01$).

Table 8.20 Correlation matrix for parallel frequency levels of the two Levels Tests

	VLT 2,000	VLT 3,000	VLT 5,000	VLT academic	VLT 10,000
PVLT 2,000	0.69	0.74	0.73	0.72	0.63
PVLT 3,000	0.65	0.78	0.77	0.74	0.72
PVLT 5,000	0.54	0.69	0.76	0.67	0.74
PVLT academic	0.60	0.73	0.76	0.73	0.69
PVLT 10,000	0.54	0.69	0.73	0.66	0.77

All values are significant at $p < 0.01$, two-tailed

In order to test research question 3d, namely the varying degrees of correlation between results obtained on the three types of vocabulary knowledge tests, the study population had to be divided into a more and a less proficient group. This was done using the subjects' VLT scores. After a careful consideration of the descriptive statistics (means, frequencies and range) of the VLT scores, a lexically less proficient group was identified as those students who scored between 15 and 80 percentage points overall on the test ($n = 181$)

and a lexically more proficient group those who scored between 81 and 98 percentage points overall on the test (n=159). 80 was chosen as a cut-off point instead of another possible one of the mean score 77, as the 80 percent points prove to be an important landmark for predicting academic success of first-year students (for more detail refer to Chapter 9). Pearson's correlation values for these two groups are reported in Table 8.20. Correlation values are significantly different for the two groups, all values are higher and statistically significant at the $p < 0.01$ level for the less proficient group. Values are weaker for the more proficient group, which means that of the three pairs of tests only one showed $p < 0.01$ level significance ($r = 0.64$), one a $p < 0.05$ significance ($r = 0.17$) and not even a weak significant difference was identified for the PVLТ and Lex30 test pairs for this group ($r = 0.10$). This implies that a specific test score predicts other test scores more strongly for students with smaller receptive vocabulary size than for those having a larger receptive vocabulary. Correlations done on the basis of this proficiency grouping correlation values are less strong in all cases than those obtained using the student population as a whole (for results see Table 8.21).

Table 8.21 Correlation matrix for the overall scores on the three tests for lexically more and lexically less proficient students

	VLT overall MP	VLT overall LP	PVLТ overall MP	PVLТ overall LP	Lex30 MP	Lex30 LP
VLT overall	1	1	0.64**	0.79**	0.17*	0.33**
PVLТ overall	0.64**	0.79**	1	1	0.10	0.37**
Lex30	0.17*	0.33**	0.10	0.37**	1	1

MP=lexically more proficient students, VLT scores between 81 and 98, n=159

LP=lexically less proficient students, VLT scores between 15 and 80, n=181

* correlations are significant at $p < 0.05$, two-tailed; ** correlations are significant at $p < 0.01$, two-tailed

8.4.4 Discussion of the results in Study 3

It was of both theoretical and practical importance to determine the relationship between the various test scores. The level of detail used in this study was necessary in order to see not only how overall results correlated, but also to identify how certain frequency levels correspond to the overall scores obtained on the same test. Data are promising in a sense that they present clear statistical relationship across variables. Results revealed a strong

relationship between the two parallel Levels Tests, and only a more moderate relationship between the association task and the vocabulary size tests. These findings are hardly surprising when one considers the fact that the three tests measure different aspects of what can be summarized as vocabulary knowledge, and use different test formats and methods. The results confirm that partial receptive knowledge of form and meaning of a word tested by the receptive Levels Tests is a prerequisite of its productive recall and use. It follows that a frequency band known in the productive test is expected to be also known in the receptive test, while partial receptive knowledge is not necessarily combined with the ability to recall this lexis in context, especially in the case of low-frequency words which are less often used. Results also suggest that, on average, students with a larger vocabulary size can be expected to produce less frequent words in an association task. It is, however, to be shown in further research in this dissertation whether students who score better on the two productive tests do use a wider range of low-frequency vocabulary in their written production tasks.

The identification of a lexically less and more proficient group on the basis of their receptive vocabulary size made it possible to determine that students with more limited vocabulary size have stronger links between their receptive and productive vocabulary size and the quality of links in their mental lexicon. There seems to be a less strong correlation between vocabulary size and the frequency of words recalled by students in a free association task. This supports the idea that a learner who knows more words has a wider range of lexical items to recall for any purpose. Therefore, an association task that uses prompt words, selected for their weak primary responses, produces a large variability of responses.

It was also revealed which of the frequency levels have the strongest correlation values with overall scores and with other, neighboring levels. This result has, first of all, a strong practical value, as it facilitates test design and organization in real-life testing situations when efficiency and time constraints need to be kept in mind, and, therefore, often a much more restricted number of tests or subtests are administered.

8.5 Concluding summary

This chapter has shown how vocabulary can be assessed using three separate tests targeting various aspects of learners' lexical knowledge. It has been possible not only to describe this knowledge of the entire study population, but we could also map changes in the results based on influencing factors such as year of enrollment in academic studies at the university, language learning background and practice in the target language. The third study in this chapter has focused on discussing test results in terms of their predictive nature of other results. Statistically significant correlations were found between scores and sub-scores of the same or parallel testing instruments, which made it possible to identify certain frequency bands of the Levels Tests the use of which could substitute for the administration of the entire test battery. This outcome, alongside with the theoretical implications, has a significant practical value for future testing among similar study populations. As has been discussed, it is often the case in institutionalized testing or in research settings that only one test or sub-tests are used, and it is important to know how these relate to a fuller picture of learners' vocabulary knowledge. Once test results have been described and possible relationships discussed, I shall now turn to the investigation of the role of vocabulary knowledge on lexical use in written assignments, its implication for reading ability and academic success in general.

CHAPTER NINE

VOCABULARY KNOWLEDGE AND USE

9.1 Introduction

This chapter will follow up on Studies 1, 2 and 3 built around research questions 1 through 3 as discussed in Chapter 8. These three studies were used to gather data on different aspects of the vocabulary knowledge of an adult population enrolled in target-language medium tertiary education. Descriptive data analysis was first provided for the entire subject population, and then differences on the basis of stages in their studies and various factors in their language learning background were explored. It was also investigated how various test results relate to each other with a view on both the theoretical and practical value of the study. The significance of this tested lexical knowledge in predicting academic success and the use of this knowledge in written text production will be the focus of the next three studies. Before turning to the discussion of each study, in order to facilitate the understanding of the complexity of questions raised, the rationale behind each of these research areas is briefly summarized in the following paragraphs.

As has been pointed out in Chapter 5, the successful understanding of a text greatly depends on the knowledge of the lexicon in a given text. Researchers have calculated that between a 95 and 98% coverage is necessary in order to read fluently and to be able to guess the remaining vocabulary from context (Hirsh & Nation, 1992; Hu & Nation, 2001). In vocabulary size, this means the recognition vocabulary of about the first 3,000 word families in English for general texts. But, as has been discussed in Chapter 5, the familiarity with the first 5,000 word families and the academic word list would be a more preferred criteria for students engaged in target language medium studies at a tertiary level, since academic texts also include topic-related technical vocabulary that students are not expected to fully know, and, therefore, fall into the two to five percent unknown vocabulary of texts. Written input is a key factor in gaining study-related information in the form of study notes, reading assignments, but also for using English outside of the university, for a variety of purposes, therefore, it is crucial that students understand texts.

By investigating the receptive version of the Levels Tests, it will be relatively easy to identify whether students meet this vocabulary threshold necessary for general reading and for their university studies. The implication of their text coverage will also be explored in terms of possible vocabulary learning from texts. Furthermore, as a next step in Study 4, the vocabulary knowledge assessed with the help of the three testing instruments will be investigated as possible predicting measures of academic success. This implies not only reading, but also the general understanding of course materials and the successful completion of the syllabus. This will be explored in terms the failing grades in the first academic year of first-year students.

Studies 5 and 6 will further explore the vocabulary knowledge of subjects by comparing their test scores to the lexical profile of their written production. As Nation (2001: 362) notes “vocabulary learning is not a goal in itself; it is done to help learners listen, speak, read or write more effectively. When testing vocabulary, it is important to distinguish between how well a word is known and how well a word is used.” With this in mind, research question 5 will investigate the relationship between vocabulary knowledge and use by analyzing how productive vocabulary (as seen by test results) may be activated to be used in a given written task. As discussed in Chapter 4, three lexical factors in the texts will be employed, namely lexical profile (the proportion of the first 2,000 word families, following Laufer [1998]), the type/token ratio and the lexical density. It will be explored whether all three measures can distinguish between a lexically rich and poor text. This investigation is a major asset of this dissertation and fills the gap in the limited literature related to the assessment of the lexical profile of student texts in terms of their vocabulary knowledge. As has been pointed out, this is rarely the case with corpus driven text analysis, as large learner corpora, in most cases, can inform us about general tendencies without reference to the underlying constructs and reasons.

In order to gain a better understanding of the results emerging in Study 5 relating to the lexicon of the written production tasks, research question 6 will take a step further in the analysis of the lexical profile of students’ essays by investigating how certain factors may influence the lexical factors of these texts. Therefore, it will be assessed whether experience with writing longer texts leads to a more sophisticated lexical use in writing, and how writing strategies related to lexical choice may influence vocabulary use. Furthermore, it will be

investigated whether students who report on concentrating more on vocabulary during their writing process than on grammar or text organization, do use a greater variety of words. Moreover, the influence of topic choice will be explored since students during data collection were provided with more than one essay prompts. It is expected that topic may have an influence on the lexical profile of texts produced, and that lexically more proficient students differ in their topic choice from their less proficient classmates.

The questions explored in this chapter will have both theoretical and practical implications for vocabulary learning and use in general, and for the target language use for academic purposes of the student population under investigation in particular.

9.2 Study 4 – Vocabulary knowledge and academic success

9.2.1 Aim

After seeing the great variability in the vocabulary test results in the previous chapters, the question emerges how vocabulary knowledge may affect the academic achievement of subjects. The main research question addressed in this section is the following:

- 4 How can vocabulary test results predict academic success in terms of reading ability and successful passing of courses?

As seen from the research question, in this section two related but separate studies will examine the role of vocabulary knowledge in the academic success in target-language medium education at a tertiary level. Academic success of students here is defined as being able to follow the assigned syllabus and completing all required courses, regardless of the nature of the passing grades.

First, it will be explored how the receptive size test results predict the difficulty of the reading of general and academic texts (Study 4A). Since the Vocabulary Levels Test consists of frequency bands, and scores are available for each, the focus will be dedicated to the 3,000, the 5,000 and the academic levels, as these show the threshold levels necessary for reading general and academic texts (for detail refer to Chapter 5). Predicted reading difficulty will also be compared to the stated difficulty with reading study materials reported by the students. In the discussion section it will be further explored what the implication of the results is on the acquisition of new lexical items and the direction of inference of unknown vocabulary from text (top-down or bottom-up, as reported by Nation [1993]).

After the investigation of the predictive role of the academic success in terms of reading ability, test results will be compared to the course grades of first year students (Study 4B). It will be explored whether it is possible to set a minimum lexical proficiency level for students without which they are likely to fail in their English language medium studies. The test scores of incoming first-year students will be analyzed with a diagnostic purpose, by comparing them to their first-semester language class results. Then, test results of students at

the end of their first-year studies will be examined in terms of their final class results at the end of their first year. This second step in the analysis is important as an overall picture of academic success, since some content classes are also part of the first year syllabus at the institution where this research took place, such as Introduction to English or to American Studies, Introduction to Linguistics, and Introduction to Literature. By the end of the first year all required language and content classes need to be passed as a prerequisite for the comprehensive language exam discussed in Chapter 7. It is possible to retake failed language classes in the second semester or sign up again for failed content class exams, therefore, taking the comprehensive exam at the end of the academic year is an indication of the successful completion of the syllabus for the given study period, considering possible retakes. In this case, only the fact of being admitted to the comprehensive exam will be explored, the exam results themselves will not be treated in detail as these would lead us far from the central discussion of the dissertation. In the next sections Studies 4A and 4B will be discussed separately in order to facilitate the discussion of the results.

9.2.2 Study 4A – Vocabulary knowledge and reading

9.2.2.1 Overview and aim

This section will treat the subjects' vocabulary knowledge as a predictive factor of written text understanding. The aim of the study is to ascertain if students possess the minimal lexical knowledge needed for fluent reading and inferring unknown lexical elements in context, which are also basic requirements for incidental lexical acquisition. It will also be explored how students themselves view their difficulty with the assigned reading material.

The following specific questions will be explored in this study:

- 4A a Do students in their first and third year of university studies meet the threshold level necessary for reading general texts? In other words, do they have the partial knowledge of meaning and form of the 3,000 word level of the receptive size test?

- 4A b Do they have the receptive vocabulary size needed to follow academic texts? In other words, do they show an adequate knowledge of the 5,000 and academic levels of the Levels Test?
- 4A c Does students' report on their reading difficulty of assigned course material match the predicted reading difficulty deduced from vocabulary test results?
- 4A d Can we set a minimum lexical proficiency level for students without which they are likely to fail in their English language medium studies?

9.2.2.2 Procedure

All three large groups of university students (n=342) described in Chapter 7 were involved in the first part of this study. The original three groups were investigated individually and not as one large group due to the nature of the questions raised. The reported reading difficulty with assigned course readings is explored only in the case of the two academically more advanced groups, as incoming first-year students at the beginning of their studies had very limited experience with course materials at the time of the data collection.

Data for predicting reading difficulty was taken from the receptive Levels Test. As it was discussed in Chapter 5, research has identified a partial receptive knowledge of a minimum of 3,000 word families as a threshold level for decoding a general text in English. This would require students to possess the vocabulary of the 2,000 and 3,000 bands of the receptive Levels Test. As Nation (1990) treated a 83% score as the indication of a weak knowledge of a certain frequency level, an acceptance level for each frequency band was set at 90%, as had been done in previous research (see e.g. Tschirner, 2004). This still leaves room for three missed or wrong items out of the 30 at each level and indicates a stronger knowledge of the band than a weaker 83%. This also allows for a more direct comparison with results of other studies. For academic reading the threshold knowledge was identified as the 5,000 and academic levels, therefore, these two bands were also closely investigated.

As a second step, the data were analyzed in terms of Question 5/C in the questionnaire filled out by Groups 2 and 3 (see discussion of the questionnaire in Chapter 7 and the questionnaire in Appendix D). The answers on a four point scale, related to students' perceived difficulty of reading course materials, were compared to their VLT scores on the

selected three bands. Answers in the questionnaire ranged from no difficulty to serious difficulty to the point of not reading the assignments.

9.2.2.3 Results

The percentage of subjects who do not reach the threshold level for reading general texts is reported in Table 9.1. The analysis of the 3,000 word level of the VLT revealed that 65.5% of the incoming-first-year students, 47.5% of students at the end of their first academic year and still 9,7% of students already at the end of their third year do not have an acceptable knowledge of this frequency level. The number of those who can be expected to have difficulty in reading due to the lack of this minimal basic vocabulary sharply decreases from year to year, the improvement is especially large between the two higher level groups, namely end of the first year and end of the third year.

Table 9.1 Percentage of students with predicted reading difficulty of general English texts

Group	Percentage of students with VLT 3,000 < 90%
Incoming first year (n=148)	65.5%
End of first year (n=101)	49.5%
End of third year (n=93)	9.7%

Although the upper groups' results may partially be a reflection of selecting criteria and drop-out rates during and after the first year, it is disappointing and alarming to see students at the end of their third academic year having difficulty with this frequency band, giving indication about general text reading. If this is the case with the 3,000 word level, even more serious problems are expected with academic reading.

Therefore, a similar analysis was performed on the threshold level for reading academic texts. This investigation related to the 5,000 word family level and the academic vocabulary is reported in Table 9.2. Figures in this case are much higher than in the case of scores calculated for general texts. 74.3% of students at the end of their first year do not have

adequate knowledge of the 5,000 level, and still 45.2% of students at the end of their third year show lack of knowledge of this level. This implies the foreseen major problems with understanding academic texts which constitutes part of students' course materials. A significant increase in the academic vocabulary is seen from year one to year three, which is shown by the sharp drop in the number of students who do not meet the set level (74.3% in Group 1, 52.5% in Group 2 and only 7.5% in Group 3).

Table 9.2 Percentage of students with predicted reading difficulty of academic texts

Group	Percentage of students with VLT 5,000 < 90%	Percentage of students with VLT academic < 90%
Incoming first year (n=148)	87.2%	74.3%
End of first year (n=101)	74.3%	52.5%
End of third year (n=93)	45.2%	7.5%

An overall report on the predicted reading difficulty for both general and academic texts is illustrated by Figure 9.1. The bars clearly indicate the decreasing number of students who do not meet the threshold levels from year to year, although figures remain high even in the case of the subjects at the end of their third academic year. As has been pointed out, this may be partly due to the vocabulary gain of all involved subjects during their academic studies and partly to the drop-out rate of weaker students. This question will be further explored in a follow-up study in Chapter 10. What is important, however, is not only the rate of subjects who are expected to have comprehension problems due to their limited vocabulary knowledge, but the fact that if percentages of these students are turned into actual numbers, very many are struggling with reading.

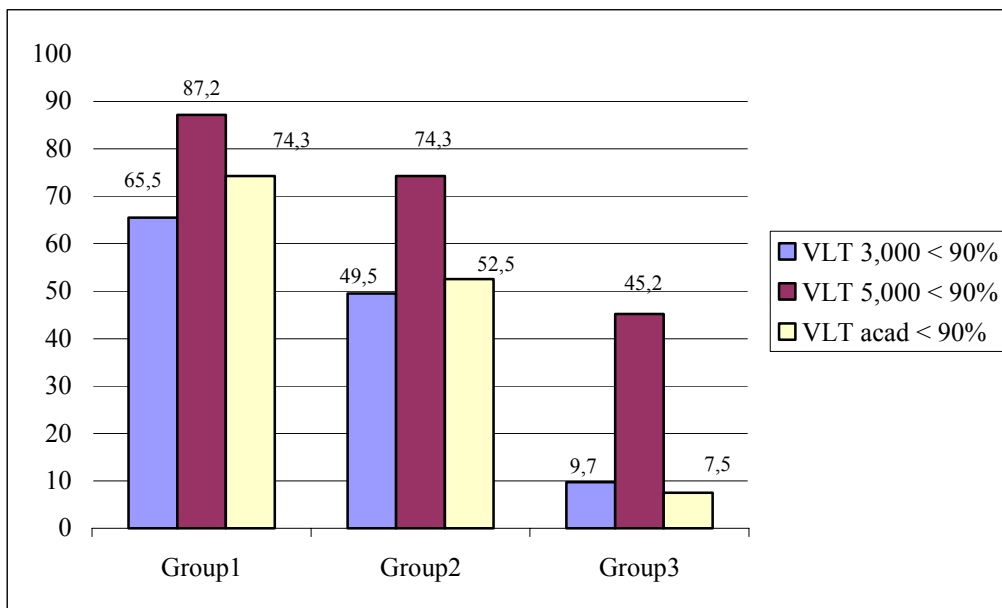


Figure 9.1 Percentage of students with predicted reading difficulty in the three study groups

Scores obtained at the three selected frequency bands of the VLT were also correlated with the reported difficulty of reading academic related texts in English. This information was gained with the help of the questionnaire. The specific question and the range of answers for the two upper student groups are reported in Table 9.3. It is clearly seen from their answers that there is a shift towards the middle answer (related to some difficulty) from year one to year three. Meanwhile, fewer students reported in the third year that they had almost never had difficulty with the assigned readings, and only two reported checking the dictionary or asking for help during reading. It needs to be stated, though, that the type of readings assigned is expected to change greatly from year one to three, from more general texts to a majority of academic texts in content classes.

Table 9.3 Reported reading difficulty of academic related texts (Question 5/C of the questionnaire)

<i>Do you have difficulties reading books/articles in English assigned as course material?</i>	End of first year n=101	End of third year n=93	Total n=194
1 = almost never	28	15	43
2 = sometimes it takes long, but I can do it without help	61	76	137
3 = I always need to check a lot of words in a dictionary or ask others to help	12	2	14

Both for Groups 2 and 3, students at the end of their first and third year, significant correlations were found between questionnaire responses and tests scores. For Group 2 correlations were moderate, -0.44 for the 3,000 level, -0.33 for the 5,000 level and -0.43 for the academic level ($p < 0.01$ for all levels). In the case of Group 3 lower, but statistically still significant correlations were identified for all three levels: -0.26 for the 3,000 level ($p < 0.05$), -0.29 for the 5,000 band ($p < 0.01$) and -0.25 for the academic word list ($p < 0.01$).

When correlating reported reading difficulty of assigned reading materials and the reported amount of reading done for academic studies, no statistically significant correlation was identified for either Group 2 or 3. This seem to suggest that those who read more academic related texts do not have less difficulty with reading this type of texts, or those who have difficulty do not strive for reading more of it for practice. However, these results need to be interpreted with some caution in the light of the data reported in Table 9.3, as the majority of students, especially in Group 3, indicated the “some difficulty” option and keeping in mind that selective readers are likely to read texts that do not pose great reading difficulties, even though none of the students chose the last option offered to this question, namely “the assigned materials are too difficult for me, therefore I do not read the majority of them”.

Now I turn to the discussion of another area of academic success, namely completing courses required by the syllabus.

9.2.3 Study 4B – Vocabulary knowledge and course failure

9.2.2.1 Overview and aim

The main purpose of this Study 4B is to explore whether low vocabulary test results are able to predict failure in academic studies in the target language. Both Laufer (1992) and Morris and Cobb (2003) suggested that reading ability, and vocabulary knowledge in general, have a predictive nature of academic ability or success on adult EFL student populations. This investigation, therefore, wishes to explore the diagnostic nature of vocabulary test results in terms of course completion or failure. Both language and content classes of first year students will be under investigation.

The following specific research questions will be explored in this section:

- 4B a How can vocabulary test results predict the academic success of first-year English majors at a Hungarian university?
- 4B b Is it possible to identify an at-risk group of students based on incoming vocabulary test results?
- 4B c Is vocabulary size a better indicator of academic success in terms of failing grades than lexical organization tested by the association test format?

9.2.2.2 Procedure

Subjects involved in this investigation were Groups 1 and 2, that is beginning and end of the first year students, as they follow a mandatory syllabus as opposed to students in Group 3 who have a wide selection of elective language and content courses. Groups 1 and 2 are treated as separate groups and scores will be calculated for each separately.

The data were analyzed in two ways. First, the overall scores obtained on the three data collection instruments of vocabulary knowledge were compared to the failing grades in first-semester language seminars for students in Group 1. A similar procedure was carried out for Group 2 who took the vocabulary tests a month before the comprehensive exam. The fact that some students did not take the mandatory exam means that they had not completed all courses required in the first academic year. Therefore, their scores are expected to have a strong diagnostic nature in foreseeing the success or failure of completing all the courses in the first-year syllabus.

9.2.2.3 Results

When looking at academic success of incoming first-year students in terms of failing grades in language classes, data show a statistically significant negative correlation between all test results and failures in the case of Group 1. Results (reported in Table 9.4) suggest that

students who do not reach a 60% overall score on the receptive Levels Test (or 80% at the 2,000 level, 60% at the 3,000 and academic level, and 50% at the 5,000 level) fail language classes in their first semester, with only some exceptions. A similar passing mark can be identified in the case of the productive Levels Test: a 30% overall score (or 60% on the 2,000 level, 30% at the 3,000 and academic levels, and 15% on the 5,000 level). The threshold level can be drawn at 35 points, 39% for the Lex30. An at-risk group with still a great number of failures can also be identified including those who score above these values, but do not go beyond an 80% on the receptive Levels Test, 50% on the productive version of the Levels Test and 44 points on the Lex30. Of course, these are overall group results, and individual variability allows for successful passing of courses with relatively low initial test scores. Similarly, since seminar grades have many components, including attendance and class work, students with little motivation are more likely to fail courses even with good vocabulary test scores.

Table 9.4 Predictive nature of academic success of vocabulary test results of incoming first-year students

Group	VLT overall scores (max. 100%)	PVLT overall scores (max. 100%)	Lex30 (max.90 points)
Likely to fail first semester courses	< 60%	< 30%	35 points
At-risk group	< 80%	< 50%	44 points

As for students at the end of the first year, the analysis of successful completion of all courses in the first academic year (indicated by the admittance to the end-of-the-year language proficiency exam) revealed that there was a statistically significant difference in the mean VLT and PVLT scores between those who completed their first year and those who did not. The Lex30 showed no predictive nature as differences in the mean scores were not statistically significant. Mean test scores for the passing and failing students are reported in Table 9.5. The relatively high standard deviation figures which indicate a partial overlap between the scores of successful and unsuccessful students, call for some caution in attributing definite and conclusive values of test scores in terms of syllabus completion. They are, indeed, meant to be indicative and not conclusive. Moreover, individual differences

within groups support the fact that vocabulary size is not the only factor influencing successful class completion. Some students who had scores well above the average were not admitted to the comprehensive exam a month later, while other students with low scores managed to complete all course requirements.

Table 9.5 Differences in mean test scores between academically successful and not successful students at the end of their first year

Group	VLT overall scores (percentage)	PVLT overall scores (percentage)	Lex30 (max.90)
Unsuccessful group n=28 (first-year courses not completed)	71.46 SD=12.977	43.36 SD=14.322	43.79 SD=8.071
Successful group n=73 (first-year courses completed)	79.58 SD=11.201	52.67 SD=15.209	44.75 SD=8.997

9.2.3 Discussion of the results in Study 4

The data gathered through vocabulary tests, questionnaires and course grades were able to answer all the specific research questions raised in this study. Test scores predicting reading difficulty show a surprisingly high proportion of students who seem to have a serious lack of recognition vocabulary to the point of not being able to fluently read even general texts in the target language. The figures prove to be even more alarming in the case of academic texts which are part of the second and third-year syllabus in content classes.

The theoretical implications of the results call for the need to reconsider the nature of reading and meaning inference from text of the student population under investigation. As discussed in Chapter 5, according to Haastrup's (1990) model, we can differentiate between 'top down' and 'bottom up' inference models. Haastrup notes that it depends on the knowledge of vocabulary and subject matter brought into the reading activities which strategy is employed during reading. The 'top down' strategy implies that readers have a very good familiarity with the subject matter, but show lack of language knowledge. In this case, a reader can infer the meaning of many unfamiliar lexical items drawing on topic knowledge in his L1, but vocabulary learning is minimal. The 'bottom up' meaning inference, on the other hand, requires good language knowledge, but little subject knowledge. In this second case,

most of the vocabulary is known to the reader and unfamiliar word meaning can be inferred and learnt. When taking a closer look at these two strategies, it can be seriously questioned that subjects involved in this research have a good topic familiarity, as they are taking classes across a variety of subject matters from linguistics to history. Second, it is also questionable, considering their Levels Test data at the three specific frequency bands, that students have the required vocabulary knowledge necessary for the 'bottom up' inference strategy. Third, as has been pointed in Chapter 5, incidental vocabulary learning from reading is possible only if we can talk about extensive reading. This would apply the continuous reading in the target language which could be expected from students of English at the tertiary level. However, the number of pages read each week for the university or for any other purposes reported in Chapter 8 seriously questions the extensive nature of reading done by the majority of the students. Only a restricted number of both first and third-year students reported reading on average more than 100 pages per week that could be already considered extensive. Fourth, as it has also been discussed in Chapter 5, repetition plays a crucial role in the acquisition and recall of new vocabulary items. If students do not read the assigned course materials and many of them do not even engage in reading for pleasure, then the acquisition of only a very limited number of new lexical items can be expected to happen. The nature of change in students' vocabulary will be treated more in detail in the following chapter, with reference to explicit and implicit learning strategies. On the whole, it can be concluded that vocabulary gain from reading is expected to be minimal and text comprehension is seriously limited in the light of the lack of the vocabulary threshold levels and academic topic matters. The picture is even more alarming if we consider the full model of reading and text comprehension presented in Chapter 5 which included motivation, reading strategy and practice. Neither strategies of meaning inference are expected to be successfully employed by students, which leads to limited reading, limited understanding of text meaning and limited gain in vocabulary size and quality.

These considerations lead to the practical implications of the results. When scores on the 3,000 level of the VLT of the incoming first-year English majors are compared to the scores reported for the German students by Tschirner (2004), results reveal that Hungarian students are at a more advantageous position compared to their German counterparts (49.4% vs. 28% reaching the set 90% acceptability mark). Still, half of the Hungarian students upon

entering the university as English majors are expected not to have the minimum receptive knowledge of form and meaning which would be the basis for the successful decoding of the meaning of even general texts. A recent study (Lehmann, 2007) conducted among first- and second-year Hungarian students of English at another university confirms the lack of vocabulary knowledge necessary for target-language tertiary studies. The author, using other testing methods, concludes that 15% of her subjects do not reach the threshold level of 5,000 base words, and 94% of the 10,000 words, a stricter level of English defined as necessary for target-language undergraduate studies. Although differences in testing methods make it difficult to directly compare the results, it is evident that many students are expected to have serious language difficulties during their studies. This may lead to the point of not being able to follow courses, fail in courses and eventually drop out of the university. This is confirmed by data related to course completion and vocabulary knowledge test scores. As has been confirmed by the results, test data have a strong predictive nature of successful academic studies, which could be used for ongoing diagnostic purposes.

9.3 Study 5 – Vocabulary knowledge and use in written text production

9.3.1 Aim

This study, similarly to Study 4, goes beyond the descriptive analysis of test data, this time by discussing written production tasks produced by the subjects. The focus of this study is the relative role of vocabulary knowledge in the lexical richness measures of L2 written production tasks. It explored whether a larger lexicon leads to more sophisticated vocabulary use in writing. As has been pointed out in Chapter 4, this type of analysis is expected to give us a different type of knowledge of student essays that large learner corpora can offer.

The main research question in Study 5 is stated as the following:

- 5 What is the relative role of vocabulary knowledge (as measured by the vocabulary tests in Study 1) in the lexical richness of L2 written production tasks? In other words, does a larger lexicon lead to more sophisticated vocabulary use during written production?

When exploring the role of vocabulary knowledge in the actual use of vocabulary in student essays, the following specific questions will be investigated:

- 5a Do results on the two types of productive tests correlate to the same degree with lexical measures in students' texts?
- 5b Do we see a higher proportion of academic words in the argumentative essays of students who have a larger productive academic vocabulary as tested by the PVLТ?
- 5c What is the degree of relationship between the different lexical measures (lexical profile, type/token ratio and lexical density)? In other words, do these measures correlate with each other?

9.3.2 Procedure

Subjects involved in this investigation were selected from Groups 2 and 3. Only those students were included in this study who had taken the comprehensive language exam at the

end of the academic year when they were tested, since essays were produced as part of this examination. This meant excluding 28 students from Group 2, and 4 students from Group 3. Therefore, 78 first-year and 89 third-year students participated in this study. Since Group 2, Group 3a and Group 3b received different writing tasks and prompts, their results will be treated separately, and not as one large group.

Of the three tests of vocabulary knowledge, the two targeting aspects of productive knowledge were selected. Overall test scores were retrieved for all selected students.

As discussed in Chapter 7, essay writing was a regular part of students' end-of-the-first and third-year Academic English comprehensive exam. This assured that all students within the same study group had the same circumstances for writing. The importance of working on the written task under controlled conditions and without the use of reference materials was pointed out in detail in Chapter 7. For first-year students, 75 minutes were given for a 220–250 word essay, and 90 minutes were provided for third-year students for writing a 300–350 word essay. Group 2 was offered four prompts, two expositions and two argumentations, each consisting of an approximately 2–3 sentences. Group 3a was provided with a choice of three prompts, all argumentative essays, with two argumentative sides to choose from. This meant altogether six choices. Students in Group 3b, those who took the comprehensive language exam a year later, had a choice of two, instead of three prompts due to changes in the exam policy. This still gave students a selection of topics (For essay prompts see Appendix E). This role of topic choice will be further analyzed in Study 6.

Due to the sensitivity of the lexical measure to text length, the essays written by the subjects had to be standardized in terms of length and spelling, as discussed in Section 7.5. This was done with the effort of leaving out the least number of words possible. This resulted in 200 word texts for the first-year students and 300 word essays for the third-year students. Since groups were treated separately, this length variable did not pose problems in the data analysis.

The lexical profile of the texts, following Laufer (1998), indicates the proportion of the first 2,000 most frequent words. The type/token value indicates the ratio of different words and all words in a text. Lexical density figures show the ratio of content words and function words.

The condensed lexical profile argued for by Laufer (1998) does not further analyze academic and off-list (less-frequent) words. It was, however, interesting to see a separate analysis of the academic vocabulary of the essays of third-year students whose data showed the highest scores on the academic word level of the three main study groups. In addition, they are the ones who all produced argumentative essays, avoiding possible genre problems present in the first-year group, some of whom wrote expositions.

9.3.3 Results

Great variability in the values related to the lexical measures of the essays was reported. Since the various groups of subjects had different writing tasks, results will be discussed in detail for each group separately in the following sub-sections. Before doing that, the range of the three lexical measure values are reported for the two groups in Table 9.6 in order to facilitate the analysis of the results discussed in the two sub-sections.

Table 9.6 Range of lexical measures of essays written by first and third-year students

Group	Lexical profile (proportion of the first 2,000 word families)	Type/token ratio	Lexical density
End of first year	84–99%	0.46–0.65	0.38–0.55
End of third year	77.67–95.37%	0.43–0.62	0.43–0.58

9.3.3.1 Results for first-year students

The lexical profile of the essays produced by first-year students at the end of the academic year (as shown in Table 9.6) ranged between 84 and 99%, showing that between 84 and 99% of all running words in the texts belonged to the most frequent 2,000 word families in English. This means that it was possible for a student to write a 200 word essay using only two words not falling into the first 2,000 most frequent English word families. The type/token ratio ranged between 0.46 and 0.65, indicating the ratio between different words and all words in the texts. The higher this figure, the less diverse the text is in terms of vocabulary. The lexical density figures showed that between 38 and 55% of the running words in the student texts were content words, as opposed to function words.

Beyond the descriptive statistics of the lexicon of the essays, attention was given to the relationship between the productive test scores and the three lexical measures. Results are reported in Table 9.7. Figures indicate that the overall productive Levels Test scores had a moderate and statistically significant effect on the lexical profile and lexical density figures of the essays ($p < 0.01$). The negative correlation values in the case of the lexical profile imply that, on average, a larger productive vocabulary size led to a small proportion of frequent words employed in essay writing. However, only a weak and statistically less significant correlation value was found for the association task. Results seem to indicate that the lexical organization measured by this test is not a strong indicator of any of the lexical measures. Indeed, no statistically significant relationship was found between scores obtained on this productive test and the lexical density and type/token figures of the essays.

Table 9.7 Correlations between productive vocabulary and lexical measures of essays produced by first-year students ($n=73$)

	PVLT overall	Lex30
Lexical profile	-0.41**	-0.27*
Lexical density	0.13	0.22
Type/token ratio	-0.01	0.12

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

It was also of interest to analyze how strong the correlation figures were among the various lexical measures, as addressed by research question number 3 in Chapter 8. When comparing the three different lexical measures of these first-year students' essays, using Pearson's correlation, a strong correlation ($r = -0.52$, $p < 0.01$) was found between the lexical profile and the lexical density values. Less strong, but still significant correlations were reported between the type/token ratio, and both the lexical profile and the lexical density of essays (see Table 9.8).

Table 9.8 Correlation matrix for the three lexical measures

	Lex. Profile 1st year	Lex. Profile 3rd year	Lex. density 1st year	Lex. Density 3rd year	Type/token 1st year	Type/token 3rd year
Lex. profile	1	1	-0.52**	-0.30**	-0.26*	-0.11**
Lex. density	-0.52**	-0.30**	1	1	0.38*	0.12
Type/token	-0.26*	-0.11	0.38*	0.12	1	1

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

9.3.3.2 Results for third-year students

A similar set of analyses was performed for the third-year students. The descriptive statistics (reported in Table 9.4) in their case also show great variability in the values of the different lexical measures. The lexical profile figures ranged between 77.67 and 95.37, the type/token ratios between .43 and .62 and the lexical density figures between .43 and .58. These mean better overall results than in the case of the first-year students.

The significance of relationship among the three lexical measures was calculated using Pearson's conduct correlation. Correlation values were less strong than for the first-year students. Table 9.9 reports the strength of relationship between productive test scores and lexical measures of texts. No statistically significant correlation was found in the type/token ratio and between the Lex30 association test and any of the three lexical measures. If compared with the correlation results of the first-year essays reported in Table 9.5, different correlation figures were seen. For both groups lexical profile seems to be correlating best with the productive test results and type/token ratio being the least influenced one of the measures by the productive vocabulary of the authors' of the texts.

Table 9.9 Correlation between productive vocabulary and lexical measures of essays produced by third-year students (n=89)

	PVLT overall	Lex30
Lexical profile	-0.36**	-0.19
Lexical density	0.28**	0.15
Type/token ratio	-0.17	0.13

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

In order to answer research question 5b, the academic vocabulary use in the third-year essays was dedicated more attention. Pearson correlation was conducted between the academic word level of the PVLТ and the proportion of the academic words used in the essays. The use of academic words in the 300 word essays ranged between 2.33 and 15%. This is a large amount of variability, meaning the use of 7 to 45 academic words within the same length of writing. When compared to the subjects' academic word knowledge, data suggest that there is only a weak, but statistically significant relationship between these values ($r=0.21$, $p<0.05$).

The analysis relating to the third research question was also carried out among this student group. When comparing the three different lexical measures of these students' essays, using Pearson's correlations, r values were found to be less strong than those seen in the first-year group (see Table 9.6). A moderate positive correlation was identified between the lexical profile and the lexical density ($r=-0.30$, $p<0.01$). Weak, but still significant correlations were reported between the type/token ratio and the lexical profile ($r=-0.11$, $p<0.05$). However, no correlation was found between lexical density and the type/token ratio.

9.3.4 Discussion of the results in Study 5

The productive vocabulary size measured by the Levels Tests seems to be a better predictive instrument of the vocabulary use in written production tasks than the associative links assessed by the Lex30. This implication is rather unexpected in the light of the aims of the test. As was discussed in detail in Chapters 4 and 7, this instrument has been designed to provide an alternative measurement instrument for assessing written assignments which contain a high proportion of function words. Test designers hypothesized to be able to predict productive vocabulary use on the basis of this type of productive test results, as Lex30 associations could be treated as forming lexically rich texts almost completely free of function words (Meara & Fitzpatrick, 2000; Fitzpatrick & Meara, 2004). Data in the present empirical investigation do not support this hypothesis. The Lex30 figures did not show strong and statistically significant correlations with the three different measures of the lexical richness in texts. Only one weak correlation figure was found and only in the case of one group and not

the others. The nature of the Lex30 will be further discussed in the overall discussion of results in Chapter 11.

Results of the two groups of subjects, namely first and third-year students, show different degrees of correlation values. The lexical profile, the proportion of frequent words in the texts, seems to correlate with the productive vocabulary size of students, even though the correlation is moderate. A weak correlation was found when looking at the academic words measured by the PVLIT and those used in the text. These findings are in line with the discussion carried out in Chapter 5, where the role of vocabulary and the complex nature of writing as a process were discussed. As has been noted in this chapter, writing involves many aspects that the writer needs to take into consideration, and lexical choice is only a part of it. How marginal this part is, though, needs to be shown in the next study, where alongside with the writing product, the process is also touched upon. This will be possible when discussing the attention students reported to pay to vocabulary during essay writing.

The lexical density and type/token ratio seem not to be significantly influenced by measured receptive vocabulary size. This may partly be due to the fact that the values move within a restricted range for these two measures.

In conclusion, the variability in the descriptive data and in the correlation figures calls for the need for further analysis of the lexical profile of the essays. It needs to be explored what may cause the differences within and between the groups. Some of the factors that are expected to have direct bearings on the lexical richness of essays produced by the subjects under controlled circumstances will be evaluated in the next sub-section.

9. 4 Study 6 – Factors influencing the lexicon of written production tasks

9.4.1 Aim

In this study some factors are investigated which are expected have a direct effect on the lexical parameters of the written production tasks. The following research question will be addressed in this section:

- 6 To what extent can differences in the vocabulary use in written production be explained by a) students' experience of producing texts in the target language, b) students' stated overall text writing strategies, including their view of the role of vocabulary in text production and c) topic choice?

The first factor to be explored is students' experience producing longer texts in English, then some of their text writing strategies, such as the avoidance of vocabulary items not fully known or language used for essay planning. It is also investigated how their view of the importance of lexical choice (compared to grammar, organization and spelling) effects their use of less frequent words in essays. Indeed, it needs to be analyzed whether those who claim to pay the most attention to vocabulary, do use a wider range of lexical items while writing. As a final area within this research question, the relationship between topic choice and vocabulary is investigated from two different points of view. First, it is explored whether lexical proficiency plays a significant role in topic choice. This investigation is made possible due to the fact that each group of students was given a choice of essay topics. It is investigated whether students with higher and lower lexical proficiency show a preference for any of the topics offered to them. Second, the effect of topic is investigated from the opposite direction by exploring whether certain topic or genre choice pre-determines the lexical richness of the essays. Answers to this last area of research will be of utmost importance for essay writing instruction and exam task elaboration.

The main area of research can be broken down into the following specific research questions:

- 6a Do students who report more experience producing longer texts in English show better lexical profile in their essays?
- 6b Do stated overall text writing strategies influence the lexical profile of written texts?
- 6c Do the prompts themselves influence the lexical profile of the written productions even within the same genre? In other words, is there a difference in the lexical profile values of texts written about different topics?
- 6d Can receptive vocabulary size be identified as an indicator of students' topic choice?

9.4.2 Procedure

Participants involved in this study were the same as those included in Study 5, more specifically 78 first-year and 89 third-year students. Special attention was paid to the differences between students drawn from Groups 3a and 3b, first, because they were given the choice of different numbers of essay prompts, and second, because they had different vocabulary test results, as was discussed in Chapter 8.

Apart from data of the lexical profile of texts, questionnaire data referring to the experience with writing texts in the target language and the stated overall writing strategies were included in the investigation. A question was also formulated in the questionnaire related to students' experience with writing longer texts (more than 10 sentences long) in their L1. This makes it possible to see whether limited L1 and L2 writing experience do correlate and add up to overall writing experience.

9.4.3 Results

9.4.3.1 The influence of writing experience

As students with more experience in writing longer texts in English were expected to produce lexically richer texts, participants were asked about this experience in the

questionnaire. Results show that students overall have very limited practice of producing texts consisting of a minimum of ten sentences in the target language. First-year students had an average of 3.55 such occasions per week, while third-year subjects reported an average of 3.45 occasions of writing longer texts per week. Figures ranged between 1 and 25 occasions. It is interesting to compare this target-language experience with the same type of writing activity carried out in Hungarian (3.15 for Group 2 and 3.61 for Group 3). These values could hardly be considered extensive writing practice in either of the languages.

When the writing experience is compared to the lexical measures of the texts, no statistically significant difference could be identified on a group level. This might suggest that students with more experience in producing texts in the target language do not necessarily write essays that are richer in vocabulary, although this interpretation of the data needs to be taken cautiously, as the analysis is based on students' stated overall writing experience without knowing the type of writing, the purpose, the length and the genre of these texts. Moreover, the majority of the students reported very limited experience which makes a difference between the "more" and the "less" writing reported with the help of the questionnaire very small.

9.4.3.2 The influence of overall writing strategies

In order to answer research question 6b, questionnaire data were again examined. Question 7 in the questionnaire (see Appendix D) referred to the attention paid to sophisticated vocabulary choice during writing in the target language in general. It was expected that those who claim to dedicate more attention to the aspect of sophisticated vocabulary use in essays (versus grammar, organization and spelling), produce lexically richer texts. Table 9.10 reports on the attention paid to these different aspects while writing target-language texts. Data show that there is a visible difference among the three groups in terms of the importance of the various factors during writing. Sophisticated vocabulary choice seems to be the most important factor. It is interesting to note that vocabulary, as the most important aspect to consider, sharply drops from year one to year three. For many students it is only a second or third aspect to consider, while for slightly over one fourth of the students' vocabulary remains the least important factor across the years. Spelling, which may influence

vocabulary choice, is ranked last by a considerable number of students. Spelling seems to lose its importance to the third year, while organization seems to recuperate its importance in text production.

As significant differences could be identified for the lexical measures of texts produced by students in Group 3a and 3b, their choice between the four factors was given some further attention. It could be seen that students in Group 3a had a more grammar oriented writing, while students in Group 3b were paying more attention to vocabulary.

To conclude, vocabulary does not seem to be the priority of choice between the several factors involved in writing, of which only four were asked to be rated. Sophisticated vocabulary choice seems to be on the third place for most students after grammar and organization.

Table 9.10 Importance of four factors in essay writing

		Group 2	Group 3a	Group 3b
Organization	1st choice	30.7	37.8	37.5
	2nd choice	14.9	13.3	25
	3rd choice	15.8	17.8	18.8
	4th choice	38.6	31.1	18.8
Vocabulary	1st choice	17.8	2.2	8.3
	2nd choice	19.8	20.5	27.1
	3rd choice	35.6	51.1	37.5
	4th choice	26.7	26.7	27.1
Grammar	1st choice	38.6	44.4	37.5
	2nd choice	34.7	42.2	37.5
	3rd choice	16.8	13.3	16.7
	4th choice	9.9	0	8.7
Spelling	1st choice	12.9	15.6	16.7
	2nd choice	32.7	24.4	12.5
	3rd choice	30.7	17.8	25
	4th choice	25.8	42.2	45.8

Values are expressed in percentages

As a second step in the investigation of overall writing strategies, students were asked to report on the strategies they use during written text production under controlled circumstances when no help is allowed to be used, such as dictionaries, sample essays, topic-related reading, friends or relatives. Subjects first reported on their strategy of L1 versus L2 use during essay planning. 7.9% of Group 2 indicated to plan and draft an essay in L1 and then translate it sentence by sentence. This type of strategy was not reported by Group 3. A mixed language use during planning was reported by 57.4% of Group 2, 84.4% in Group 3a

and 66.7% of Group 3b. The third option, according to which they write down only ideas that come to their mind directly in English, was chosen by 34.7% of students in Group 2, 15.6% of students in Group 3a and 33.3% of students in Group 3b. Although these results show a visible difference between individuals within and among the groups, no statistically significant difference could be identified between this planning strategy and the lexical parameters of essays.

Similarly, statistically significant differences were not found in the case of the other two strategies related to lexical choices during essay writing. The overwhelming majority of students in all groups reported using another lexical item close in meaning to one that does not come to their mind (93.1% in Group 2, 95.6% in Group 3a and 95.8% in Group 3b). Only the remaining few percent claimed to leave the information out if an English equivalent of a Hungarian word or phrase does not come to their mind during writing.

The last strategy, related to possible spelling problems, revealed that around 70% of the students in all groups use a lexical item close in meaning if the given item's spelling is problematic. The remaining 30% guesses the spelling and uses the vocabulary item in their text.

9.4.3.3 Topic choice as an influencing factor of the lexical profile of texts

The previous area of investigation revealed a possible influencing factor of vocabulary size on the choice of topic to be elaborated on. A crucial point in the investigation of the role of topic choice is whether the prompts themselves determine the vocabulary used in the essays. This question was again explored independently for the three groups due to the different number and types of prompts.

9.4.3.3.1 Students from Group 2

Table 9.11 reports on the lexical parameters of texts written for the four different topics. The proportion of the four chosen topics seems uneven, with very few students elaborating on topics 1 and 2, the expository essays. A continuous increase in the percentages of the most frequent 2,000 word families is seen in the results. For the lexical density measures, a clear difference is seen between the two argumentative and the two expository

essays. The type/token ratio also shows identical mean values for two prompts (topics 2 and 4) and different values for the other two prompts. The statistically significant nature of these differences was tested using ANOVA. The analysis of variance ascertained the statistically significant nature of the difference in the case of the lexical profile and lexical density ($F=14,931$, $DF=3$, $p<0.01$ and $F=4.964$, $DF=3$, $p=0.003$, respectively). In order to reveal whether the significant difference was there for all pairs of topics, Tukey's post hoc test was performed. As for the lexical profile, differences between topic pair 1–3 ($p=0.025$), pair 1–4 ($p<0.01$), pair 2–4 ($p<0.01$) and pair 3–4 ($p<0.01$) were found to be highly significant. A similar post hoc test was run on the lexical density of essays, which revealed statistically significant difference only for topic pairs 1–3 ($p=0.014$) and 1–4 ($p=0.016$). A similar set of relationship could not be identified for the type/token ratio.

Table 9.11 Lexical measures of essays written by Group 2 compared to topic choice

Topic	Lexical profile	Lexical density	Type/token ratio
1 (n=3)	88.33	0.53	0.59
2 (n=7)	90.57	0.50	0.57
3 (n=27)	92.58	0.46	0.55
4 (n=36)	95.22	0.46	0.57

In order to partially balance for the small number of items in the first two topics and to capture a more general difference between the two expository and the two argumentative essays, the four groups of topics presented in Table 9.7 were condensed into two groups, based on the two genres (topics 1 and 2 being expositions and topics 3 and 4 being argumentations). Independent samples t-test found statistically significant differences in the lexical measure values between the two groups as follows: for the lexical profile $t=4.51$, $DF=70$, $p<0.01$; for lexical density $t=-3.64$, $DF=70$, $p<0.01$. No statistically significant difference was found for the type/token ratio. Whether these differences could be partly explained by the genre differences or by the prompted topic in general, cannot be fully

ascertained. The genre difference, though, will be removed in the two third-year groups, discussed in the next sections.

9.4.3.3.2 Students from Group 3a

Results for Group 3a are reported in Table 9.12 where scores between essays written in the three topics show clear and visible differences, especially regarding the lexical profile and the type/token ratio. The ANOVA showed the statistically significant nature of the difference for these two lexical measures, showing values $F= 43.932$, $DF= 2$, $p<0.01$ for the lexical profile and values $F=5.368$, $DF=2$, $p<0.01$ for the type/token ratio. Tukey' post hoc test revealed that the statistically significant relationship is valid for all topic pairs in the lexical profile, but only for topic pair 1–3 in the type/token ratio. Statistically significant relationship could not be identified for the lexical density measure, as these values were very similar for all three topics.

Table 9.12 Lexical measures of essays written by Group 3a compared to topic choice

Topic	Lexical profile	Lexical density	Type/token ratio
5 (n=13)	82.44	0.51	0.50
6 (n=16)	89.09	0.51	0.53
7 (n=11)	91.88	0.49	0.55

9.4.3.3.3 Students from Group 3b

As data in Table 9.13 show, the choice between the two prompts offered to this group was very much balanced. Furthermore, the essays written on the basis of the two prompts showed almost identical lexical parameters. In order to test this close similarity, independent samples t-test was performed which supported what was visible in the data: no statistically significant difference was found between the lexical measures of the texts written on the two topics.

Table 9.13 Lexical measures of essays written by Group 3b compared to topic choice

Topic	Lexical profile	Lexical density	Type/token ratio
8 (n=23)	89.493	.54	.51
9 (n=24)	90.362	.54	.51

If results of the three groups are compared, we can conclude that topic or the actual prompt have an influence on the variety of words used in essays. The difference between the topics was best seen when three choices were given to third year students. It needs to be stated, however, that it is not simply the number of choices offered to students, but also the close similarity between the assumed difficulty of the topics is crucial, especially in the case of exams, so that students are likely to write essays similar in their lexical measures regardless of the chosen topic, as was seen with Group 3b.

9.4.3.4 *Vocabulary knowledge as an influencing factor of topic choice*

The fact that students were given a choice of topics to select from leads us to ask whether vocabulary knowledge had a significant influence on the topic choice. This question had to be treated again in three separate discussions, since the three groups of participants were at different levels of lexical proficiency shown by results discussed in Chapter 8 and were asked to make a choice between varying numbers of topics. For these reasons the results of the three groups are treated and discussed separately in this section.

9.4.3.4.1 *Students from Group 2*

As has been seen in the previous section, subjects in Group 2 had a choice of four prompts, two expository and two argumentative essays. Topic choice and the vocabulary test scores of students who chose them are reported in Table 9.14.

Table 9.14 Relationship between vocabulary knowledge and topic choice for Group 2

Topic	VLT overall	PVLT overall	Lex30
1 (n=3)	88.33 SD=8.33	39.33 SD=34.08	52.59 SD=3.40
2 (n=7)	87.14 SD=5.40	61.57 SD=12.27	51.27 SD=13.33
3 (n=27)	84.04 SD=8.84	58.70 SD=14.95	49.88 SD=10.11
4 (n=36)	74.03 SD=11.23	47.53 SD=11.46	49.08 SD=9.84

Scores are expressed in percentage points

In order to test whether these differences in topic choice could be influenced by vocabulary knowledge to a statistically significant degree, one-way analysis of variance (ANOVA) was performed. Results show a significant difference in the VLT overall scores ($F=7.748$, $DF=3$, $p<0.01$) and similarly significant scores for the PVLT test scores ($F=5.101$, $DF=3$, $p<0.01$). Difference between pairs of groups was tested using Tukey's post hoc test. Both for the VLT and PVLT scores significant differences were found only between topic pair 3–4 ($p<0.05$). The Lex30 scores were found not to be clear influencing factors in topic choice.

9.4.3.4.2 Students from Group 3a

Students in this group were given a choice of three prompts, all argumentative essays. Topic 5 was elaborated by 13 students, topic 6 was chosen 18, while topic 7 was used by 11 subjects (see Table 9.15). A one-way analysis of variance revealed values slightly over the statistically significant values for the VLT ($p=0.069$), and showed significant difference in the case of the PVLT ($p=0.028$). Tukey's post hoc test was able to identify the pair of topics responsible for the difference. This was clearly seen only between the first and third topics. The Lex30 scores were found not to be clear influencing factors in topic choice.

Table 9.15 Relationship between vocabulary knowledge and topic choice for Group 3a

Topic	VLT overall	PVLT overall	Lex30
5 (n=13)	89.08 SD=5.99	71.92 SD=11.72	54.02 SD=9.58
6 (n=18)	86.28 SD=6.30	64.89 SD=8.94	50.80 SD=7.13
7 (n=11)	82.27 SD=8.80	60.91 SD=8.95	46.57 SD=14.11

Scores are expressed in percentage points

9.4.3.4.3 Students from Group 3b

The second sub-group of third-year students, tested a year later, could choose between two prompts. This resulted in a more balanced topic choice, as topic 8 was elaborated by 23 and topic 9 by 24 participants (see Table 9.16). An analysis using independent samples t-test indicating vocabulary knowledge test scores as test variables and topic choice as grouping variable revealed no statistically significant difference between the vocabulary knowledge of those who chose one or the other topic. Balanced topic choice was not clearly influenced by the lexical knowledge of the subjects, as measured by tests.

Table 9.16 Relationship between vocabulary knowledge and topic choice for Group 3b

Topic	VLT overall	PVLT overall	Lex30
8 (n=23)	90.04 SD=6.34	69.74 SD=9.73	55.31 SD=8.60
9 (n=24)	88.92 SD=4.06	68.75 SD=7.73	52.97 SD=9.17

Scores are expressed in percentage points

Overall, it seems from these analyses that topic choice is not in all cases clearly driven by the lexical knowledge of the subjects, as measured by tests, but neither is a casual choice driven by liking a certain prompt. However, this relationship seems to be greatly influenced by the number of prompts and the prompts themselves. When 3 and 4 choices were given, a preference was identified for certain topics based on the receptive and productive vocabulary test scores. When two prompts were offered to third-year students, topic choice and students'

lexical knowledge were more balanced, in other words, the two prompts were chosen by almost the same number of students involved in the investigation, and lexical measures also showed almost identical values.

Moreover, if topic choices are further analyzed at a subject level, while examining students with the highest and lowest test scores, it is revealed that some students with weak test scores produced essays with excellent lexical profiles on certain topics, while some other students with high test scores wrote essays that show weak lexical profiles. From this it seems that some prompts are more likely to allow for a more informal style and less sophisticated vocabulary choice and others push even weaker students to recall and employ more less frequent words. This was clearly seen in the case of Group 3a, where topic 5 prompted lexically weaker essays and topic 7 forced students to use more sophisticated vocabulary.

9.4.4 Discussion of results in Study 6

In Study 6, which targeted possible influencing factors of lexical measures in student essays, a rich set of results were identified. Writing experience in the target language and stated overall writing strategies related to language use did not show a statistically significant influence on the lexical measures of texts written under controlled circumstances. These results may be interpreted in different ways.

First, writing practice remains limited for most of the students, as they report on a few occasions per week of longer text production. For many of them this cannot be considered extensive writing practice. Second, stated overall strategies may not be so clear-cut during actual writing, but rather show a mixed nature. Third, as the data revealed, students do not consider vocabulary choice as a primary concern during text production, and related figures seem to decline as students advance in their studies. The difference between a more grammar and a more vocabulary oriented writing practice is evident when the two third-year groups are compared. Data analysis in Chapter 8 showed that Group 3b produced lexically richer texts, which can be partly explained by focus on vocabulary in the light of the questionnaire data showing more attention paid to vocabulary by these students.

The analysis of topic choice also led to important findings. In the case of the first two groups where three or four prompts were provided to choose from, data revealed a significant

influence of vocabulary size on the choices made. This supports the tendency seen in essay writing classes, namely that students, if tasks allow it, set up a difficulty hierarchy among prompts, and lexically less proficient students tend to choose a topic which can be completed with less sophisticated language use. Individual differences though, suggest that weaker students who chose a more challenging topic (in other words topics that cannot be discussed with limited vocabulary) can produce lexically rich texts and some of the most proficient students who elaborate on seemingly less demanding topics (those that allowed for more informal language use and limited vocabulary) produced lexically poor texts. This calls for the need of analyzing how certain prompts may effect or determine vocabulary use, as determined by word choice. For this investigation three different lexical measures were employed, of which the lexical profile and the lexical density show statistically significant differences between essays written on the different prompts. Topic effect on either of the lexical measures was not identified for third-year students who were given two prompts only.

It needs to be further noted, that for the discussion of the issue under investigation *topic* as a general term was employed, though students in each case were presented with a prompt and a task specifying the essays they were required to write within a wider topic. Piloting of the methodology has shown the importance of providing students with specific prompts rather than a single word or short sentence as a title. The correlations found between vocabulary measures, test scores and choice of prompts seem to suggest that in many cases topic choice is not simply a matter of liking or not liking the provided topic or prompt.

The implication of these results can be drawn as the following. On a more theoretical level, vocabulary use seems to clearly differ from vocabulary knowledge targetable by language tests. Even tests that are labeled as productive and are designed to resemble real language use cannot fully predict the vocabulary activated for certain situations. Tasks and specific prompts seem to have a direct bearing on the language use of learners, some asking for more sophisticated vocabulary use than others. On a more practical level, one of the major implications of these studies is the need to sensitize students to vocabulary use. Not only do they need to practice text production, but practice it while dedicating more attention to vocabulary choice alongside other parameters of writing. Furthermore, providing students with challenging writing tasks seems to push them towards more sophisticated language use and a greater variety of vocabulary choice.

9.5 Concluding summary

The aim of this chapter was to assess the lexical knowledge measured by tests in terms of its implications for the language use of subjects involved in the present empirical investigation. The significance of this tested lexical knowledge in predicting academic success and the use of this knowledge in written text understanding and production was the primary focus of the last three studies. The multifaceted nature of research questions made it possible to assess both theoretical and practical implications of vocabulary knowledge of students of English. Study 5 investigated the predictive nature of test results of vocabulary knowledge in terms of academic success in the target language. At-risk groups could be identified regarding text comprehension and course completion. Studies 5 and 6 directed the vocabulary use essays written under controlled circumstances and some influencing factors of vocabulary choice. It was investigated how essay writing strategies, practice and topic choice may influence three specific lexical measures of these essays.

My predictions concerning language use were only partly asserted by data analysis. Findings confirmed the intuition that students who enter a university as language majors bring very different levels of overall language proficiency with them, including lexical knowledge. This gap, however, seems to remain in the more advanced years. A gap that is large enough to be accounted for when designing the syllabus.

A practical outcome of this study is the information gained about aspects of Hungarian university students' vocabulary knowledge and language use that we could, so far, only make assumptions about on the basis of the coursework they were doing. As many students lack the threshold vocabulary knowledge needed for successful academic work, focused help needs to be provided in the form of extra classes that directly focus on building academic vocabulary. In addition, ways of improving students' lexis in the form of individual work can be suggested.

CHAPTER TEN

STUDY 7: CHANGES IN VOCABULARY KNOWLEDGE OVER ONE ACADEMIC YEAR

10.1 Introduction

The ability to write argumentative essays or develop the breadth or depth of one's vocabulary is a gradual process rather than consisting of large jumps which reflects the incremental nature of vocabulary development (Waring, 1999). In the previous chapters a gain in vocabulary knowledge was reported from study group to study group on the basis of their time spent in the academic environment. It has been shown that there is a positive change in vocabulary knowledge between incoming first-year students, those tested at the end of their first year and those investigated at the end of their third year. The question, however, inevitably emerges whether the difference between the results of one group and another one can be explained by a continuous gain in vocabulary with time or rather with a selection of more proficient students in the upper years, a high drop-out rate of less proficient students in the first year of their studies or a combination of these factors. It is of interest to see how real-time changes in vocabulary happen compared to the apparent-time changes investigated in the previous research questions. In the literature related to vocabulary knowledge reviewed in Part II of this dissertation, a need for longitudinal research has been voiced by many authors (e.g. Read, 2000). Compared to the large body of research concerning the descriptive data of different study populations, very little is known at all about the longitudinal changes in students' vocabulary knowledge, especially involving diverse knowledge types (Dóczi, 2007). The analysis in this chapter aims to contribute to the need for filling this gap by the parallel discussion of three sets of data of a small group of participants. This is the first time that these three test results are followed up in a longitudinal study supported by background information on students' target language use and learning strategies. This small-scale study, therefore, should be of both major theoretical and practical value for understanding the nature of vocabulary change and the implications that can be turned into syllabus design and instruction.

10.2 Aim

Study 7 is a case study involving 15 first-year students (indicated as Group 4 in Chapter 7), with the aim of following up on their vocabulary change over a period of one academic year. These students who were all tested in September 2006 as part of Group 1 were retested using all three testing instruments in May 2007. Moreover, background information was collected from them in order to explain possible reasons behind changes in their lexicon. This study aims to contribute to the need for understanding real-time changes in students' lexicon and also gives an opportunity to do qualitative investigation alongside with the statistical analyses.

As part of this chapter, the following broad research question will be investigated:

- 7 How does the lexical knowledge of English majors, as measured by vocabulary tests, change over one academic year?

There are at least two ways of looking at pattern changes. The first involves investigating the amount of change and the second is the direction of change. The amount of change is discussed in terms of change in the percentage point of the Levels Tests and change in the points awarded in the Lex30. The analysis of the Levels Tests results permits the discussion of each level of the VLT and PVLТ separately and of the overall scores. Data for group means is reported first, but since this can shadow some of the individual variability in the data, scores for each subject are also analyzed at the test and sub-test level. Individual variability is not dealt with at the item level, as we are not interested in each item separately, but in seeing overall patterns of change. The term 'change' has been and will be used in this chapter intentionally, as it needs to be verified whether an expected increase is the case for all levels and all subjects over one academic year.

The broad research question, therefore, will be divided into specific questions as follows:

- 7a What is the overall change for the group as a whole on each test and all levels of the two Levels Tests?

- 7b What are the individual differences among subjects?
- 7c How can individual differences in change be explained on the basis of a) incoming vocabulary knowledge, that is the assessed vocabulary knowledge of students in the first weeks of their tertiary education; b) time dedicated to learning and practicing the target language; and c) stated overall vocabulary learning strategies?

10.3 Procedure

Selection of the subjects into this case study was based on classroom participation, more precisely, all 15 students were enrolled in a Communication Skills language seminar in the second semester of their studies (14 females and 1 male). This class has the aim of target language development, including all four basic skills with a special focus on vocabulary building. The selection of students proved to be appropriate as the group includes both lexically weak and strong students. During the analysis of individual students, they will be referred to by numbers, such as S1. It will be of utmost importance to investigate what can explain the expected individual variability in the change in result patterns. Therefore, questionnaire data discussed in the previous chapter will be supported by structured interview data (for a description of the interviews, refer to Chapter 7).

10.4 Results

10.4.1 Changes at a group level

As a first step, descriptive statistics were carried out on the three test results for the subjects as a group in order to identify overall changes in the vocabulary knowledge of the study population. Minimum, maximum, mean scores and standard deviations are reported in Table 10.1 for the Vocabulary Levels Test, including both the test and the re-test data (September and May of the same academic year). This enables us to compare test results and see the changes in the scores. Overall group results at each level show a positive change in the results, all gains are significant at the $p < 0.05$ level, as shown by paired samples t-tests, except for the 2,000 level. Changing results do not only show a gain at each level in both the

minimum and maximum scores, but also decreasing standard deviation figures which suggest that the student group got more homogeneous in their results over the one-year period. Increasing figures are especially visible in the upper level minimum scores.

Table 10.1 Overall group changes in the Vocabulary Levels Test results over one academic year

	Minimum	Maximum	Mean	Std. Deviation
VLT overall	50	89	67.53	12.16
VLT overall/2	67	95	77.00	8.04
VLT 2,000	77	100	91.33	7.86
VLT 2,000/2	83	100	94.20	4.75
VLT 3,000	50	97	74.13	12.79
VLT 3,000/2	67	100	85.53	8.63
VLT 5,000	20	93	56.40	19.78
VLT 5,000/2	53	97	71.80	13.28
VLT 10,000	7	63	39.07	16.85
VLT 10,000/2	30	80	48.13	12.76
VLT academic	50	93	73.87	12.77
VLT academic/2	60	100	85.20	10.30

Scores are expressed in percentage points.

A similar descriptive statistical data set is reported for the productive version of the Levels Test in Table 10.2. Contrary to the productive test, the productive version did not show a gradual increase in both minimum and maximum scores was not found. An increase in the minimum scores was reported on all levels except for the academic band. Maximum scores changed in a most unpredictable way: scores at the 2,000, 3,000 and academic levels and the overall scores did not show higher maximum scores than those in September. On the other hand, the 5,000 level reported a higher maximum score, while the 10,000 level showed a decrease. Paired samples t-tests show significant difference in the mean scores of the two testing sessions for all pairs ($p < 0.05$), except for the 10,000 level. Standard deviation figures got smaller on each band, except for the academic level which remained identical in minimum and maximum scores. This implies that students' knowledge at the first three bands was consolidated, but whether this is a small gain for some and a large gain for others, or rather a

large gain for initially weak students and minimal gain for lexically more proficient ones, will need to be explored when analyzing individual student scores in the next section.

Table 10.2 Overall group changes in the Productive Vocabulary Levels Test results over one academic year

	Minimum	Maximum	Mean	Std. Deviation
PVLT overall	23	79	40.40	13.28
PVLT overall/2	40	79	50.87	9.95
PVLT 2,000	56	100	78.07	11.84
PVLT 2,000/2	72	100	85.93	8.87
PVLT 3,000	22	83	42.47	17.68
PVLT 3,000/2	39	83	59.40	12.22
PVLT 5,000	11	61	24.47	13.98
PVLT 5,000/2	17	72	31.93	13.14
PVLT 10,000	6	67	20.73	15.12
PVLT 10,000/2	11	50	24.87	12.32
PVLT academic	17	83	35.80	16.60
PVLT academic/2	17	83	35.80	16.60

Scores are expressed in percentage points

When comparing the two sets of data for the Lex30 association task, a gain in both the minimum and maximum scores is found (see Table 10.3). Paired samples t-tests show a significant difference between the two mean values ($p < 0.05$). However, contrary to the two Levels Tests, standard deviation figure increases and not decreases. This means that differences between individual scores can be expected to grow with time. This factor will be further analyzed when comparing test and re-test results at the subject level.

Table 10.3 Overall group changes in the Lex30 test results over one academic year

	Minimum	Maximum	Mean	Std. Deviation
Lex30	31	63	46.30	9.37
Lex30/2	40	74	55.41	10.76

Scores are expressed in percentage points

However, if the gain scores on the different tests and sub-tests are closely examined (as indicated in Table 10.4), the difference between minimum and maximum changes on certain test levels asks for a closer examination. Also, negative minimum values seem to indicate that the positive changes on the overall Levels Tests scores are not an indication of gains on all levels of the tests.

Table 10.4 Gain in scores on the three vocabulary tests

Changes in	Minimum	Maximum	Mean	Std. Deviation
VLT overall	2	23	9.47	6.39
VLT 2,000	-7	13	2.93	5.48
VLT 3,000	-3	27	11.47	8.27
VLT 5,000	-10	43	15.40	11.53
VLT academic	-7	37	11.20	11.55
VLT 10,000	-10	33	9.13	14.47
PVLT overall	-3	20	10.47	6.95
PVLT 2,000	-6	28	7.87	9.67
PVLT 3,000	-5	34	16.93	12.32
PVLT 5,000	-11	22	7.47	7.76
PVLT acad	-5	44	16.80	12.91
PVLT 10,000	-23	22	4.13	11.70
Lex30	-1	33	8.20	8.27

Scores are expressed in percentage points

A final issue that was raised concerning the results in change over time in the lexical knowledge of the subjects was whether factors, such as the time spent with English (both during the week and over the weekend) and the amount of overall and study-related reading could have an influence on the degree of change in subjects' test scores. Table 10.5 reports the correlation matrix for these factors. Results indicate that only the amount of time spent with English during the weekdays has a significant role in the amount of change in subjects' vocabulary knowledge measured by the VLT and the PVLT ($r=0.60$ and $r=0.58$, $p<0.05$, respectively). These issues will be further discussed in the following section which is dedicated to the discussion of changes in test scores at an individual level.

Table 10.5 Correlations for influencing factors on the changes in vocabulary test results

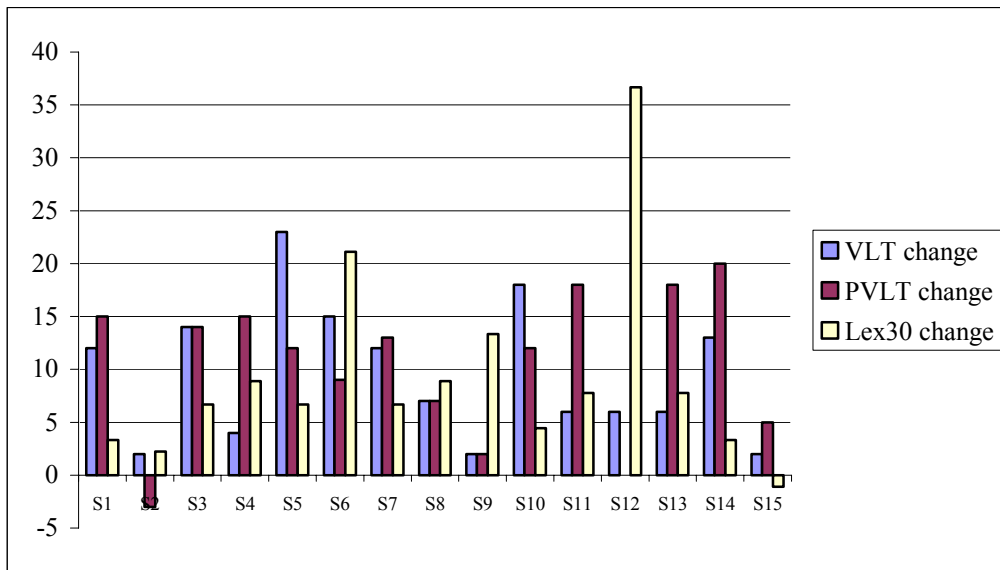
	Hours of English per day on a weekday	Hours of English per day at a weekend	Number of pages read per week	Number of pages read for university studies per week
Change in VLT overall scores	0.60*	0.34	-0.15	0.05
Change in PVL overall scores	0.58*	0.28	-0.01	-0.03
Change in Lex30 scores	-0.07	-0.16	0.28	0.16

* correlations are significant at $p < 0,05$, two-tailed

10.4.2 Changes at an individual level

The present section will summarize and comment on the changes in different types of vocabulary knowledge of students while drawing on data elicited by not only test results and questionnaire information, but also by short interviews to facilitate the understanding of individual variability. Results are reported and discussed for each subject individually.

As a first step, the overall changes in all three tests were analyzed for each individual subject. Results are summarized in Figure 10.1. What is evident from the data is that the different tests show a different degree of overall change with the PVL showing the most homogeneity, with changes ranging mostly between 5 and 20 percentage point gains. However, one student (S2) showed not a gain, but a minor loss in her productive vocabulary size, and another student (S15) in her Lex30 score. Changes in VLT scores proved to have a larger variability, but no negative figures can be seen. Lex30 changes are the ones with the largest swing in the percentage point change, ranging between 33 and -3.



Changes are expressed in percentage points.

Figure 10.1 Changes in the three test scores over one academic year by subjects

As has been discussed in the previous paragraphs, great variability is seen in how the vocabulary knowledge assessed by tests change over one academic year. On the basis of the individual changes, 7 subjects were chosen for the interesting nature of the change patterns they produced. Two cases of close to no change on the Levels Tests (Subjects 9 and 15), two students who produced the largest percentage point gains on the overall Levels Tests (Subject 5 on the VLT and Subject 14 on the PVL), two students who had the largest gain in the academic levels (Subjects 1 and 11), and the student who had the largest gain in the Lex30 (Subject 12). The detailed test results of these subjects are reported in Table 10.6. Results of these seven participants will be explored individually in the following paragraphs in order to find possible explanations for their special cases, by considering initial test scores, questionnaire and interview data.

Table 10.6 Changes in the two Levels Test scores of 7 subjects

	2,000 level		3,000 level		5,000 level		10,000 level		Academic vocabulary		Overall change	
	VLT	PVLT	VLT	PVLT	VLT	PVLT	VLT	PVLT	VLT	PVLT	VLT	PVLT
S1	-3	11	13	17	20	6	20	11	10	34	12	15
S5	7	6	27	11	20	17	33	6	27	23	23	12
S9	0	6	10	0	-10	0	0	-5	10	11	2	2
S11	-3	11	7	28	13	5	-3	0	17	44	6	18
S12	0	0	3	0	3	11	17	-23	7	11	6	0
S14	3	28	27	28	17	6	-3	17	23	23	13	20
S15	-7	11	10	12	13	-11	17	0	10	11	2	5

Changes in scores are expressed in percentage points.

Subject 9, who demonstrated no gain in her scores, had been studying English for 12 years at the time of the second testing session. In her questionnaire she reported on three hours per day activities in English on weekdays, and only one hour per weekends. She also indicated reading 30 pages per week in English. Her first VLT score was 78 and the PVL T score 41. This means a VLT score seven percentage points above group mean of the 148 first-year students, and two percentage points lower than mean PVL T score. She expected no changes in her scores as discussed in the interview. She reported to read “a lot”, but not to learn vocabulary through explicit strategies. Overall, it seems that the limited amount of time dedicated to English and the lack of explicit vocabulary reading may explain the fact that her assessed vocabulary size had practically remained the same over the academic year (two percentage point gain on both Levels Tests). She could not turn reading practice into implicit vocabulary gain.

Similarly, Subject 15 was the other student with no gain in vocabulary size. She had started university with ten years of English learning background. She indicated spending four hours on weekdays and ten hours at weekends with English. She read fifteen pages per week in the target language. She had initial 78 VLT and 41 PVL T overall scores (compared to the 71 and 43 group means). What is interesting, she herself expected no change in her vocabulary size as she reported paying no attention to explicit vocabulary learning strategies. She indicated reading texts on the Internet without much systematic attempt to gain vocabulary from these texts. Again, the lack of explicit vocabulary learning seems to be a plausible explanation for the lack of lexical gain.

Subject 5, on the other hand, showed the largest gain in her receptive vocabulary size (23 percentage points in her VLT). She had had ten years of English studies, and reported on eight hours of activities in English on weekdays, and 20 hours at weekends, which is a significant exposure to the target language compared to the previous two subjects. She indicated to read 20 pages per week in English. She had very low first test scores, 50 on the VLT and 36 on the PVL T. This means that even with this large gain, her vocabulary size is still below the group mean. At the interview she indicated these initial scores as “surprisingly low” for herself, and for that reason she had put extra effort to improve her lexicon. She reported to take notes of all unknown lexical items that she meets and to make all effort to

learn them. On the basis of these indications, it is not surprising that she had a large receptive vocabulary gain.

Subject 14, on the other hand, is the one who showed the most gain in the productive size test scores. She had been studying English for only five years and produced very low test scores in September (58 on the VLT and only 32 on the PVLТ). Her attempt to make up for falling behind her classmates was reported on by the nine hours with English per weekday and eight hours at weekends. Moreover, she indicated reading 60 pages per week in the target language. At the interview she reported feeling a large gain in her vocabulary thanks to the university language classes with focus on direct vocabulary building, and her effort to read journals and work out explicit learning strategies she was missing at the end of the academic year.

Two students with the largest gain in their productive academic vocabulary (34 and 44 percentage points, compared to the 13 points average gain) also deserve extra attention. Subject 1 had a few percentage points above average scores in September (77 on the VLT and 47 on the PVLТ). She reported ten years of English studies, seven hours target language-related activities on weekdays and eight hours at weekends. She also indicated reading fifteen pages in English per week. During the interview she explained her gain in vocabulary size by the fact that, first, she had had the opportunity for overall language proficiency gain having taken all language classes offered in the first semester, and, second, by the fact of using English excessively on a daily basis for various activities (including watching films and using the Internet).

Subject 11, contrary to Subject 1, had very low initial test scores (67 on the VLT and 39 on the PVLТ). This could be partly explained by the fact that she had been studying English only for four years. She reported on a significant amount of time spent with English, eight hours on weekdays and sixteen hours at weekends, including 50 pages reading per week. During the interview, similarly to Subject 14, she reported on putting a lot of effort into improving her English. She said that vocabulary building in language classes such as Writing Skills or Communication Skills were very effective. She also found more proficient classmates to be a boosting power in her effort to improve.

The last subject to be treated in detail was chosen because she showed large improvement in her Lex30 scores compared to the limited change in her other two test results.

S12's minor gain in the size test scores could be explained by the fact that she was one of the lexically strongest students when tested in September (89 VLT and 79 PVLТ scores). The following May, at the end of her first year, she reported on six years of previous English studies, the majority of which spent in a Hungarian–English bilingual high school. This could account for her well above the average test scores. High initial test scores could then explain the fact that there was little gain in her Levels Tests scores. However, the strong overall language skills, 50 pages per week reading and university language classes brought to a great change in her association test results. This seems to indicate that a large vocabulary size, daily language use for various purposes and explicit vocabulary building could explain the unexpectedly large change in her associations.

10.5 Discussion of results

Results in this follow-up case study, first of all, showed an overall group gain in all vocabulary test scores. Standard deviation figures at lower frequency levels decreased while at upper frequency bands they increased. This seems to suggest that the student group became more homogeneous in their vocabulary size at lower frequency bands, by weaker students trying to gain more scores through explicit learning strategies. Gains in the vocabulary size of the 5,000 and 10,000 word levels, however, are harder to predict, with some students improving significantly, while others even showing decrease in their scores. This leads to a more heterogeneous nature of these two advanced frequency level scores, showing larger standard deviation figures.

While overall group results seemed to be promising in the sense that they suggested overall gains on all tests and sub-tests, a closer investigation into the individual cases revealed that the change in many times was negative. This is in line with previous research which underlines the incremental nature of vocabulary knowledge. As Nations argues, “if learners have a sufficiently large vocabulary but they are not given the opportunity to put this vocabulary to use and develop skill in use, their growth in knowledge and further vocabulary growth will not be achieved” (1983: 132). As has been pointed out in the introductory chapter of this dissertation, the transition for these students from secondary school level to tertiary education means that they are expected to become more independent learners and develop

study strategies that work for their needs. This was supported by the experience of those students who had considerable gain in their vocabulary size over one academic year, while students with practically no gain were not putting effort and explicit strategies in learning new lexical items. The loss in some scores is not surprising in the light of the limited reading practice of students and the lack of explicit learning strategies that could balance for the insignificant gain through incidental vocabulary acquisition. Findings are in line with Hunt and Beglar's (2005) conclusion which confirms that implicit vocabulary learning should be combined with explicit learning strategies.

Subject-level analysis, furthermore, revealed that questionnaire and interview data are valuable methods to gain background information on subjects and they help to explain empirical results that are not evident from test scores alone. The data suggest that changes in vocabulary do not happen in an accidental nature and students in most cases are aware of the importance of strategies and target language exposure in their language proficiency improvement.

10.6 Concluding summary

This chapter has investigated the change in the vocabulary knowledge of a small student population of first-year English majors. We could find overall positive group changes, but varying degrees of negative and positive changes at a subject level. The reported patterns of change were found not to be erratic and could be explained by students' strategies of exposure to the target language, by using explicit vocabulary learning strategies, and by taking university language seminars focusing on vocabulary building. The findings of this study provided a valuable insight into the nature of vocabulary learning of students with time, of which very limited information was available so far. The small-scale study allowed for the identification of group tendencies, but also made possible the arrangement of personal interviews and the explanation of some of the most interesting individual cases.

CHAPTER ELEVEN

OVERALL DISCUSSION OF RESULTS

11.1 Introduction

Based on the multifaceted nature of vocabulary knowledge discussed in Chapter 3, a multidimensional approach was chosen for the investigation of the vocabulary knowledge and use of English majors. While discussing the findings of the various research questions, several issues have come forth that need to be further explored.

11.2 Tests of vocabulary knowledge

The fact of using parallel testing methods proved to have an implication not only for practical assessment, but also for theoretical issues of understanding the complex nature of lexical knowledge. Scores of the VLT and the PVLТ could partially be expected on the basis of previous research and the course work of English majors at the study site. However, the Lex30 test seemed to be the “the odd one out” in many respects, showing weak or non-significant correlation values with the other two tests or with predicted influencing factors. This calls for further consideration of what this test actually measures and what the scores reflect. Indeed, the nonconformity of this test with the Levels Tests is not surprising as they test different types of knowledge (as discussed in Chapters 4 and 7). Vocabulary size, either receptive or productive, is easier to measure and much more straightforward to interpret than associations.

A closer look at the association types may give us an answer why correlation values with certain predicted influencing factors that work for the other two tests, do not work for the Lex30. First, answers to prompt words may be given on the basis of different aspects, such as meaning or form. As has been discussed earlier, the associative answers given to the prompts in the Lex30 do not generate strong first responses, but recall a number of different words (Meara & Fitzpatrick, 2000; Doró, forthcoming). The problem with scoring comes from the fact that some of the responses form compounds or multi-word units with the prompt word

(trade→[trade]mark, board→[to board a] ship). Although some of these examples are reflections of more sophisticated vocabulary knowledge, parts of these multi-word units may be built up of the most frequent words of English, therefore, no scores are awarded to them.

Second, due to the written format, it is not possible to time answers or to assure for the first associations to be recorded. Scores in written association format can be awarded only on the basis of the students' final answers. More associations, each representing a different type of link, can be activated at the same time. While this may seem to be a drawback regarding the scoring of the test, it has a positive effect in terms of test design and methodology. Studies on L2 vocabulary acquisition most often tend to focus on the lexical knowledge learners have stored, but the Lex30 recognizes the fact that the access to this lexical knowledge also plays a determining role in communicative competence. As DeKeyser states, "without automatization no amount of knowledge will ever translate into the levels of skill required for real life use" (2001: 126). When reading a word, more types of links of the same word are activated, and focus is not necessarily on the meaning associations only (for instance, synonyms, antonyms and meronyms). An example of this multiple activation is often seen for the word *board* (*snowboard, board meeting, to board a ship*). Also, as this example shows, multiple meanings of the same words are often activated simultaneously.

11.3 Theoretical issues

Concerning the theoretical questions addressed in this dissertation, one of the most important issues is the receptive–productive dichotomy of vocabulary knowledge. While the Levels Test results are in line with previous studies that target the gap between the two parallel scores, with the use of three tests the simple treatment of this issue is seriously challenged. As has been discussed in Chapter 4, different tests that are labeled as productive do not test the same underlying constructs of knowledge, therefore, it cannot be expected of them to behave the same way. On the basis of the various results it can be concluded that the lexical organization assessed by the Lex30 and its scoring method do not reflect immediate change in the learners' vocabulary size.

Another key theoretical concept discussed in this dissertation is vocabulary knowledge versus use. This proved to be a more challenging question than simply treating it

as a competence versus performance issue. First, because test scores already reflect a type of test taking performance and underlying mental processes, as knowledge is not directly accessible. Second, the assessed knowledge and use may not have a direct relationship with each other, as use requires the ability to make use of the knowledge. Two broader issues of written vocabulary use were directly discussed in this dissertation, namely reading and writing. In both cases vocabulary is one of many factors involved.

The causative links between vocabulary knowledge and reading performance discussed in Chapters 5 and 9 led to some crucial implications. As has been shown, understanding of a text is possible only if a threshold level of words in the texts is met, if reading is practiced and the reader has a good topic familiarity. After evaluating all these aspects on the basis of the data gained, it is seriously questioned whether the majority of the student population under investigation can have adequate understanding of the texts they are reading. Neither the ‘top-down’, nor the ‘bottom-up’ strategy may fully work for them. If reading is not meaningful and is not changed from intensive to extensive reading, then implicit vocabulary gain is expected to be limited. The discussion of the case study has further revealed that only students who reported explicit learning together with extensive reading showed significant change in their vocabulary with time.

Under the broad heading of lexical use, the importance of assessing writing both as a product and as a process was also explored. It could be concluded that vocabulary choice in written performance may not only be due to the transition between knowledge and use, but to the attention vocabulary is dedicated to during the writing process. This implies that if the task does not require sophisticated language use and the writer is not paying enough attention to word choice, even excellent vocabulary knowledge will not be reflected in the lexical profile of essays.

11.4 Assessment implications

Directly following from these theoretical issues, assessment implications of this study could also be drawn. Most importantly, the parallel use of multiple testing methods with a relatively large student population has formed a body of data and allowed for data analysis that can be regarded as a basis for future assessment with a similar student population. The

conclusions drawn here will greatly facilitate the interpretation of future data, even if only limited data sets or number of instruments are used. One of the most important findings on an assessment nature of this research project is the degree of correlation that was revealed among the various tests. It is often the case that in real-life testing situations efficiency and time constraints need to be kept in mind, therefore, a much more restricted number of tests or subtests are administered. Based on the findings of this study we now have great predictive measures in hand to extrapolate results obtained on any part of the above discussed tests into the remaining knowledge types of a learner's lexical proficiency. However, it needs to be stated that the specific research questions were based on questionnaire data the conclusions of which are indicative rather than conclusive and call for the need of further empirical investigation which could focus more on some of these marginally treated aspects of knowledge and use.

Moreover, it needs to be stated, that the instruments employed for data collection can be easily integrated into the everyday instruction practice with university students. The tests can be used as valuable source of feedback to students (both of diagnostic and of predictive nature). Students seemed to accept vocabulary assessment as a separate construct and treated the results and criticism more positively than the feedback provided to them on other, more grammar-oriented language tests. Furthermore, these tests can be used in follow-up assessment, either as part of class work or as study material by the students individually. This is facilitated by the online access to the tests and by the parallel versions of the Levels Tests. Test results should also be taken into consideration in syllabus design, as Levels Tests scores, for instance, can be interpreted in terms of reading difficulty or general academic achievement. As has been pointed out, even those at the end of their third year can be expected to show serious lack of threshold vocabulary for reading. Even if we view this threshold level not as an "all or nothing phenomenon", but as a probabilistic boundary without which inferring of meaning is seriously limited (following Hu & Nation [2001]), more attention should be paid to vocabulary building. As some students may lack the skill or motivation to combine implicit vocabulary learning strategies with explicit ones, help should be provided in the form of separate language classes, and as part of all classes in general. As has been explored in earlier chapters, content knowledge is not gained either if students are unable to read the assigned texts meaningfully. And this sets up a vicious circle which is

harder and harder to get out of, as academic reading assignments and content knowledge requirement increase with time.

Similarly to the tests, the VocabProfile is also a meaningful tool that can be easily integrated into everyday class use. It can check the predicted reading difficulty of texts to be used or assess student essays. This proved to work well in essay writing classes in the previous academic years at the institution where the study took place. Now the large number of texts analyzed should provide an excellent basis for comparison.

With these practical implications in mind, I now turn to the final chapter of review and conclusions.

PART V

REVIEW AND CONCLUSIONS

CHAPTER TWELVE

REVIEW AND CONCLUSIONS

12.1 Introduction

This final chapter has the aim of providing a review of previous chapters, by touching on the major issues and findings of each. This will be followed by drawing some conclusions, discussing limitations, and making suggestions for directions of future research.

12.2 Overview of the dissertation

The dissertation was divided into five main parts, each having a distinct role of general introduction, literature review, research questions and methodology, results and discussion, and review and conclusions. This section aims to summarize the content of the dissertation by following the order of the chapters.

The first chapter had the main role of introducing the issue of written assessment of vocabulary knowledge and use in an English for academic purposes context. After some notes on the general importance of the domain, the challenging task of defining the notion of vocabulary to be used was met. Then the Hungarian context was introduced, following the explanation of the main research areas of the dissertation.

The Literature review chapters started out with a reflection on the impossible nature of giving a detailed account of all areas of the growing body of research concerning L2 vocabulary studies. The main purpose of the literature review was that of exploring major cornerstones in the research directly relating to the empirical investigation.

In Chapter 2 the difficult issue of construct definition was dealt with in detail. It was pointed out that the way *word* and other related terms are interpreted may have a direct bearing on research methods and analysis. The chapter then went on to explore the role of

derivational affixes and multi-word units. Finally, the role of multiple meanings of words was explored.

Chapter 3 started out with the evaluation of a quote which suggested that the blurring of terminology may be a deliberate choice made by some researchers. It was concluded that the ill-defined nature of vocabulary is not universally purposeful, but it surely poses difficulty in interpreting the accumulated body of knowledge in certain areas. The chapter discussed various vocabulary knowledge frameworks and aspects, all relating back to Richard's (1976) model. These frameworks were found to have many overlapping points of which the aspects of receptive and productive, partial and precise knowledge, the accessibility of the lexicon and lexical organization were elaborated on.

Chapter 4 was dedicated to the discussion of the assessment of learners' vocabulary. First, a theoretical background was provided in the form of frameworks of vocabulary assessment, emphasizing the need for a careful selection of instruments. The chapter then followed on with the discussion of test types, paying special attention to the three tests employed for data collection in the present empirical investigation. Then, I presented the lexical measures used for assessing written texts, with detailed explanation of those selected for this study. Some consideration was also given to corpus studies and to the role of word frequency in assessment.

In Chapter 5, the last chapter of the literature review, the first part was dedicated to the discussion of the relationship between vocabulary and reading. Issues concerning the vocabulary gain from reading, the threshold level of words needed for decoding meaning, the inferring of unknown lexical items from text, and the effect of repetition on vocabulary learning were discussed. The second part of the chapter examined models of the role of the lexicon in writing. The difference between seeing writing as a product and as a process was further elaborated on. The remaining parts of the chapter reviewed some empirical studies with direct indication of the methodology and research questions used for the proposed investigation.

Chapter 6 presented the seven main research questions, explaining the rationale behind each of them and discussing specific questions related to all research areas. It highlighted the importance of a multidimensional study, which uses a variety of data collection instruments to capture rich data.

Chapter 7 explained all issues concerning the methodology of the investigation, including subjects, research instruments, procedure, data handling and scoring, and the detailed piloting. It was explained that a major strength of the present research is the fact that learners' lexical knowledge is targeted from various angles and not limited to one or two aspects. Qualitative data analysis, alongside with the statistical one, was proposed.

The first part of Chapter 8 provided the descriptive statistics for the three tests of vocabulary knowledge, the Vocabulary Levels Test, the Productive Vocabulary Levels Test and the Lex30 association task. It was shown that the treatment of the over 300 participants as one overall group allowed only for general conclusions. Therefore, in the second part of this chapter, different groupings of the study population were done on the basis of several factors that were expected to have an influence on the test results. It was explained why multiple factors are more likely to have a significant effect on test scores than individual factors. In the third part of the chapter, correlations were calculated using overall and sub-scores of the three tests. This had both theoretical and assessment related importance.

The first section of Chapter 9 investigated the role of vocabulary knowledge on the academic success of English majors. The advancement in studies was investigated, first, on the basis of predicted reading difficulty, and second, on the basis of failing or passing courses. As for reading, many of the students could be expected to show reading problems of general and academic texts due to the high proportion of unknown vocabulary. This is combined with the limited amount of overall reading done by most of the subjects. As for writing, the influence of vocabulary knowledge and topic choice in the lexical measure of texts was investigated. Students on average reported to pay less attention to vocabulary than to grammar or text organization during writing.

Chapter 10 reported on the results of a follow-up case study involving 15 first-year students tested at the very beginning and at the end of their first academic year. Overall positive group changes were found to hide great individual differences. Both negative and positive values were identified. Seven subjects showing extreme cases in their vocabulary gain were further examined using questionnaire and structured interview data.

Chapter 11 provided an overall discussion of some areas of the study that deserved further treatment, such as the role of the Lex30 test or the theoretical and practical implications of the data.

12.3 Conclusions and implications

The main aim of this dissertation has been to add to the body of work carried out so far on written vocabulary assessment for both research and pedagogical purposes. While keeping in mind theoretical issues, such as the complex nature of vocabulary knowledge, the relationship between lexical knowledge assessed by tests and the vocabulary use seen in the case of reading or writing at front, future assessment purposes and practical issues could not be left out of consideration.

This dissertation was written with both theoretical and practical needs in mind, on the one hand asking for the clarification of research instruments and data interpretation, and, on the other hand, having the need of simple testing methods that are easily available and readily usable in language instruction and syllabus planning. Rather than designing alternative tests, the dissertation wished to employ data collection methods and research tools that had already been partly used in university contexts, both in Hungary and elsewhere, and that could be easily introduced in a wider context either as diagnostic measures of students' vocabulary, as tools to be integrated in instruction, or as help that could be directly used by students during their individual language studies. It was of both theoretical and practical value to gain information on how these instruments can inform us about students' lexical knowledge and use of which, so far, we could only have partial insight or assumptions based on the limited amount of published data and on the everyday observation of instructors.

My teaching experience with students involved in the studies described, their daily needs in terms of class materials that can maximize their language gain, proved to provide valuable insight for this investigation. The implications of the research, however, could be valid in other, similar educational contexts in Hungary or elsewhere in the world. The rich data collected allowed for many major and minor questions to be addressed. Further research would be needed to follow up on the longitudinal study done at an individual level, as it has been shown that treating the subjects one-by-one might provide more meaningful data than group means. The analysis of student essays could be continued both for research and pedagogical purposes, involving also the process part of the writing through the analysis of various drafts or the following up on students' progress in essay writing classes (as has been piloted by myself). Answers based on limited questionnaire data could also be further

explored in order to strengthen the predictive nature of certain influencing factors on vocabulary knowledge or use. Texts analysis could also be done on a multi-word level alongside with the word-level analyses. As a final point to be raised, the associative answers prompted by the Lex30 could be further analyzed to better understand the nature of the responses (as was done in Doró [forthcoming] for some of the stimuli).

A major pedagogical implication of the study is the large individual variability in vocabulary knowledge that could be documented. Many first-year students seem to lack the threshold lexical knowledge needed to read general and academic texts and to function in a target-language medium education. While upper grade groups reached better test scores, these can be partially attributed to the high drop-out rate of weak students early in their studies. Based on the case study we can conclude that only those who do excessive language practice and employ explicit learning strategies show meaningful gain in their vocabulary. To overcome the primary language difficulties, the first 5,000 word families, alongside with basic technical vocabulary of course materials, should be systematically and explicitly learned, as suggested by Schmitt (2007). Moreover, strategies to cope with less frequent words should be learnt, but strategies that are not restricted to the ignorance of all unknown vocabulary, because that is unlikely to lead to any lexical gain over time. Since most students arrive at the university with no useful strategy of independent vocabulary learning, help should be offered to them in the form of extra vocabulary building classes or by directing them to materials available for individual study.

REFERENCES

- Agustín Llach, M. P., Moreno Espinosa, S., & Fernández Fontecha, A. (2005). Responding to different composition topics: A quantitative analysis of lexical error production. *Glosas Didácticas*, 13, 128–140.
- Aitchison, J. (1994). *Words in the mind. An introduction to the mental lexicon*. Oxford: Blackwell.
- Alderson, J. C., & Huhta, A. (2005). The development of a suite of computer-based diagnostic tests based on the Common European Framework. *Language Testing*, 22(3), 301–320.
- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14, 115–129.
- Altenberg, B., & Granger, S. (2001). The grammatical and lexical patterning of MAKE in native and non-native student writing. *Applied Linguistics*, 22(2), 173–195.
- Anthony, E. M. (1975). Lexicon and vocabulary: Some theoretical and pedagogical aspects of word meaning. *RELC Journal*, 6(6), 2–30.
- B. Fejes, K. (1981). *Egy korosztály írásbeli nyelvhasználatának fejlődése*. Budapest: Tankönyvkiadó.
- Bachman, L. F. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language Testing*, 17(1), 1–42.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Barratt, L. B. & Kontra, M. (1996). Matching Hungarian and English color terms. *International Journal of Lexicography*, 9(2), 102–117.
- Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography*, 6(4), 253–279.
- Beglar, D., & Hunt, A. (1999). Revising and validating the 2000 word level and university word level vocabulary tests. *Language Testing*, 16(2), 131–162.
- Benssoussan, M., & Laufer, B. (1984). Lexical guessing in context in EFL reading comprehension. *Journal of Research in Reading*, 7, 15–32.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London: Longman.
- Breland, H., Lee, Y.-W., & Muraki, E. (2005). Comparability of TOEFL CBT essay prompts: Response-mode analyses. *Educational and Psychological Measurement*, 65(4), 577–595.
- Brutten, S. R. (1981). An analysis of student and teacher indications of vocabulary difficulty. *RELC Journal*, 12(1), 66–71.
- Bukta, K. (2007). *Processes and outcomes in L2 English written performance assessment: Raters' decision-making processes and awarded scores in rating Hungarian EFL learners' compositions*. Unpublished PhD dissertation, University of Pécs, Hungary
- Cameron, D. (2002). Measuring vocabulary size in English as an additional language. *Language Teaching Research*, 6, 145–73.

- Campion, M.E., & Elley, W.B. (1971). *An academic vocabulary list*. Wellington: NZCER
- Carlo, M. S., August, D., McLaughlin, B., Snow, C. E., Dressler, C., & Lippman, D. N. (2004). Closing the gap: Addressing the vocabulary needs of English-language learners in bilingual and mainstream classroom. *Reading Research Quarterly*, 39(2), 188–215.
- Chapelle, C.A. (1994). Are C-tests valid measures for L2 vocabulary research? *Second Language Research*, 10, 157–187.
- Chapelle, C.A. (1998). Construct definition and validity inquiry in SLA research. In L. F. Bachman, & A. D. Cohen (Eds.), *Interfaces between second language acquisition and language testing research*. (pp. 32–70). Cambridge: Cambridge University Press.
- Chin, C. H. L., Donn, J. S., & Conry, R. F. (1991). Effects of computer-based tests on the achievement, anxiety, and attitudes of grade 10 science students. *Educational and Psychological Measurement*, 51(5), 735–745.
- Choi, I.-C., Kim, K. S., & Boo, J. (2003). Comparability of a paper-based language test and a computer-based language test. *Language Testing*, 20(3), 295–320.
- Chujo, K. & Utiyama, M. (2005). Understanding the role of text length, sample size and vocabulary size in determining text coverage. *Reading in a Foreign Language*, 17, 1–22.
- Chung, T. M., & Nation, P. (2003). Technical vocabulary in specialised texts. *Reading in a Foreign Language*, 15, 103–116.
- Coady, J. & Nation, I.S.P. (1988). Vocabulary and reading In R. Carter & M. McCarthy (Eds.), *Vocabulary and language teaching* (pp. 97–110). London: Longman.
- Cobb, T. (2008a). Complete Lexical Tutor [computer software]. Available at <http://www.lextutor.ca>. Access January 30, 2008.
- Cobb, T. (2008b). Web VocabProfile [computer software]. Available at <http://www.lextutor.ca/vp>. Access January 30, 2008.
- Council of Europe (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Coxhead, A. (1998). *An academic word list*. LALS Occasional Publication. Wellington: Victoria University of Wellington.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213–238.
- Csapó, B. (Ed.) (2002). *Az iskolai műveltség*. Budapest: Osiris.
- Csernoch, M. & Korponayné, N. I. (2005). A comparative analysis of the word types of the New Headway Coursebooks. *Publicationes Universitatis Miskolciensis*, 13–23.
- Csernoch, M., & Korponayné, N. I. (2007). Az új kétszintű érettségi: Európai kitekintés. In P. Heltai (Ed.), *Nyelvi modernizáció. Szaknyelv, fordítás, terminológia. A XVI. Magyar Alkalmazott Nyelvészeti Kongresszus előadásai Gödöllő, 2005. április 7–9*. Vol. 2/1. (pp. 267–276). Pécs – Gödöllő: MANYE – Szent István Egyetem.
- Day, R. R., & Bamford, J. (1998). *Extensive reading in the second language classroom*. Cambridge: Cambridge University Press.
- DeKeyser, R. M. (2001). Automaticity and automatization In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 125–151). Cambridge: Cambridge University Press.

- Dóczi, B. (2007). Az idegen nyelvi mentális lexikon fejlődése. Paper presented at the State of English in Hungary Conference, Budapest, November 16, 2007.
- Doró, K. (2001). *“The mystery of the black cloud”*: The development of American children’s written narratives. Unpublished MA thesis, University of Szeged, Hungary
- Doró, K. (2006). Angolul tanuló középiskolások iskolán kívüli nyelvhasználata. In K. Klaudy, & Cs. Dobos (Eds.), *A világ nyelvei és a nyelvek világa. Soknyelvűség a gazdaságban, a tudományban és az oktatásban. A XV. Magyar Alkalmazott Nyelvészeti Kongresszus előadásai Miskolc, 2006. április 10–12.* Vol. 3/1 (pp. 197–201). Pécs – Miskolc: MANYE – Miskolci Egyetem.
- Doró, K. (2007a). Aktív és passzív szókincs a középiskolai idegennyelv-tanulásban és tanításban. In P. Heltai (Ed.), *Nyelvi modernizáció. Szaknyelv, fordítás, terminológia. A XVI. Magyar Alkalmazott Nyelvészeti Kongresszus előadásai Gödöllő, 2005. április 7–9.* Vol. 2/1 (pp. 277–281). Pécs – Gödöllő: MANYE – Szent István Egyetem.
- Doró, K. (2007b). The use of high- and low-frequency verbs in English native and non-native student writing. In Zs. Lengyel, & J. Navracsics (Eds.), *Second language lexical processes: Applied linguistic and psycholinguistic perspectives* (pp. 117–129). Clevedon: Multilingual Matters.
- Doró, K. (2007c). Kerekasztal-beszélgetés a kollokációkról. *Modern Nyelvoktatás, XIII/2–3*, 133–143.
- Doró, K. (2007d). Választékos idegen nyelvi szókincs: Előny vagy feltétel? In T. Váradi (Ed.), *I. Alkalmazott Nyelvészeti Doktorandusz Konferencia* (pp. 24–33). Budapest, MTA Nyelvtudományi Intézet. Available at <http://www.nytud.hu/alknyelvdok/proceedings07/Doro.pdf>. Access January 30, 2008.
- Doró, K. (2007e). Produktív szókincs és szövegalkotás. Paper presented at the 17th MANYE Congress, Siófok, Hungary, April 20, 2007.
- Doró, K. (forthcoming) Productive vocabulary and mental lexicon in light of a word association task. In Zs. Lengyel, & J. Navracsics (Eds.), *Proceedings of the 10th Summer School of Psycholinguistics*.
- Dörnyei, Z. (2003). *Questionnaires in second language research: Construction, administration and processing*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Dörnyei, Z., & Katona, L. (1992) Validation of the C-test amongst Hungarian EFL learners. *Language Testing, 9*, 187–206.
- Dörnyei, Z., Durow, V., & Zahran, K. (2004). Individual differences and their effects on formulaic sequence acquisition. In N. Schmitt (Ed.), *Formulaic Sequences. Acquisition, processing and use* (pp.87–106). Amsterdam: John Benjamins.
- Dycus, D. (1997). Guessing word meaning from context: Should we encourage it? *Literacy Across Cultures, 1*(2). Available at <http://www2.aasa.ac.jp/~dcdycus/LAC97/guessing.htm>. Access January 30, 2008.
- Eckes, T., & Grotjahn, R. (2006). A closer look at the construct validity of C-tests. *Language Testing, 23*, 290–325.
- Ellis, N. C. (1994). Introduction: Implicit and explicit learning – an overview. In N.C. Ellis (Ed.) *Implicit and explicit learning of languages* (pp. 1–31). London: Academic Press.
- Fitzpatrick, T., & Meara, P. (2004). Exploring the validity of a test of productive vocabulary. *Vigo International Journal of Applied Linguistics, 1*, 55–74.

- Flower, L., & Hayes, J. (1981). The cognition of discovery: Defining a rhetorical problem. *College Composition and Communication*, 31, 21–32.
- Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage*. Boston: Houghton Mifflin Company.
- Fujieda, Y. (2006). A brief historical sketch of second language writing studies: A retrospective. Available at <http://www.kyoai.ac.jp/college/ronshuu/no-06/fujieda.pdf>. Access December 14, 2007.
- Gardner, D. (2007). Validating the construct of word in applied corpus-based vocabulary research: A critical survey. *Applied Linguistics*, 28(2), 241–265.
- Goldberg, A. L., & Pedulla, J. J. (2002). Performance differences according to test mode and computer familiarity on a practice graduate record exam. *Educational and Psychological Measurement*, 62(6), 1053–1067.
- Goodfellow, R., Jones, G., & Lamy, M.-N. (2002). Assessing learners' vocabulary using the lexical frequency. *ReCALL*, 14(1), 133–145.
- Goodfellow, R., & Lamy, M.-N. (1998). Learning to learn a language – at home and on the Web. *ReCALL*, 10(1), 68–78.
- Grabe, W., & Kaplan, R.B. (1996). *Theory and practice of writing*. London and New York: Longman.
- Granger, S. (1996). From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg & M. Johansson (Eds.), *Languages in Contrast. Text-based cross-linguistic studies. Lund Studies in English 88*. (pp. 37–51). Lund: Lund University Press.
- Granger, S. (2003). Error-tagged learner corpora and CALL: A promising synergy. *CALICO Journal*, 20(3), 465–480.
- Granger, S. (2004). Computer learner corpus research: Current status and future prospects. In U. Connor, & T. Upton (Eds.), *Applied corpus linguistics: A multidimensional perspective* (pp. 123–145). Amsterdam & Atlanta: Rodopi.
- Granger, S. (2005). Pushing back the limits of phraseology. How far can we go? In C. Cosme, C. Gouverneur, F. Meunier, & M. Paquot (Eds.), *Proceedings of the Phraseology 2005 Conference, Louvain-la-Neuve, 13–15 October 2005* (pp. 165–168). Louvain-la-Neuve: Centre for English Corpus Linguistics.
- Granger, S. (Ed.). (1998). *Learner English on computer*. New York: Addison Wesley Longman Inc.
- Granger, S., & Wynne M. (2000). Optimizing measures of lexical variation in EFL learner corpora In J.M. Kirk (Ed.), *Corpora galore* (pp. 249–258). Amsterdam: Rodopi.
- Haastrup, K. (1990). Inferring from context. Paper presented at the 9th World Congress of Applied Linguistics, Thessaloniki, Greece.
- Haastrup, K., & Henriksen, B. (1998). Vocabulary acquisition: From partial to precise understanding. In Haastrup, K., & Viberg, Å. (Eds.), *Perspectives on lexical acquisition in a second language* (pp. 97–124). Lund: Lund University Press.
- Hazenbarg, S., & Hulstijn, J. H. (1996). Defining the minimal receptive second-language vocabulary for non-native university students: An empirical investigation. *Applied Linguistics*, 17/2, 145–163.

- Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition*, 21(2), 303–317.
- Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language*, 8(2), 689–696.
- Honeyfield, J. (1977). Word frequency and the importance of context in vocabulary learning. *RELC Journal*, 8(1), 35–42.
- Horowitz, D. (1986). Process, not product: Less than meets the eye. *TESOL Quarterly*, 20, 141–144.
- Horst, M., Cobb, T., & Meara, P. (1998). Beyond A Clockwork Orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, 11, 207–223.
- Horváth, J. (2000). A JPU korpusz kialakítása és tartalma. *Modern Nyelvoktatás*, VI(2), 42–49.
- Horváth, J. (2001). *Advanced writing in English as a foreign language: A corpus-based study of processes and products*. Pécs: Lingua Franca.
- Horváth, J. (2007). Modern nyelvoktatás a JPU korpusz internetes gyűjteményével. *Modern Nyelvoktatás*, XIII(4), 19–25.
- Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics*, 19(1), 24–44.
- Hu, H. M., & Nation, I.S.P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a foreign language*, 13(1), 403–430.
- Hulstijn, J. H. (1992). Retention of inferred and given word meaning. Experiments in incidental vocabulary learning. In P. Arnaud, & H. Bejoint (Eds.), *Vocabulary and applied linguistics* (pp. 113–25). London: McMillan.
- Hulstijn, J. H. (2001). Implicit and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity In P. Robinson (Ed.). *Cognition and second language instruction* (pp. 258–287). Cambridge: Cambridge University Press.
- Hunt, A. & Beglar, D. (2005). A framework for developing EFL reading vocabulary. *Reading in a Foreign Language*, 17(1), 23–59.
- Ittzész, K.K. (1989). A magyar nyelvi tudás mérése Amerikában integratív nyelvi tesztekkel. *Magyar Nyelvőr*, 2, 228–237.
- Jackson, H. (2007). Does frequency matter? Paper presented at the 10th Summer School of Psycholinguistics, Balatonalmádi, Hungary, June 5, 2007.
- Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, 21(1), 47–77.
- Johnson, K., & Johnson, H. (Eds.) (1999). *Encyclopedic dictionary of applied linguistics*. Oxford: Blackwell Publishing Ltd, 341–348.
- Kádárné, Fülöp J. (1979). Az angol nyelv tanításának eredményei In Á. Kiss, S. Nagy, & J. Szarka (Eds.), *Tanulmányok a neveléstudomány köréből 1975–1976* (pp. 276–341). Budapest: Akadémiai Kiadó.
- Kiszely, Z. (2006). Magyar középiskolások angol fogalmazási stratégiái. *Nyelvinfó* 14/2, 8–12.

- Kiszely, Z. (2007). Középiskolások magyar és angol nyelvi írásfolyamatai. In P. Heltai (Ed.), *Nyelvi modernizáció. Szaknyelv, fordítás, terminológia. A XVI. Magyar Alkalmazott Nyelvészeti Kongresszus előadásai Gödöllő, 2005. április 7–9.* Vol. 2/1 (pp. 282–287). Pécs – Gödöllő: MANYE – Szent István Egyetem.
- Kormos, J., Kontra E. H., & Csölle, A. (2002). Language wants of English majors in a non-native context. *System*, 30, 517–542.
- Kučera, H., & Francis, W.N. (1967). *A computational analysis of presentday American English*. Providence, RI: Brown University Press.
- Laufer, B. (1992). Reading in a foreign language: How does L2 lexical knowledge interact with the reader's general academic ability? *Journal of Research in Reading*, 15, 95–103.
- Laufer, B. (1997a). The lexical plight in second language reading: Words you don't know, words you think you know and words you can't guess. In J. Coady, & T. Huckin (1997). *Second language vocabulary acquisition* (pp. 20–34). Cambridge: Cambridge University Press.
- Laufer, B. (1997b). What's in a word that makes it hard or easy: Some intralexical factors that affect the learning of words. In N. Schmitt, & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 140–155). Cambridge: Cambridge University Press.
- Laufer, B. (1998). The development of passive and active vocabulary in a second language. Same or different? *Applied Linguistics*, 19(2), 255–271.
- Laufer, B. (2005). Lexical frequency profile. From Monte Carlo to the real world. A response to Meara (2005). *Applied Linguistics*, 26(4), 582–588.
- Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, 54(3), 399–436.
- Laufer, B., & Nation, P. (1995). Lexical richness in L2 written production: Can it be measured? *Applied Linguistics*, 16(3), 307–312.
- Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, 16(1), 33–51.
- Laufer, B., & Paribakht, T. S. (1998) The relationship between active and passive vocabularies: Effects of language learning context. *Language Learning*, 48/3, 365–391.
- Laufer, B., & Yano, Y. (2001). Understanding unfamiliar words in a text: Do L2 learners understand how much they don't understand? *Reading in a Foreign Language*, 13(2), 549–566.
- Lehmann, M. (2003). The lexis of writing and vocabulary size: The relationship between receptive knowledge and productive use. In J. Andor, J. Horváth, & M. Nikolov (Eds.), *Studies in English theoretical and applied linguistics* (pp. 172–181). Pécs: Lingua Franca.
- Lehmann, M. (2006). Vocabulary as a filter with first-year English majors. In M. Nikolov, & J. Horváth (Eds.), *UPRT 2006: Empirical studies in English applied linguistics* (pp. 121–137). Pécs: Lingua Franca Csoport. Available at <http://www.pte.hu/uprt/2.2%20Lehmann.pdf>. Access January 30, 2008.
- Lehmann, M. (2007). The size of an English receptive vocabulary. An exploratory study of the EFL vocabulary size of English majors at the University of Pécs. In R. Sheorey, & J.

- Kiss-Gulyás (Eds.), *Studies in applied and theoretical linguistics* (pp. 25–38). Debrecen: Kossuth Egyetemi Kiadó.
- Little, D., & Singleton, D. (1992) The C-test as an elicitation instrument in second language research. In R. Grotjahn (Ed.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen. Vol. I.* (pp. 173–190). Bochum: Brockmayer.
- Lumley, T. (2002). Assessment criteria in a large-scale writing test: What do they really mean to the rater? *Language testing*, 19(3), 246–276.
- Lynn, R.W. (1973). Preparing word lists: A suggested method. *RELC Journal*, 4, 25–32.
- Maeda, J. & Hobara, Y. (1999). Frequency level checker [computer software]. Available at <http://language.tiu.ac.jp/flc/index.html>. Access January 30, 2008.
- McCrostie, J. (2007). Investigating the accuracy of teachers' word frequency intuitions. *RELC Journal*, 38(1), 53–66
- Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjaev, & J. Williams (Eds.), *Performance and competence in second language acquisition* (pp. 35–53). Cambridge: Cambridge University Press.
- Meara, P. (2005). Lexical frequency profile: A Monte Carlo analysis. *Applied Linguistics* 26(1), 32–47.
- Meara, P., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing*, 4, 142–45.
- Meara, P., & Fitzpatrick, T. (2000). Lex30: An improved method of assessing productive vocabulary in an L2. *System*, 28, 9–30..
- Meara, P., & Jones, G. (1988). Vocabulary size as a placement indicator. In P. Grunwell, (Ed.), *Applied linguistics in society* (pp. 80–87). London: CILT.
- Meara, P., & Jones, G. (1990). *The Eurocentres vocabulary size test 10KA*. Zurich: Eurocentres.
- Meara, P.M., & Milton, J. L. (2003). *X_Lex: The Swansea vocabulary levels test*. Newbury: Express.
- Melka, F. (1997). Receptive vs. productive aspects of vocabulary. In N. Schmitt, & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 84–102). Cambridge: Cambridge University Press.
- Messick, S. (1989). Validity. In Linn, R.L. (Ed.), *Educational measurement* (pp. 13–103). New York: Macmillan.
- Mochida, A., & Harrington, M. (2006). The Yes/No test as a measure of receptive vocabulary knowledge. *Language Testing*, 23(1), 73–98.
- Moon, R. (1997). Vocabulary connections: Multi-word items in English. In N. Schmitt, & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 40–63). Cambridge: Cambridge University Press.
- Morgan, B. Q., & Oberdeck, L. M. (1930). Active and passive vocabulary In E. W., Bagster-Collings (Ed.), *Studies in Modern Language Teaching* (pp. 213–221). New York.
- Morris, L., & Cobb, T. (2003). Vocabulary profiles as predictions of academic performance of Teaching English as a Foreign Language trainees. *System*, 32(1), 75–87.

- Muncie, J. (2002). Process writing and vocabulary development: Comparing Lexical Frequency Profiles across drafts. *System*, 30(2), 225–235.
- Na, L., & Nation, I.S.P. (1985). Factors affecting guessing vocabulary in context. *RELC Journal*, 16(1), 33–42.
- Nagy, W. (1997). On the role of context in first and second language learning. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 64–83). Cambridge: Cambridge University Press.
- Nagy, W. E., Diakidoy, I. N., & Anderson, R. C. (1993). The acquisition of morphology: Learning the contribution of suffixes to the meanings of derivatives. *Journal of Reading Behaviour*, 25(2), 155–80.
- Nassaji, H. (2004). The relationship between depth of vocabulary knowledge and L2 learners' lexical inferencing strategy use and success. *The Canadian Modern Language Review*, 61(1), 107–134.
- Nation, I.S.P. & Waring, R. (1997). Vocabulary size, text coverage and word lists. In N. Schmitt, & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy*. (pp. 6–19). Cambridge: Cambridge University Press.
- Nation, I.S.P. (1983). Testing and teaching vocabulary. *Guidelines*, 5, 12–25.
- Nation, I.S.P. (1990). *Teaching and learning vocabulary*. New York: Newbury House.
- Nation, I.S.P. (1993). Vocabulary size, growth, and use In R. Schreuder, & B. Weltens (Eds.), *The bilingual lexicon* (pp. 115–134). Amsterdam and Philadelphia: John Benjamins.
- Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nesslerhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24(2), 223–242.
- Nikolov, M., & Józsa, K. (2003). *Idegen nyelvi készségek fejlettsége angol és német nyelvből a 6. és 10. évfolyamon a 2002/2003-as tanévben. Szakmai beszámoló országos adatok alapján*. Budapest: OKÉV.
- Nikolov, M., & Józsa, K. (2006). Relationship between language achievements in English and German and classroom-related variables. In M. Nikolov, & J. Horváth (Eds.), *UPRT 2006: Empirical studies in English applied linguistics* (pp. 179–207). Pécs: Lingua Franca Csoport.
- Paribakht, T. S., & Wesche, B. M. (1993). Reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal*, 11(1), 9–29.
- Peckham, D. (2007). Self regulation of vocabulary learning at the university level. Paper presented at the 17th MANYE Congress, Siófok, Hungary, April 21, 2007.
- Peckman, D. (2006). The effect of vocabulary size on reading comprehension and subsequent memory for unknown words. Paper presented at the 16th MANYE Congress, Gödöllő, Hungary, April 11, 2006.
- Pitts, M., White, H., & Krashen, S. (1989). Acquiring second language vocabulary through reading: A replication of the Clockwork Orange study using second language acquirers. *Reading in a Foreign Language*, 5(2), 271–275.

- Polio, C. (2001). Review of test pilot. *Language Learning & Technology*, 5(2), 34–37.
- Praninskas, J. (1972). *Academic university word list*. London: Longman.
- Qian, D. D., & Schedl, M. (2004). Evaluation of an in-depth vocabulary knowledge measure for assessing reading performance. *Language Testing*, 21(1), 28–52.
- Qian, D. D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review*, 56(2), 283–307.
- Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning*, 52, 513–536
- Read, J. (1988). Measuring the vocabulary knowledge of second language learners. *RELC Journal*, 19(2), 12–25.
- Read, J. (1993). The development of a new measure for L2 vocabulary knowledge. *Language Testing*, 10(4), 355–371.
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Read, J., & Chapelle, C. A. (2001). A framework for second language vocabulary assessment. *Language Testing*, 18(1), 1–32.
- Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quarterly*, 10(1), 77–89.
- Rieder, A. (2002). A cognitive view of incidental vocabulary acquisition: From text meaning to word meaning? *VIEWS, Vienna English Working Papers*, 11(1–2), 53–71.
- Rieder, A. (2003). Implicit and explicit learning in incidental vocabulary acquisition. *VIEWS, Vienna English Working Papers*, 12(2), 24–39.
- Sag, I. A., Baldwin, T., Bond, F., Copestake, A., & Flickinger, D. (2002) Multiword expressions: A pain in the neck for NLP In A. F. Gelbukh (Ed.), *Proceedings of the third international conference on intelligent text processing and computational linguistics (CICLING 2002)* (pp. 1–15), Mexico City, Mexico. Available at <http://lingo.stanford.edu/pubs/WP-2001-03.pdf>. Access January 30, 2008.
- Saragi, T., Nation, P., & Meister, G. F. (1978). Vocabulary learning and reading. *System*, 6, 72–78.
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- Schmitt, N. (2007). Current perspectives on vocabulary teaching and learning. In J. Cummings, & C. Davison (Eds.), *International handbook of English language teaching. Vol II.* (pp. 827–841). Norwell, MA: Springer.
- Schmitt, N., Dörnyei, Z., Adolphs, S., & Durow, V. (2004). Knowledge and acquisition of formulaic sequences In N. Schmitt (Ed.), *Formulaic Sequences. Acquisition, processing and use* (pp.55–86). Amsterdam: John Benjamins.
- Schmitt, N., & Dunham, B. (1999). Exploring native and non-native intuitions of word frequency. *Second Language Research*, 15(4), 389–411.
- Schmitt, N., & McCarthy, M. (Eds.). (1997). *Vocabulary: Description, acquisition and pedagogy*. Cambridge: Cambridge University Press.

- Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework: Word association and verbal suffixes. *Studies in Second Language Acquisition*, 19, 17–36.
- Schmitt, N., & Zimmerman, C. B. (2002). Derivative word forms: What do learners know? *TESOL Quarterly* 36(2), 145–171.
- Schmitt, N., Schmitt, D., & Claphan, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18(1), 55–88.
- Shaw, P., & Liu, E. T-K. (1998). What develops in the development of second-language writing? *Applied Linguistics*, 19(2), 225–254.
- Shiotsu, T., & Weir, C. J. (2007). The relative significance of syntactic knowledge and vocabulary breadth in the prediction of reading comprehension test performance. *Language Testing*, 24(1), 99–128.
- Silva, T., & Matsuda, P. K. (Eds.). (2001). *On second language writing*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Sinclair, J. (Ed.) (1995). *Collins COBUILD English Dictionary*. London: Collins.
- Singleton, D. (1999). *Exploring the second language mental lexicon*. Cambridge: Cambridge University Press.
- Singleton, D., & Singleton, E. (1998). The C-test and L2 acquisition/processing research. In J. A. Coleman (Ed.), *University language testing and the C-test* (pp. 150–178). Portsmouth: University of Portsmouth.
- Smith, F. P., & Campbell, H. (1942). Objective achievement testing in French: Recognition versus recall tests. *Modern Language Journal*, 26, 192–198.
- Smith, K. (2004). School to university: An investigation into the experience of first-year students of English at British universities. *Arts and Humanities in Higher Education*, 3(1), 81–93.
- Stalnaker, J. M., & W. Kurath. (1935). A comparison of two types of foreign language vocabulary test. *Journal of Educational Psychology*, 26, 435–442.
- Summers, D. (Eds.). (1995). *Longman dictionary of contemporary English*. London: Longman.
- Sutarsyah, C., Nation, I.S.P., & Kennedy, G. (1994). How useful is EAP vocabulary for ESP? A corpus based case study. *RELC Journal*, 25(2), 34–50.
- Szirmai, M. (2001). *The theory and practice of corpus linguistics*. Debrecen: Debreceni Egyetem Kossuth Egyetemi Kiadó.
- Szirmai, M. (2005). *Bevezetés a korpusznyelvészethez*. Budapest: Tinta Kiadó.
- Tadamitsu, K. (2001). An examination of Nation's (1990) vocabulary levels test. Available at <http://www1.harinet.ne.jp/~waring/vocab/colloquium/tad2004.htm>. Access January 30, 2008.
- Thorndike, E. L. & Lorge, I. (1944). *The teacher's word book of 30,000 words*. New York, NY: Teachers College, Columbia University.
- Tsang, W.-K. (1996). Comparing the effects of reading and writing on writing performance. *Applied Linguistics* 17(2), 210–223.

- Tschirner, E. (2004). Breadth of vocabulary and advanced English study: An empirical investigation. *Electronic Journal of Foreign Language Teaching*, 1, 27–39.
- Váradi, T. (1980a). A guide to Hungarian–English errors. In L. Dezső, & W. Nemser (Eds.), *Studies in English and Hungarian contrastive linguistics* (pp. 513–590). Budapest: Akadémiai Kiadó.
- Váradi, T. (1980b). Strategies of target language learner communication: Message adjustment. *IRAL*, 18, 59–71.
- Wall, D. (1997). Impact and washback in language testing. In C. Clapham & D. Corson (Eds.), *Encyclopedia of language and education. Vol. 7: Language testing and assessment*. (pp. 291–302). Dordrecht: Kluwer.
- Wang, S., Jiao, H., Young, M. J., Brooks, T., & Olson, J. (2007). Comparability of computer-based and paper-and-pencil testing in K–12 reading assessments: A meta-analysis of testing mode effects. *Educational and Psychological Measurement*, 68, 5–24.
- Waring, R. (1997a). A comparison of the receptive and productive vocabulary sizes of some second language learners. *Immaculata: The occasional paper at Notre Dame Seishin University*. Available at <http://www1.harenet.ne.jp/~waring/papers/vocsize.html>. Access January 30, 2008.
- Waring, R. (1997b). *Tasks for assessing second language receptive and productive vocabulary*. Unpublished doctoral dissertation. University of Wales, United Kingdom.
- Waring, R. (2000). The ‘State rating task’ – An alternative method of assessing receptive and productive vocabulary. *Kiyo, Notre Dame Seishin University: Studies in Foreign Language and Literature*, 24(1), 125–15. Available at <http://www1.harenet.ne.jp/~waring/papers/kiyo2000.html>. Access January 30, 2008.
- Waring, R. (2002). Scales of vocabulary knowledge in second language vocabulary assessment. *Kiyo, The occasional papers of Notre Dame Seishin University*. Available at <http://www1.harenet.ne.jp/~waring/papers/scales.html>. Access January 30, 2008.
- Waring, R., & Nation, P. (2004). Second language reading and incidental vocabulary learning. *Angles on the English-speaking world*, 4, 11–12.
- Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, 15, 1–27.
- Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, 28(1), 46–65.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.
- Wesche, M. B., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review*, 53, 13–40.
- West, M. (1953). *A general service list of English words*. London: Longman, Green & Co.
- Witalisz, E. (2004). *An analysis of the written English produced by Polish EFL learners applying to the English Department of the Jagiellonian University*. Unpublished doctoral dissertation. Jagiellonian University, Krakow, Poland.
- Wolter, B. (2001). Comparing the L1 and L2 mental lexicon. A depth of individual word knowledge model. *Studies in Second Language Acquisition*, 23, 41–69.
- Wray, A. (2000). Formulaic sequences in second language teaching: Principles and practice. *Applied Linguistics*, 21(4), 463–89.

- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
- Xing, P., & Fulcher, G. (2007). Reliability assessment for two versions of Vocabulary Levels Tests. *System*, 35(2), 182–191.
- Xue, G., & Nation, I.S.P. (1984). A university word list. *Language Learning and Communication*, 3, 215–229.
- Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *Canadian Modern Language Review*, 57(4), 541–572.
- Zareva, A. (2005). Models of lexical knowledge assessment of second language learners of English at higher levels of language proficiency. *System*, 33(4), 547–562.
- Zimmerman, K. J. (2004). *The role of vocabulary size in assessing second language proficiency*. Unpublished MA thesis, Brigham Young University, USA. Available at <http://contentdm.lib.byu.edu/ETD/image/etd518.pdf>. Access January 30, 2008.

LIST OF TABLES AND FIGURES

Tables

Table 2.1 Difficulty of inflectional and derivational suffixes summarized by Bauer and Nation (1993)	17
Table 3.1 Richards' vocabulary knowledge framework (1976).....	23
Table 3.2 Nation's word knowledge model (2001: 27)	29
Table 4.1 Read's dimensions of vocabulary assessment (2000: 9).....	34
Table 4.2 A framework for vocabulary testing proposed by Read and Chapelle (2001: 10).....	36
Table 4.3 Lexical Proficiency Profile for a learner text calculated by version 2.5 of the VocabProfile	54
Table 5.1 Change in the number of unknown words with increase in text coverage	65
Table 7.1 Major groups of participants	93
Table 7.2 Comprehensive summary of the testing instruments: Vocabulary Levels Test (VLT), Productive Vocabulary Levels Test (PVL) and Lex30.....	97
Table 8.1 Vocabulary Levels Test scores for all subjects	113
Table 8.2 Productive Vocabulary Levels Test scores for all subjects.....	114
Table 8.3 Lex30 test scores for all subjects	114
Table 8.4 Vocabulary Levels Test results for the three main subject groups	119
Table 8.5 Productive Vocabulary Levels Test results for the three main subject groups	120
Table 8.6 Lex30 results for the three main subject groups	121
Table 8.7 Mean values for influencing factors for the three main subject groups	122
Table 8.8 Test of vocabulary knowledge results for the two subgroups of Group 3.....	123
Table 8.9 Mean values for influencing factors for Groups 3a and 3b.....	124
Table 8.10 The influencing factor of years of English studies.....	125
Table 8.11 The influencing factor of time spent in an English-speaking country	127
Table 8.12 The influencing factor of time spent with English on weekdays	128
Table 8.13 The influencing factor of time spent with English at weekends	129
Table 8.14 The influencing factor of all readings done in English	130
Table 8.15 The influencing factor of reading done in English in connection with university studies.....	131
Table 8.16 Statistically significant combinations of influencing factors on vocabulary test results	132
Table 8.17 Correlation matrix for the overall scores on the three tests for the entire study population	137
Table 8.18 Correlation matrix for all frequency levels on the Vocabulary Levels Test	137
Table 8.19 Correlation matrix for all frequency levels on the Productive Vocabulary Levels Test.....	138
Table 8.20 Correlation matrix for parallel frequency levels of the two Levels Tests	138

Table 8.21 Correlation matrix for the overall scores on the three tests for lexically more and lexically less proficient students.....	139
Table 9.1 Percentage of students with predicted reading difficulty of general English texts	148
Table 9.2 Percentage of students with predicted reading difficulty of academic texts	149
Table 9.3 Reported reading difficulty of academic related texts (Question 5/C of the questionnaire).....	150
Table 9.4 Predictive nature of academic success of vocabulary test results of incoming first-year student.....	153
Table 9.5 Differences in mean test scores between academically successful and not successful students at the end of their first year.....	154
Table 9.6 Range of lexical measures of essays written by first and third-year students.....	159
Table 9.7 Correlations between productive vocabulary and lexical measures of essays produced by first-year students (n=73).....	160
Table 9.8 Correlation matrix for the three lexical measures.....	161
Table 9.9 Correlation between productive vocabulary and lexical measures of essays produced by third-year students (n=89).....	161
Table 9.10 Importance of four factors in essay writing	167
Table 9.11 Lexical measures of essays written by Group 2 compared to topic choice	169
Table 9.12 Lexical measures of essays written by Group 3a compared to topic choice	170
Table 9.13 Lexical measures of essays written by Group 3b compared to topic choice	171
Table 9.14 Relationship between vocabulary knowledge and topic choice for Group 2	172
Table 9.15 Relationship between vocabulary knowledge and topic choice for Group 3a	173
Table 9.16 Relationship between vocabulary knowledge and topic choice for Group 3b	173
Table 10.1 Overall group changes in the Vocabulary Levels Test results over one academic year	180
Table 10.2 Overall group changes in the Productive Vocabulary Levels Test results over one academic year.....	181
Table 10.3 Overall group changes in the Lex30 test results over one academic year.....	181
Table 10.4 Gain in scores on the three vocabulary tests	182
Table 10.5 Correlations for influencing factors on the changes in vocabulary test results	183
Table 10.6 Changes in the two Levels Test scores of 7 subjects	184

Figures

Figure 4.1 Sample test item from the Vocabulary Levels Test.....	41
Figure 4.2 Sample test items from the Productive Vocabulary Levels Test.....	43
Figure 4.3 Test format of the Vocabulary Knowledge Scale test (Wesche & Paribakht, 1996)	49
Figure 4.4 Test format of the Self Rating Task (Waring, 1997b).....	50
Figure 5.1 The component of writing (adopted from Johnson & Johnson, 1999: 344).....	74
Figure 5.2 The Flower and Hayes writing process model (Flower & Hayes, 1981: 370)	75
Figure 9.1 Percentage of students with predicted reading difficulty in the three study groups	150
Figure 10.1 Changes in the three test scores over one academic year by subjects.....	184

APPENDIX A: VOCABULARY LEVELS TEST

Válaszd ki a megadott szavak közül azokat, amelyek a jobb oldalon feltüntetett magyarázatoknak a megfelelői. Írd a szavak számát a jelentések előtti vonalra. A hat közül három szó nem illik oda, azoknak a jelentését nem kell megadnod.

Példa:

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

<p>1-10</p> <ol style="list-style-type: none"> 1. birth 2. dust ___ game 3. operation ___ winning 4. row ___ being born 5. sport 6. victory <ol style="list-style-type: none"> 1. belt 2. climate ___ idea 3. executive ___ inner surface of your hand 4. notion ___ strip of leather worn around 5. palm ___ the waist 6. victim <ol style="list-style-type: none"> 1. balloon 2. federation ___ bucket 3. novelty ___ unusual interesting thing 4. pail ___ rubber bag that is filled with air 5. veteran 6. ward <ol style="list-style-type: none"> 1. benefit 2. labor ___ work 3. percent ___ part of 100 4. principle ___ general idea used to guide 5. source ___ one's actions 6. survey <ol style="list-style-type: none"> 1. antics 2. batch ___ foolish behavior 3. connoisseur ___ a group of things 4. foreboding ___ person with a good knowledge 5. haunch ___ of art or music 6. scaffold 	<ol style="list-style-type: none"> 1. choice 2. crop ___ heat 3. flesh ___ meat 4. salary ___ money paid regularly 5. secret ___ for doing a job 6. temperature <ol style="list-style-type: none"> 1. acid 2. bishop ___ cold feeling 3. chill ___ farm animal 4. ox ___ organization of framework 5. ridge 6. structure <ol style="list-style-type: none"> 1. alcohol 2. apron ___ stage of development 3. hip ___ state of untidiness or dirtiness 4. lure ___ cloth worn in front to protect 5. mess ___ your clothes 6. phase <ol style="list-style-type: none"> 1. element 2. fund ___ money for a special purpose 3. layer ___ skilled way of doing something 4. philosophy ___ study of the meaning of life 5. proportion 6. technique <ol style="list-style-type: none"> 1. auspices 2. dregs ___ confused mixture 3. hostage ___ natural liquid present in 4. jumble ___ the mouth 5. saliva ___ worst and most useless parts 6. truce ___ of anything
---	---

11-20

1. cap
2. education _____ teaching and learning
3. journey _____ numbers to measure with
4. parent _____ going to a far place
5. scale
6. trick

1. bench
2. charity _____ long seat
3. jar _____ help to the poor
4. mate _____ part of a country
5. mirror
6. province

1. apparatus
2. compliment _____ expression of admiration
3. ledge _____ set of instruments or machinery
4. revenue _____ money received by
5. scrap
6. tile

1. consent
2. enforcement _____ total
3. investigation _____ agreement or permission
4. parameter _____ trying to find information
5. sum _____ about something
6. trend

1. casualty
2. flurry _____ someone killed or injured
3. froth _____ being away from other people
4. revelry _____ noisy and happy celebration
5. rut
6. seclusion

21-30

1. cream
2. factory _____ part of milk
3. nail _____ a lot of money
4. pupil _____ person who is studying
5. sacrifice
6. wealth

1. apartment
2. candle _____ a place to live
3. draft _____ chance of something happening
4. horror
5. prospect _____ first rough form of something written
6. timber

1. attack
2. charm _____ gold and silver
3. lack _____ pleasing quality
4. pen _____ not having something
5. shadow
6. treasure

1. boot
2. device _____ army officer
3. lieutenant _____ a kind of stone
4. marble _____ tube through which blood flows
5. phrase
6. vein

1. bulb
2. document _____ female horse
3. legion _____ large group of soldiers or or people
4. mare
5. pulse _____ a paper that provides information
6. tub

1. decade
2. fee _____ 10 years
3. file _____ subject of a discussion
4. incidence _____ money paid for services
5. perspective
6. topic

1. apparition
2. botany _____ ghost
3. expulsion _____ study of plant
4. insolence _____ small pool of water
5. leash
6. puddle

1. adopt
2. climb _____ go up
3. examine _____ look closely
4. pour _____ be on every side
5. satisfy
6. surround

1. betray
2. dispose _____ frighten
3. embrace _____ say publicly
4. injure _____ hurt seriously
5. proclaim
6. scare

1. concrete		1. blend	
2. era	_____ circular shape	2. devise	_____ mix together
3. fiber	_____ top of a mountain	3. hug	_____ plan or invent
4. loop	_____ a long period of time	4. lease	_____ hold tightly in your arms
5. plank		5. plague	
6. summit		6. reject	
1. colleague		1. achieve	
2. erosion	_____ action against the law	2. conceive	_____ change
3. format	_____ wearing away gradually	3. grant	_____ connect
4. inclination	_____ shape or size of something	4. link	_____ finish successfully
5. panel		5. modify	
6. violation		6. offset	
1. arsenal		1. acquiesce	
2. barracks	_____ happiness	2. bask	_____ to accept without protest
3. deacon	_____ difficult situation	3. crease	_____ sit or lie enjoying warmth
4. felicity	_____ minister in a church	4. demolish	_____ make a fold on cloth or paper
5. predicament		5. overhaul	
6. spore		6. rape	
31-40		1. burst	
1. bake		2. concern	_____ break open
2. connect	_____ join together	3. deliver	_____ make better
3. inquire	_____ walk without purpose	4. fold	_____ take something to someone
4. limit	_____ keep within a certain	5. improve	
5. recognize	_____ size	6. urge	
6. wander		1. assist	
1. encounter		2. bother	_____ help
2. illustrate	_____ meet	3. condemn	_____ cut neatly
3. inspire	_____ beg for help	4. erect	_____ spin around quickly
4. plead	_____ close completely	5. trim	
5. seal		6. whirl	
6. shift		1. bleed	
1. abolish		2. collapse	_____ come before
2. drip	_____ bring to an end by law	3. precede	_____ fall down suddenly
3. insert	_____ guess about the future	4. reject	_____ move with quick steps and
4. predict	_____ calm or comfort someone	5. skip	_____ jumps
5. soothe		6. tease	
6. thrive		1. anticipate	
1. convert		2. compile	_____ control something skillfully
2. design	_____ keep out	3. convince	_____ expect something will happen
3. exclude	_____ stay alive	4. denote	_____ produce books and newspapers
4. facilitate	_____ change from one thing	5. manipulate	
5. indicate	_____ into another	6. publish	
6. survive			

1. blaspheme		1. clinch	
2. endorse	_____ slip or slide	2. jot	_____ move very fast
3. nurture	_____ give care and food to	3. mutilate	_____ injure or damage
4. skid	_____ speak badly about God	4. smolder	_____ burn slowly without flame
5. squint		5. topple	
6. straggle		6. whiz	
41-50			
1. original		1. brave	
2. private	_____ first	2. electric	_____ commonly done
3. royal	_____ not public	3. firm	_____ wanting food
4. slow	_____ all added together	4. hungry	_____ having no fear
5. sorry		5. local	
6. total		6. usual	
1. annual		1. dim	
2. concealed	_____ wild	2. junior	_____ strange
3. definite	_____ clear and certain	3. magnificent	_____ wonderful
4. mental	_____ happening once a year	4. maternal	_____ not clearly lit
5. previous		5. odd	
6. savage		6. weary	
1. casual		1. gloomy	
2. desolate	_____ sweet-smelling	2. gross	_____ empty
3. fragrant	_____ only one of its kind	3. infinite	_____ dark or sad
4. radical	_____ good for your health	4. limp	_____ without end
5. unique		5. slim	
6. wholesome		6. vacant	
1. equivalent		1. alternative	
2. financial	_____ most important	2. ambiguous	_____ last or most important
3. forthcoming	_____ concerning sight	3. empirical	_____ something different that
4. primary	_____ concerning money	4. ethnic	_____ can be chosen
5. random		5. mutual	_____ concerning people from a
6. visual		6. ultimate	_____ certain nation
1. auxiliary		1. dubious	
2. candid	_____ bad-tempered	2. impudent	_____ rude
3. luscious	_____ full of self-importance	3. languid	_____ very ancient
4. morose	_____ helping, adding support	4. motley	_____ of many different kinds
5. pallid		5. opaque	
6. pompous		6. primeval	

APPENDIX B: PRODUCTIVE VOCABULARY LEVELS TEST

Fejezd be az elkezdett szavakat úgy, hogy értelmes és nyelvileg helyes mondatokat kapjunk. A hiányzó betűk száma változó. Minden esetben csak egy szó hiányzik.

1. I'm glad we had this opp..... to talk.
2. There are a doz..... eggs in the basket.
3. He has a successful car..... as a lawyer.
4. The thieves threw ac..... in his face and made him blind.
5. Soldiers usually swear an oa..... of loyalty to their country.
6. The voter placed the ball..... in the box.
7. There has been a recent tr..... among prosperous families toward a smaller number of children.
8. The ar..... of his office is 25 square meters.
9. The baby is wet. Her dia..... needs changing.
10. The prisoner was released on par.....
11. Every working person must pay income t.....
12. The pirates buried the trea..... on a desert island.
13. To improve the country's economy, the government decided on economic ref.....
14. She wore a beautiful green go..... to the ball.
15. They keep their valuables in a vau..... at the bank.
16. A bird perched at the window led.....
17. Phil..... examines the meaning of life.
18. According to the communist doc....., workers should rule the world.
19. Second year university students in the US are called soph.....
20. Her favourite flowers were or.....
21. Her beauty and ch..... had a powerful effect on men.
22. La..... of rain led to a shortage of water in the city.
23. The government tried to protect the country's industry by reducing the imp..... of cheap goods.
24. The children's games were amusing at first, but finally got on the parents' ner.....
25. The kitten is playing with a ball of ya.....
26. The thieves have forced an ent..... into the building.
27. Spending many years together deepened their inti.....
28. He usually read the sports sec..... of the newspaper first.
29. The insect causes damage to plants by its toxic sec.....
30. The evacu..... of the building saved many lives.
31. He takes cr..... and sugar in his coffee.
32. The rich man died and left all his we..... to his son.
33. The lawyer gave some wise coun..... to his client.
34. Many people in England mow the la..... of their houses on Sunday morning.
35. The small hill was really a burial mou.....
36. We decided to celebrate New Year's E..... together.
37. Because of the doctors' strike, the cli..... is closed today.
38. There are several misprints on each page of this te.....
39. For many people, wealth is a prospect of unimaginable felic.....
40. She found herself in a pred..... without any hope for a solution.
41. Pup..... must hand in their papers by the end of the week.
42. This sweater is too tight. It needs to be stret.....
43. The farmer sells the eggs that his he..... lays.
44. Sudden noises at night sca..... me a lot.
45. The soldier was asked to choose between infantry and cav.....
46. This is a complex problem that is difficult to compr.....

47. The suspect had both opportunity and mot..... to commit the murder.
48. They insp..... all products before sending them out to stores.
49. The deac..... helped with the care of the poor of the parish.
50. The hurricane whi..... along the coast.
51. Ann intro..... her boyfriend to her mother.
52. Teenagers often adm..... and worship pop singers.
53. France was proc..... a republic in the 18th century.
54. Many people are inj..... in road accidents every year.
55. The angry crowd sho..... the prisoner as he was leaving the court.
56. Don't pay attention to this rude remark. Just ig..... it.
57. A considerable amount of evidence was accum..... during the investigation.
58. The victim's shirt was satu..... with blood.
59. Some coal was still smol..... among the ashes.
60. The dead bodies were mutil..... beyond recognition.
61. If you blow up that balloon any more it will bu.....
62. In order to be accepted into the university, he had to impr..... his grades.
63. Suddenly he was thru..... into the dark room.
64. He perc..... a light at the end of the tunnel.
65. The management held a secret meeting. The issues discussed were not disc..... to the workers.
66. We could hear the sergeant bel..... commands to the troops.
67. He is irresponsible. You cannot re..... on him for help.
68. It's impossible to eva..... these results without knowing about the research methods that were used.
69. She was sitting on a balcony and bas..... in the sun.
70. For years waves of invaders pill..... towns along the coast.
71. The telegram was deli..... two hours after it had been sent.
72. The differences were so sl..... that they went unnoticed.
73. Children are not independent. They are att..... to their parents.
74. She showed off her sle..... figure in a long narrow dress.
75. The boss got angry with the secretary and it took a lot of tact to soo..... him.
76. We do not have adeq..... information to make a decision.
77. He finally att..... a position of power in the company.
78. The story tells about a crime and subs..... punishment.
79. The rescue attempt could not proceed quickly. It was imp..... by bad weather.
80. I wouldn't hire him. He is unmotivated and indo.....
81. The dress you're wearing is lov.....
82. He wasn't very popu..... when he was a teenager, but he has many friends now.
83. She has been changing partners often because she cannot have a sta..... relationship with one person.
84. You must wear a bathing suit on a public beach. You're not allowed to bath na.....
85. She is not a child, but a mat..... woman. She can make her own decisions.
86. The prisoner was put in soli..... confinement.
87. In a hom..... class all students are of a similar proficiency.
88. The urge to survive is inh..... in all creatures.
89. Computers have made typewriters old-fashioned and obs.....
90. Watch out for his wil..... tricks.

APPENDIX C: LEX30

A következő egy asszociációs feladat. Írj a megadott szavak mellé három olyan angol szót, ami eszedbe jut a megadottakról! Bármilyen szófajú szót leírhatsz, de ne használj rövidítéseket (pl. EU), tulajdonneveket (John), vagy több szóból álló kifejezéseket, címeket. Sorban haladj!

attack
board
close
cloth
dig
dirty
disease
experience
fruit
furniture
habit
hold
hope
kick
map
obey
pot
potato
real
rest
rice
science
seat
spell
substance
stupid
television
tooth
trade
window

APPENDIX D: QUESTIONNAIRES

1. Questionnaire used with Groups 2 and 3

1. Hány éve tanulsz angolul? _____ év
2. Összesen eddig mennyi időt töltöttél angol nyelvű országokban? _____ hónap

3. Egy szokásos hétköznapi általában hány órát töltesz az angollal (beleértve a tanórákat, órai készülést és minden egyéb tevékenységet, ami angolul folyik)? _____ óra
4. Egy-egy hétvégén (szombat + vasárnap együtt) általában mennyi időt töltesz az angollal? _____ óra

- 5/A. Hetente átlagosan hány oldalt olvasol el angolul (1 oldal = kb. egy A4-es, 30 soros szöveg)? _____ oldal
- 5/B. Ebből mennyi kapcsolódik az egyetemi tanulmányaidhoz? _____ oldal
- 5/C. Van nyelvi nehézséged az angol órákon kiadott könyvek/cikkek olvasásával?
 - a/ szinte soha
 - b/ néha sok időt vesz igénybe az olvasás, de segítség nélkül boldogulok
 - c/ mindig sokat kell szótároznom és mások segítségét kérem az értelmezéshez
 - d/ a kiadott szövegek többsége túl nehéz számomra, ezért ezeket nem olvasom el

- 6/A. Hetente általában hányszor fogalmazol hosszabb (min. 10 mondatos), összefüggő angol szöveget (házi esszét, e-maileket, leveleket, stb.)? _____ alkalommal
- 6/B. Hetente általában hányszor fogalmazol hosszabb (min. 10 mondatos), összefüggő magyar szövegeket (esszét, e-maileket, leveleket, stb.)? _____ alkalommal

7. Amikor angolul írsz hosszabb szövegeket, milyen szempontokra figyelsz a leginkább?
A következő szempontokat állítsd sorrendbe, a legfontosabbat 1-essel jelöld, a legkevésbé szem előtt tartottat 4-essel:
 - ___ a szöveg megszerkesztettsége
 - ___ választékos szókinés
 - ___ nyelvtani helyesség
 - ___ helyesírás

8. Általában milyen segítséget használsz, ha angolul fogalmazol? A gyakoriságot számmal jelöld! 2 = gyakran, 1 = ritkán, 0 = soha

- ___ egynyelvű angol szótárt,
 - ___ magyar–angol szótárt,
 - ___ angol–magyar szótárt,
 - ___ angolul nálam jobban tudó barátot/családtagot,
 - ___ szövegmintákat,
 - ___ a témához kapcsolódó internetes szövegeket
- Egyéb: _____

Amikor nem tudsz ilyen segítséget használni (pl. vizsgán) mit teszel a következő helyzetekben:

9. Melyik módszert alkalmazod a szöveg megtervezésénél és az írás során a leggyakrabban?

- a/ magyarul kitalálom és mondatról mondatra megpróbálom angolra fordítani
- b/ a mondatok egy része rögtön angolul ugrik be, egy részét pedig magyarról fordítom
- c/ csak olyat írok le, ami rögtön angolul jutott az eszembe

10. Amikor írás közben egy magyar szó/kifejezés legközelebbi angol megfelelője nem jut az eszedbe, mit teszel leggyakrabban?

- a/ jelentésben közel álló másik szót/kifejezést használok
- b/ kihagyom a szövegből az ide vonatkozó információt

11. Ha nem vagy biztos egy használandó szó helyesírásában, de nem tudod azt leellenőrizni, melyik módszert alkalmazod a leginkább?

- a/ megcélzom a szót/kifejezést és használom a szövegben
- b/ más, értelmileg hasonló szóval/kifejezéssel helyettesítem

2. Questionnaire used with Groups 1 and 2 (English translation)

1. Number of years English studied: _____

2. Number of months spent abroad in an English speaking country: _____

3. Average number of hours spent with English on weekdays (including classes, preparation for classes and extracurricular activities): _____

4. Average number of hours spent with English/week at weekends (Saturday + Sunday together):

5/A. Average number of pages read in English/week (1 page= 1 A/4 size page, with approx. 30 lines):

5/B. How much of these are related to your university studies? _____

5/C. Do you have any language problems while reading assigned books/articles?

a/ almost never

b/ sometimes it takes long, but I can manage without help

c/ I always have to check dictionaries and ask others to help

d/ most of the assigned reading materials are too difficult for me, therefore, I do not read them

6/A. Number of times you write longer texts/week in English (including homework assignments, e-mails, letters, etc., containing a minimum of 10 sentences): _____

6/B. Number of times you write longer texts/week in Hungarian (including essays, e-mails, letters, etc., containing a minimum of 10 sentences): _____

7. During essay writing what aspects do you focus on the most? Put the following aspects in order, marking the most important aspects with 1 and the least important aspect with 4.

organization _____

sophisticated vocabulary _____

grammar _____

spelling _____

8. What form of help do you usually use when you write essays? (write 2 if you use it often, 1 if rarely, 0 if never)

- English monolingual dictionary
- Hungarian–English dictionary
- English–Hungarian dictionary
- friend with better English proficiency
- sample essays
- Internet articles on the topic

other: _____

When you cannot use any help (e.g. exam situation) what do you do in the following situations?

9. What is the way you plan and write your essay more often?

- a/ I plan it in Hungarian and then translate it into English sentence by sentence
- b/ part of it comes directly in English, part of it is translated in my head
- c/ I write down only the ideas that come in English

10. When during writing the closest equivalent of a Hungarian word/expression does not come to your mind in English, what do you do more often?

- a/ use another word/expression close to it in meaning
- b/ leave out this part of the information from the essay

11. If you are not sure of the spelling of a word you want to use and cannot check it, which strategy do you use more often:

- a/ guess the spelling of this word/expression and use it
- b/ replace the word/expression with another one close in meaning

3. Questionnaire used with Group 1

1. Hány éve tanulsz angolul? _____
2. Éltél-e hosszabb ideig angol nyelvterületen? _____ (hét/hónap)
3. Van-e valamilyen angol nyelvvizsgád? _____
Milyen típusú?, _____
4. Mire használod az angol nyelvet az egyetemen kívül?

4. Questionnaire used with Group 1 (English translation)

1. Number of years English studied: _____
2. Number of weeks/months spent abroad in an English speaking country: _____
3. Type of English language proficiency exam: _____
4. Activities English is used for outside of the university:

APPENDIX E: WRITTEN PRODUCTION TASKS

Prompts used with first-year students

1. The Department of Psychology wants to find out how people understand some basic concepts in life. They have asked students to write essays in which they define and explain what “peer pressure” is.
2. If you could invent something new to greatly improve the lives of average Hungarians, what would it be? For a document to be sent to the “Inventions Registry Office” explain what your invention would do and how it would work.
3. The university has a plan for students to evaluate teachers who teach them. Each teacher would be evaluated by students who they teach. Write an argumentation to be published in the student newspaper in which you support this idea.
4. It is often said that “experience is the best teacher”. That is, it is claimed that what is learned outside the classroom through real life experiences is more important than what is learned inside the classroom. The Department of Education is collecting essays on this topic. Write an argumentation where you disagree with this idea.

Prompts used with third-year students (Group 3a)

5. Renewable and alternative sources of power – such as windmills and solar panels– have been popular areas of research. But technologists admit that these sources have not been able to compete with power derived from fossil fuels to supply large amounts of energy. Making power from higher-cost alternative sources is often less attractive.

Write an essay to be published in an English-language paper in which you

A. argue why you believe that the government should finance the research on and the introduction of new sources of power.

OR

B. argue why you believe that time and money should be spent on the optimal use of already available sources of power.

6. Regardless of the fact that many countries have laws that prohibit the indication of the gender, age and other personal information of the candidates in job advertisements, we can often read the following and similar messages: “...company is seeking men ages 30–40...”

Write an essay to be published in an English-language paper in which you

A. argue why you think that the above mentioned example is unacceptable and the laws should be enforced more severely.

OR

B. argue why you believe that the indication of the gender and age of potential candidates for jobs should be acceptable and can help future employees find the most suitable position.

7. An article has been published in the magazine *Health and the Countryside*, which argues that although spending time outside is mentally and physically healthy, and that it is constructive as well, these days it may not only be a positive experience. Sprays, disinfectants, fertilizers, etc. pollute the air, and very often it may be safer to stay indoors.

Write a letter to the Editor of the magazine expressing your views and concerns on this matter.

A. Argue why open air activities are a must for everybody and why chemicals and air pollution should not discourage people from staying outside.

OR

B. Argue why staying indoors these days is safer than staying outside.

Prompts used with third-year students (Group 3b)

8. Recently the phenomenon of the 'brain drain' has become familiar to Hungarians: many of our doctors and engineers have taken up employment in Western Europe hoping to make a better living. Some people are concerned that this is an irreversible process which will do considerable harm to both our society and our economy.

Write an essay to be published in a newspaper in which you

A. argue that something should be done to prevent highly qualified professionals from leaving the country for a better future in the West.

OR

B. argue that there is nothing to be feared as professionals go abroad to gain experience, and when they return, they can work more efficiently in Hungary.

9. An article has recently been published in a national newspaper in Britain suggesting that the British Library (one of the largest copyright libraries in the world) should stop issuing real books to readers, but make all its books instantly available as e-books, making enormous savings on staffing, storage and transportation costs.

Write an essay to the newspaper either

A. supporting the move towards the e-book and the 'paperless library' as cutting-edge technology and cultural democratization

OR

B. arguing for the preservation and continual use of printed books