University of Pécs

Doctoral School of Linguistics

Doctoral Programme in Applied Linguistics

ASSESSING ENGLISH MAJORS' VOCABULARY
AT THE UNIVERSITY OF PÉCS

WORKING TOWARDS A CORPUS-BASED APPROACH

Lehmann Magdolna

Consultant: Dr. Andor József

Pécs

2009

TABLE OF CONTENTS

Acknowledgements

List of abbreviations and acronyms

AWL	Academic Word List
CORES	Corpus of Readings in English Studies
EAP	English for academic purposes
EFL	English as a foreign language
ESL	English as a second language
FL	foreign language
GSL	General Service List
L1	first language/mother tongue
L2	second language/ foreign language
SLA	Second Language Acquisition
SPSS	Statistical Package for Social Sciences
TEFL	Teaching English as a Foreign Language
UP	University of Pécs
VLT	Vocabulary Levels Test
VKS	Vocabulary Knowledge Scale

List of tables

List of figures

Introduction

The mysteriously convoluting strings of letters in black ink against white paper have fascinated my imagination from the first encounters on. I remember begging my parents to teach me the letters of the alphabet one by one, who were rather unwilling as at that time school teachers did not like first-graders who could read and write. But I insisted as although picture books were colourful enough to catch my attention, I felt convinced that the knowledge of those curlicued shapes would take me to an unknown territory I was eager to discover. Since then words have proved to be not mere strings of letters but gates to understanding the world around me and it appears that the more I learn about words the more improbable it becomes to ever be able to fully grasp all dimensions of the worlds they promisingly open windows to.

One such micro-world that raised my particular interest as a researcher and university lecturer is the domain of words in an English for academic purposes (EAP) context. The rationale for the studies reported in this dissertation was the observation that Hungarian students of English seem to have difficulties in text comprehension and academic writing; a phenomenon that may root in the lack of sufficient or adequate vocabulary knowledge.

This dissertation is divided into two parts and seven chapters, as shown in Table 1. The three chapters in Part I provide the theoretical background to the subsequent empirical studies and centre around the three major issues in current vocabulary research: vocabulary description, acquisition and assessment. The first chapter introduces the essential terminology for vocabulary studies and overviews the most influential theories on how words are defined, what knowing a word may constitute and how words are stored in and retrieved from the mental lexicon. The second chapter synthesises the findings of research on how vocabulary is acquired, touching

upon similarities and differences in the mechanisms and scope of first language (L1) and second language (L2) vocabulary acquisition, and the much debated questions of how vocabulary is learnt most efficiently. The third chapter overviews current theories, methods and instruments available for assessing various levels of lexical knowledge.

These issues have inspired and constitute the leitmotives of years of empirical research conducted at the Department of English Applied Linguistics, University of Pécs reported in the following chapters in Part II of the present dissertation. The studies aim to explore the vocabulary knowledge of English majors as compared to vocabulary requirements imposed on them by curricular requirements, the challenging reading load and writing assignments. Chapter 4 sets the context for the empirical studies by introducing the participants, the curriculum and the background to vocabulary testing at the department, and by providing a methodological overview of the studies. The empirical chapters follow the developmental sequence of the test battery devised to assess the vocabulary knowledge of first-year students of English.

Table 1. The data sources and methods of analysis used for the main research questions

	Research question	**Data sources**	**Methods of analysis**
Pilot Study 1 93 participants	How many words do first-year English majors know in the English language? In other words, what is their vocabulary size and is it sufficient for coping with academic readings?	Vocabulary size test (Goulden, Nation & Read, 1990)	Descriptive statistics
Pilot study 2 33 participants	What is the relationship between vocabulary size and free productive vocabulary use? Do learners with a larger vocabulary show a wider range of lexical expression in their writings, than their fellow students with smaller vocabulary?	Vocabulary size test (Goulden, Nation & Read, 1990) Student essays	Corpus analysis of student scripts Correlational analysis
Pilot study 3 33 participants	Is intentional vocabulary learning and conscious preparation for regular vocabulary tests more effective in the long-term retention of words than incidental vocabulary learning as a by-product of reading only?	Vocabulary size test (Goulden, Nation & Read, 1990) Vocabulary progress tests Qualitative background questionnaire	Correlational analysis

	Research question	Data sources	Methods of analysis
Stage 1 220 participants	How familiar are English majors with academic vocabulary essential for pursuing university studies?	Vocabulary Levels Test (Schmitt, Schmitt & Clapham, 2001)	Descriptive statistics Correlational analysis
	How well does the academic section of the Vocabulary Levels Test (Schmitt, Schmitt & Clapham, 2001) discriminate between candidates with larger and smaller vocabularies?	Vocabulary Levels Test (Schmitt, Schmitt & Clapham, 2001)	Item analysis
Stage 2 211 participants	Is academic or low-frequency vocabulary a better indicator of student vocabulary for the purposes of filtering first-year English majors lacking in lexis necessary for pursuing academic studies? Is it academic or low-frequency vocabulary that differentiates among candidates better?	Vocabulary Levels Test (Schmitt, Schmitt & Clapham, 2001)	Correlational analysis Item analysis
Stage 3 135 participants	Is receptive or productive vocabulary a more reliable predictor of scores on a reading comprehension, a grammar and usage, a listening comprehension and a writing test? How do receptive and productive vocabulary measures relate to general language proficiency?	Proficiency test A test of lexis intuitively perceived by the researcher as useful in English studies Proficiency test A test of lexis intuitively perceived by the researcher as useful in English studies	Corpus analysis of student scripts Correlational analysis Correlational analysis
	What is the relationship between receptive and productive vocabulary measures?	Proficiency test A test of lexis intuitively perceived by the researcher as useful in English studies	Linear regression analysis
	Which one of the 9 measures investigated in the study is a good predictor of overall academic performance?	Proficiency test A test of lexis intuitively perceived by the researcher as useful in English studies	Linear regression analysis
Stage 4 134 participants	What lexis do English majors at UP need to be familiar with to be likely to comprehend the required compulsory readings?	Corpus of Readings in English Studies at UP (CORES)	Corpus analysis Vocabulary profiling
	Based on evidence from corpus analysis, to what extent may the knowledge of the words included in the Academic Word List (Coxhead, 2000) be beneficial for students in text comprehension?	Corpus of Readings in English Studies at UP (CORES)	Corpus analysis Vocabulary profiling
	What low-frequency words are students likely to encounter in their compulsory readings?	Corpus of Readings in English Studies at UP (CORES)	Corpus analysis Vocabulary profiling
	How familiar are the students with the most frequent academic and low-frequency words occurring in the corpus of compulsory readings?	Corpus-based vocabulary test	Descriptive statistics Item analysis

The first of three pilot studies reported in Chapter 5 discusses how many words first-year students know in the English language to find out whether the size of their receptive vocabulary would make them capable of pursuing academic studies in English. Results are compared to the latest findings of research on native speaker vocabulary size and the minimal vocabulary size necessary for learners of English as a foreign language for academic purposes. Besides estimating the receptive vocabulary size of the participants, however, further effort is devoted to initial explorations into the relationship between receptive vocabulary size and productive vocabulary use, as well as incidental and intentional vocabulary learning in two other studies.

The development of a test is at best characterized by constant improvement adjusted to changing needs and conditions and incorporates the latest achievements of relevant research. On the basis of this assumption Chapter 6 is devoted to introducing the reader to the developmental stages of the new battery which involves introducing vocabulary testing as a separate sub-test of the long-existing proficiency test developed and administered at the Department of English Applied Linguistics. Two trials of an acknowledged and widely used vocabulary test were followed by devising and administering a test for assessing the vocabulary intuitively perceived as indispensable in pursuing English studies.

The aim of the study introduced in Chapter 7 is to inquire into the nature of the lexis English majors are expected to be familiar with in the course of their studies, therefore, the estimation of the students' vocabulary size is followed by a corpus analysis of a representative sample of compulsory readings. The results obtained from the analysis of the corpus will then serve as a basis of comparison with the words the students know on various levels of word knowledge on the receptive/productive continuum and the development of a new test battery better serving the lexical needs of students of English than the widely used vocabulary tests available at present.

Part I

Vocabulary in second-language acquisition research

Chapter 1

What is vocabulary?

1.1 Introduction

T he study of words and word knowledge is a common concern of numerous interrelated disciplines, such as morphology, semantics, pragmatics, psycholinguistics, sociolinguistics, neurolinguistics, language pedagogy, and more recently even translation studies, showing the vast complexity and richness of the field. The present work is an endeavour of a much narrower focus into the pedagogical and assessment aspects of vocabulary studies. However, before embarking on a detailed investigation of how we learn words and how our knowledge of words can be assessed for pedagogical purposes, in this first chapter it will be useful to look at and clarify some terms and notions readers of the field under investigation might encounter. On the one hand, I aim to provide a general introduction for the reader and set the theoretical background for the subsequent chapters, where these terms are used as self-explanatory. On the other hand, this overview will provide the context for the general problems that arise in linguistics, applied linguistics and the related disciplines posed by empirical data.

1.2 Recurring terms in vocabulary acquisition research

To begin with the broadest term appearing first in the title of this section, 'vocabulary' is often referred to as the total number of words of a language, a text, or the total

number of words a native speaker or a language learner of a target language knows. Resorting to the use of the Greek word 'lexicon' for dictionary, the term 'mental lexicon' denotes the words present in our minds. Williams (1994, p. 7), however, emphasises the importance of making a clear distinction between two notions of the lexicon, which have been identified as one in modern times. One is the Bloomfieldian lexicon, which he calls "a repository of all of a language's idiosyncracies" (Williams, 1994, p. 7), the other is the grammatical lexicon, which is interpreted as "the linguist's theory of the category of linguistic object we call a 'word'" (Williams, 1994, p. 7). The domain of the lexicon is defined by Singleton (1999, p. 15) as

> ... that component of a language which has to do with what one might call local phenomena – the meanings of particular elements of a given language, the phonological and orthographic forms of these elements, and the specific ways in which they collocate and colligate.

In his view, these constitute the various dimensions of the lexicon. Lexis is the Greek word for 'word'; therefore, 'lexicology' is the study of words and the lexicon, while lexicography concerns the theory and practice of writing dictionaries (Martsa, 2007, p. 5).

The question of what a word is may be of special interest to linguists on a theoretical level and to applied linguists from the perspective of language teaching, learning and testing. However, the concept of a word is not easy to define either in theoretical terms or for practical purposes. Researchers and readers of the field encounter recurring terms such as lexeme, word, word family, word type and token, lemma, lexical item, lexical phrase or multi-word unit. Singleton (1999, p. 10) points out that although the word is central to both specialist and non-specialist understanding, it is the level of abstraction that determines how we interpret it.

To the non-specialist, looking at words as strings of letters divided from other strings of letters by space is one possible level of abstraction in orthographic terms. On the basis of simply counting how many words a sentence or a text is constructed of, the

famous line of a Beatles song "She loves you, yeah, yeah, yeah" thus comprises six words altogether, so it can be said to contain six tokens (i.e. the total number of words in a text). However, the word *yeah* appears three times, so there are four different words in it, these are called word types (Davies, Brown, Elder, Hill, Lumley, & McNamara, 1999).

To the specialist, Carter defines words from a semantic point of view as "the minimum meaningful unit of language" (1998, p. 5), while he also highlights problems with this definition supposing clear relations between words and meanings, citing units of meaning which are represented by more than one words such as *bus conductor* or *school teacher* (Carter, 1998, p. 5). Based on this approach, there are lexical words conveying some semantic meaning (often referred to as content words or full words, e.g., *book*, *shadow*) as opposed to grammatical words signalling some grammatical relationship or function (also known as function words, empty words or form words, e. g., *of, to, and, the).*

Others claim that the smallest unit of language with a lexical and/or grammatical meaning is a morpheme (Caron, 1992, p. 48.; Martsa, 2007, p. 11). Bloomfield (1933) defined words on a grammatical level as a minimal free form, proposing that a word is a word if it can stand on its own and cannot be subdivided further. Carter (1998), however, warns that this definition assumes a basic stability of the word and explains that although some words such as *my* or *because* may stand on their own as a reply to a question, their meanings cannot be independent from their context and relation to other words. Or let us consider idioms consisting of several orthographic (written) forms: they may stand on their own and be substituted with one word, but their meanings cannot be reduced without loss (e.g., *to rain cats and dogs = to rain heavily*).

To solve this problem, abstract units based on a collection of possible forms they take are usually called lexemes, these are what we usually find listed in a dictionary (Martsa, 2007, p. 10); whereas its concrete possible representations, "the graphic

forms realizing and constituting the inflectional forms of lexemes" (Martsa, 2007, p. 10) are called word forms (see also Carter, 1998; Singleton, 1999): the lexeme BRING may display word forms such as *'brought'* or *'brings'*. Specialists further divide the level of lexical representation and collectively call the base and inflected forms of a word (e.g., *go, going, goes, went, gone*) a lemma (Caramazza, 1999; Kempen & Huijbers, 1983). Base words do not only take inflectional endings, but have various derivatives, such as *happy, happily, happiness*; these sets of words together with the base form are known as a word family. Read (2000, p. 20) explains that vocabulary consists of more than just single words. For instance, phrasal verbs, compound nouns or idioms are recognized as lexical units or lexical items (Carter, 1998, p. 7) consisting of more than one word, named multi-word units or multi-word items, and these phrases or sentences have meanings that cannot be guessed by knowing only the meaning of the individual words constituting them (e.g., *to kick the bucket*). That is, such expressions are non-compositional in nature.

Researchers found that speakers have a large amount of such prefabricated language or lexical chunks in stock at their disposal. Pawley and Syder (1983, p. 208) pointed out that "memorised sentences and phrases are normal building blocks of fluent discourse" and at the same time they are models for the creation of many possible other phrases that may finally become lexicalized, i.e. enter common usage. Wray (2002, p. 9) agrees and proposes the term formulaic sequence instead, defined as

> 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.

These formulaic sequences form the basis of Wray's (2002) revolutionary model of the mental lexicon discussed in section 1.6.2 of this chapter.

Nattinger and DeCarrico (1992, pp. 36-47) developed a similar concept, lexical phrase, which they define as "lexico-grammatical units that occupy a position somewhere

between the traditional poles of lexicon and syntax". They explain that lexical phrases are treated as one unit, but can be derived from regular rules of syntax and their use is restricted by principles of pragmatic competence. They emphasize that the main difference between conventionalized or frozen forms like idioms or clichés and lexical phrases is that the latter are used to perform certain pragmatic functions in discourse, e.g., saying '*How are you doing*?' serves the function of greeting.

Categorizing lexical phrases based on their length and grammatical status, canonical or non-canonical shape, being variable or fixed and continuous or discontinuous (i.e. whether or not interrupted by variable lexical fillers), Nattinger and DeCarrico (1992, pp. 36-47) identify the following four categories. The first of these, polywords, are short fixed expressions that are used to perform a wide variety of functions, such as qualifying (*for the most part*), summarizing (*in a nutshell)*, shifting topic (*by the way*), marking approval, agreement or disagreement (*hold your horses*), relating (*for that matter*), parting (*so long*), evaluating (*strictly speaking*) or clarifying (*you know*), including those relator polywords that are written as one, such as *nevertheless*, or *moreover*.

The second category is named institutionalized expressions, which are defined as sentence length lexical phrases functioning as separate utterances, such as proverbs, aphorisms or formulas for social interaction. Their functions may be greeting (*How do you do?, How are you?*), parting (*Have a nice day!*), expressing approval or disapproval (*there you go*), closing (*nice meeting you*), giving advice (*a watched pot never boils),* or concession *(be that as it may).*

Category three involves phrasal constraints. These are short or medium length phrases which may allow for substitutable variations, such as in *a _____ ago*, with possible variations like *a day ago, a year ago, a long time ago*. Finally, category four is sentence builders, i.e. lexical phrases which provide the framework for whole sentences and contain slots for "parameters and arguments for expression of an entire idea" (Nattinger & DeCarrico, 1992, p. 42), where such slots for substitution are

marked by capital letters (*I think that X; not only X, but also Y; my point is that X; I am a great believer in X;* or *let me start by/with X*). The functions each phrase performs may be, among others, that of assertion, relation, summarizing, or marking topic, respectively.

As can be seen from the above, the terminology of vocabulary studies is wide and far from being unified in its present state; definitions and usage vary with the interpretation of the relationship of word and meaning, lexis and grammar. In the literature of the field the use of 'lexicon', 'lexis' and 'lexical item' usually signals a wider perspective than the general association of vocabulary with single words, whereas 'vocabulary' is most often referred to as the subject matter of research. 'Word' is generally used to refer to individual units, unless specialized fields of abstraction, such as multi-word units or items need to be specified. In my discussions I intend to follow this approach in the use of these terms.

1.3 Vocabulary and language competence

In the previous section, we have seen what basic terms we may encounter in the study of words and I attempted to clarify the somewhat overlapping multitude of definitions and the probably baffling picture this dynamically developing area of linguistics, and more recently applied linguistics, displays at present. I have yet to explain, however, what vocabulary knowledge is and how it fits into the wider picture of language competence.

Models of communicative competence have evolved through various stages (Bachman, 1990; Canale, 1983; Canale & Swain, 1980; Celce-Murcia, Dörnyei & Thurrell, 1995; Hymes, 1971; Munby, 1978; Swain, 1985) to be defined as divisible into two major components: organizational knowledge (or competence) and pragmatic knowledge (Bachman & Palmer, 1996, p. 68). Organisational knowledge is responsible for how utterances or sentences and texts are organised, while pragmatic knowledge is concerned with how utterances or sentences and texts are related to the

communicative goals of the language user and to the features of the language use setting.

Organizational competence comprises those abilities that are involved in "controlling the formal structure of language for producing or recognizing grammatically correct sentences, comprehending their propositional content, and ordering them to form texts" (Bachman, 1990. p. 87). It involves grammatical knowledge about how individual sentences and utterances are organized; and textual knowledge describing how utterances or sentences form texts. Pragmatic competence, on the other hand, comprises two categories: functional and sociolinguistic knowledge. This model describes vocabulary knowledge as part of grammatical knowledge within the field of organizational knowledge, together with the knowledge of syntax and phonology/graphology, as indicated in Table 2, based on the competencies described earlier by Widdowson (1978):

Table 2. Areas of language knowledge (Bachman & Palmer, 1996, p. 68)

_____

Organisational knowledge
(how utterances or sentences and texts are organised)
    Grammatical knowledge
    (how individual utterances or sentences are organised)
    Knowledge of vocabulary
    Knowledge of syntax
    Knowledge of phonology/graphology
    Textual knowledge
    (how utterances or sentences are organised to form texts)
    Knowledge of cohesion
    Knowledge of rhetorical or conversational organisation
Pragmatic knowledge
(how utterances or sentences and texts are related to the communicative goals of the language user and to the features of the language use setting)
    Functional knowledge
    (how utterances or sentences and texts are related to the communicative goals of language users)
    Knowledge of ideational functions
    Knowledge of manipulative functions
    Knowledge of heuristic functions
    Knowledge of imaginative functions
    Sociolinguistic knowledge
    (how utterances or sentences and texts are related to feature of the language use setting)
    Knowledge of dialects/varieties
    Knowledge of registers
    Knowledge of natural or idiomatic expressions
    Knowledge of cultural references and figures of speech

_____

Although Bachman (1990) emphasises that this table of hierarchical relations displays the competencies as separate and independent, while in language use these all interact with each other and the situation of language use, the model has been widely criticised on various grounds both by linguists and applied linguists. On the one hand, from the applied linguist's perspective, Read (2000) argues that the restrictive categorization of vocabulary knowledge under grammatical knowledge is an understatement of the extent by which vocabulary contributes to language knowledge. He claims that vocabulary knowledge is largely represented in sociolinguistic knowledge as defined in the Bachman and Palmer (1996) model in the form of 'natural or idiomatic expressions', 'cultural references and figures of speech', as well as 'registers' associated with varieties of language use by particular users in particular contexts. McCarthy's argument (1990, pp. 61-64) supports this claim by pointing out that the primary characteristic feature of a register is the distinctive words and phrases in it (on register see also Biber & Finnegan, 1994; and Lee, 2001).

The Bachman and Palmer (1996) model has been disputed by pragmatists and lexical semantists as well, pinpointing the controversy in the definitions of the subcategories: utterance is a pragmatic term, whereas sentence is not. It also lacks reference to the role of morphological knowledge and the type of lexical knowledge connected to it (for an overview see Martsa, 2007). Similarly to the views of applied linguists on the role of context, one of the most recent interpretations of how pragmatics and semantics become integrated in discourse, Pustejovsky's theory of the Generative Lexicon (1995, 2001) discusses how meaning is built up and modified by context, which the Bachman and Palmer (1996) model treats separately.

The same problem arises in the separation of the lexicon and grammar, which has more recently been seen as one unit, termed lexicogrammar by Halliday (1985) within the framework of his Systemic Functional Theory of Language, combining syntax, lexicon and morphology again treated separately by Bachman and Palmer (1996) under the umbrella of grammatical knowledge. For evidence, Biber, Conrad and Reppen (1998, p. 84) explain that two nearly synonymous words may be distinguished

by their grammatical constructions, i.e. their different preferences for attributive or predicative positions (e.g., *little/small; begin/ start*); or two nearly synonymous constructions are distinguished by their associations with different classes of words (e.g., *that/to* clauses).

Halliday (1985) explains that language is stored in larger, often unanalysed chunks. Chapelle (1998, p. 2) adds that assuming the existence of a single lexicogrammar can explain why "some lexical/syntactic patterns are more likely to co-occur than others depending on register" in corpus-based linguistic research. But before discussing how language, or more closely, words are stored in the mental lexicon to be overviewed in the next section of this chapter, it is necessary to explain what components vocabulary knowledge may comprise.

1.4 Components of vocabulary knowledge

In line with the above outlined views on the role of context, Chapelle (1994) developed the Bachman (1990) model further and proposed a comprehensive definition of vocabulary knowledge or vocabulary ability including both knowledge about language and knowledge about how to use language. The three components of this framework are: (1) the context of vocabulary use; (2) vocabulary knowledge and fundamental processes; and (3) metacognitive strategies for vocabulary use. Let us briefly examine each.

1.4.1 Context of vocabulary use

Language use is never used in a vacuum but in context. As Pinker (2007, p. 9) explains, words are tied to reality, to things and situations in the world around us. Moreover, words are anchored to people, emotions and social relations determining how reality is interpreted in our heads. As for the context of vocabulary use, Chapelle (1994, p. 164) applies the framework of Halliday and Hasan (1989) to explore the social factors affecting word choice: field, tenor and mode. The element of 'field' refers to the subject matter and the type of activity the language user is engaged in.

'Tenor' means the role relationships and the relative social status of the language user; while 'mode' concerns the channel of communication and the features that distinguish spoken and written language use. In a communicative view, explains Read (2000, p. 29), the context of vocabulary use, i.e. the social and cultural situation in which a word is used, has a significant influence on its meaning.

Read (2000, p. 31) identifies three areas where context in this sense can be influential on word choice: (1) differences across generations and between colloquial and more formal uses of words; (2) differences in interpretation across language varieties; and (3) differences between everyday usage and specialized terminology in particular fields of study. Pointing to the social linkage of lexical items, Skehan (1998, p. 37) further claims that besides the considerable difference in the size of the mental lexicon of a language, lexical choice is what distinguishes between a language learner and a native speaker. This view is widely accepted and relied upon in tests of vocabulary knowledge to be introduced in detail in Chapter 3, whereas the biological underpinnings of vocabulary acquisition are further discussed in Chapter 2.

Context cannot be interpreted only in the sociolinguistic sense. Vocabulary learning is also affected by the linguistic environment of the word, such as the syntactic or morphological context it is surrounded by (see Bloom, 2000; Forster, 2006; Martsa, 2007). Formal/logical semantics operates with truth conditions and propositions to describe meaning, while sentence-based cognitive semantics defines it in relation to how we perceive the world and picture it in our mental representations. As we will see later in this chapter, Jackendoff (1990, pp. 84-87) in a generative approach with a conceptual viewpoint assumes a direct relationship between entities and verbs/events in his definition of conceptual meaning. Wierzbicka (1996), a current representative of componential analysis, operates with semantic primitives, whereas in pragmatics meaning is defined by the context of speech as described by Grice's (1975) maxims and speech act theory (see also Austin, 1962 and Searle, 1969). But as Bloom (2000, p. 212) explains, linguistic information is only one source used integrated with information obtained from other inferential sources of context to figure out the

meanings of unknown words.

In the discussion of the various views on communicative competence we have seen how important context may be in modifying meaning for vocabulary use, however, it may play a crucial role in other aspects as well, within the domains of applied linguistics. The effects of context in vocabulary acquisition will be discussed in further detail in relation to vocabulary learning strategies in Chapter 2 and to assessing vocabulary knowledge in Chapter 3.

1.4.2 Metacognitive strategies for vocabulary use

The other major component of vocabulary ability in Chapelle's (1994) framework, metacognitive strategies for vocabulary use, covers what Canale and Swain (1980) call strategic competence, i.e. being able to use vocabulary knowledge with appropriateness in communication. Read (2000) points out that most of these strategies, such as paraphrase, language switch to L1, use of superordinate terms to replace a specific, unknown word, or appeal to authority by asking '*How do you say ...?*', are mostly used to overcome the gap in vocabulary knowledge. However, some communication situations may require more conscious metacognitive strategy use, e.g., in the case of a reading or writing task, or conversing with a child or a non-native speaker. For a comprehensive overview of studies on vocabulary strategies see Chapter 2.

1.4.3 Vocabulary knowledge and fundamental processes

The final major component of Chapelle's (1994) model involves four subcategories: (1) vocabulary size; (2) knowledge of word characteristics; (3) the organization of the lexicon; and (4) fundamental vocabulary processes. In the next section I will introduce research on vocabulary size, what it means to know a word, how the lexicon is organized and what fundamental processes may take place in vocabulary use, before discussing these issues more deeply in the subsequent two chapters.

Research on vocabulary size attempts to explore how many words speakers or learners of a language know. Estimates may serve as guidelines for foreign language learners in what goals to set in the course of their studies. There is an ongoing and unsettled debate (see Chapter 3) on how many words a native speaker of English may know and how the vocabulary size of a language learner may be proportional to it, as well as how many words a language learner is required to know to be able to read and comprehend certain types of texts. The importance of information on vocabulary size is unquestionable; nevertheless, such data should be handled with caution and can only be useful if information on various other aspects of vocabulary use is also available, such as what words a learner needs to know for various purposes, like reading authentic texts or pursuing academic studies at a university in the target language. A commonly used and accepted test of vocabulary size, for instance, was developed by Goulden, Nation and Read in 1990 based on taking representative samples of words from large dictionaries to measure the vocabulary size of native speakers of English. An adapted version of this test was applied for the purposes of the present study to estimate the vocabulary size of first-year students of English at the University of Pécs.

The focus of studies on knowledge of word characteristics is generally paraphrased as what it means to know a word. Speakers of a language know more about some words than about other words (Read, 2000), and researchers tend to agree that word knowledge is not static, but develops along several possible aspects of how well a word is known, i.e. how 'deep' word knowledge is (Henriksen, 1999; Huckin & Coady, 1999; Laufer, 1998). A common feature of the analyses to be introduced below is that words are defined with a set of properties the language user knows or should know about the word, be it conscious or unconscious knowledge.

One of the earliest such sets of criteria was defined by Richards in the 1970s, who made the following basic assumptions about what it means to know a word:

1. Vocabulary knowledge of native speakers expands throughout their lives.
2. Knowing a word means knowing the degree of probability of encountering that word in speech or print. For many words we also know the sort of words most likely to be found associated with the word.
3. Knowing a word implies knowing the limitations on the use of the word according to variations of function and situation.
4. Knowing a word means knowing the syntactic behaviour associated with the word.
5. Knowing a word entails knowledge of the underlying form of the word and the derivations that can be made from it.
6. Knowing a word entails knowledge of the network of associations between that word and other words in the language.
7. Knowing a word means knowing the semantic value of a word.
8. Knowing a word means knowing many of the different meanings associated with the word.

<div align="right">(Richards, 1976, p. 83)</div>

It is important to note here that although the majority of his statements touching upon various aspects of word frequency, collocation patterns, the meaning modifying role of context, as well as knowledge of the syntax, morphology, and the semantic features of the word, were widely debated and modified later, the first assumption on the life-long expansion of native speaker vocabulary knowledge has been commonly accepted and supported by corpus studies. Nation (1990) incorporated some of Richards' (1976) assumptions and arrived at a checklist-like set of questions for determining how much we know about a word (Table 3).

Nation clearly indicates what Richards' list only implies: the distinction of receptive and productive types of knowledge in all four categories of the form, position, function and the meaning of the word, which is a relevant addition, as for learners of a language it may be important what type of word knowledge is required for certain purposes only, e.g., to understand everyday oral conversation or to be able to write a business letter.

Table 3. Components of word knowledge (Nation, 1990, p. 31)

Key: R = receptive, P = productive

_____

Form
    Spoken form   R   What does the word sound like?
                              P   How is the word pronounced?
    Written form   R   What does the word look like?
                              P   How is the word written or spelled?
Position
    Grammatical patterns   R   In what patterns does the word occur?
                              P   In what patterns must we use the word?
    Collocations   R   What words or types of words can be
                                             expected before or after the word?
                              P   What words or types of words must we use
                                       with this word?
Function
    Frequency   R   How common is the word?
                              P   How often should the word be used?
    Appropriateness   R   Where would we expect to meet this word?
                              P   Where can this word be used?
Meaning
    Concept   R   What does the word mean?
                              P   What word should be used to express this
                                     meaning?
    Associations   R   What other words does this word make us
                                     think of?
                              P   What other words could we use instead of this one?

_____

Another related question that might arise from the obligation implied in the model by the use of the word 'should' is the level of consciousness, i.e. how much a language user/learner is aware of this knowledge and whether this knowledge is applied consciously in communication. Later, Laufer (1997b) in her overview of approaches to definitions of a word describes possessing the following information on word characteristics as necessary in order to know a word:

    a) Form - spoken and written; that is pronunciation and spelling.
    b) Word structure – the basic free morpheme (or bound morpheme) and the common derivations of the word and its inflections.
    c) Syntactic pattern of the word in a phrase and sentence.
    d) Meaning: referential (including multiplicity of meaning and metaphorical extensions of meaning), affective (the connotation of the word), and pragmatic (the suitability of the word in a particular situation).
    e) Lexical relations of the word with other words, such as synonymy, antonymy, hyponymy.
    f) Common collocations.

Laufer (1997b) seems to have neglected the importance of word frequency information mentioned in both the Richards and the Nation lists, whereas she has kept pragmatic appropriateness and the meaning modifying role of how the word may become combined with other words. In category d) she included referential meaning, but denotative meaning is not mentioned. She argues that knowing a word implies knowing all the above described features, which may be the case of an educated native speaker. However, it is important to point out, as we will see later in the third chapter, that language learners are often at various stages of knowledge about words before arriving at the final stage of knowing all the possible features described above (Henriksen, 1999).

Similar essential components of word knowledge are identified by Singleton (2000. pp. 161-162) based on observing what aspects of a profile of a word we need to be familiar with in order to be able to cope with it in ordinary language use. In his view, these aspects are the following:

> (1) knowing what it sounds like – so that we can recognize it and produce it in speech;
> (2) (at least in literate societies) knowing its written form – so that we can recognize it and produce it in writing;
> (3) knowing what it means – so that we can understand it and deploy it appropriately;
> (4) knowing how it behaves morphologically – so that we can recognize and use its different forms (singular, plural etc.);
> (5) knowing how it behaves syntactically – so that we can identify its function in phrases and sentences and so that we can use it in different roles in phrases and sentences;
> (6) knowing what 'company it keeps' – that is knowing about the various effects which accrue when the word participates in specific compounds, collocations, fixed expressions and idioms and knowing about the impact on its meaning and usage resulting from participating in such combinations;
> (7) knowing how its interpretation shifts in accordance with the different contexts in which it may occur;
> (8) knowing its social associations and knowing the kinds of social contexts in which it would and would not be likely to occur.
>
> (Singleton, 2000, pp. 161-162)

In this interpretation the first two categories refer to what Laufer (1997b) calls form. He also defines morphology, syntax and common combinations as essential

information about a word comprising Laufer's last three categories. However, the last two categories of contextual information and the possible social associations of the particular word add new dimensions to the study of word characteristics moving towards the domains of sociolinguistics. These aspects of word knowledge become especially relevant in the context of vocabulary testing, which is investigated further in Chapter 3.

Based on the four types of knowledge (lexical, encyclopaedic, generic and individual) distinguished by frame semantics, a field of linguistics emerging from the late 1970s (see Andor, 1985; Clark, 1992, pp. 35-38; Fillmore, 1985; Wierzbicka, 1996), Andor (1998, pp. 90-91) proposes that disctionary entries should include the following features of lexical items: (a) denotation-based content features, (b) referential features, (c) category features, (d) prominent syntactic markedness, (e) features of collocation range, (f) features of sub-categorization, (g) features of selection, and (h) features of frame knowledge.

So far we have seen how much a native speaker or a language learner needs to know about a word to be able to use it appropriately in the relevant context. We have yet to consider how words are stored and organized in the brain and what mental processes take place in language perception and production, i.e. the two components of vocabulary knowledge that Chapelle (1994) named the organization of the lexicon and fundamental vocabulary processes, respectively. Aitchison (1994a), Caron (1992), Cruise (1986), Garrod and Pickering (1999), Gleitman and Landau (1994), Grodzinsky, Shapiro, and Swinney (2000), Jackendoff (1990. 1993), Kiefer (2007), Levelt (1989), Langacker (1987, 1991), Palmer (1981), Pinker (1989, 1999, 2007), Robinson (2001b), Singleton (1999, 2000), and Skehan (1998) provide a comprehensive overview of the field and these issues are further discussed in section 1.5 of this chapter.

1.5 Models of lexical access in the mental lexicon

This human word store (Aitchison, 1994a), the mental lexicon, may be likened to a dictionary, but the differences between the two are multi-fold, involving content, the storage and the organization of words, as well as access to them (Martsa, 2007). As Aitchison (1994a) explains, dictionaries list words in neat, alphabetical order, while 'slips of the tongue' imply that besides word initial sounds or spelling, words in the mental lexicon are stored and looked up based on other features as well, involving the sound structure knowledge of the word, such as stress, vowel patterns and endings. An oft-cited metaphor of phonological access is sitting in a bath-tub. The so called 'bath-tub effect' implies a selective attention in the perception of words. Retention experiments have shown that word beginnings seem to be more prominent than word ends in the perception of sound differences, and less attention is paid to the middle segments (Aitchison, 1994a; Pinker, 1999, 2007).

The fact that speakers often confuse words with similar meanings suggests that the organization of the human mental lexicon is probably more complex than that of a printed dictionary, especially in aspects of content. While the content of a print dictionary, once published, is fixed and static, the human mind is capable of flexibly adding new items and altering meanings or pronunciation at any moment, and while looking up a word in a heavy dictionary may prove to be a tedious enterprise, a speaker is able to find a word in the mental lexicon in a fraction of a second. Considering all these factors, however, the main difference between a dictionary and the mental lexicon is in the amount of information stored about each entry (Aitchison, 1994a).

Other frequently quoted metaphors of a possible image of the mental lexicon are the London underground system, a spider's web, the structure of an atom, or a library (Aitchison, 1994a; Gósy, 1999; Pinker, 1999; Singleton, 1999, 2000). All models of the mind, however, have one thing in common: "they are simplified diagrams which encapsulate crucial features of something that is in reality considerably more complex"

(Aitchison, 1994a, p. 36).

Researchers working on the borderlines of psychology, neuroscience and linguistics have developed several models on how words are represented in, accessed and retrieved from the mental lexicon (for overviews see Aitchison, 1994a; Emmorey & Fromkin, 1988; Garman, 1990; Garrod & Pickering, 1999; Jackendoff, 1993; Levelt, 1989; Pinker, 2007; Reeves, Hirsh-Pasek & Golinkoff, 1998; Robinson, 2001b; Singleton 1999, 2000; Tannenhaus, 1988; Wray, 2002). The models discussed below address issues of what components the mental lexicon may comprise, as well as how these components relate to and co-operate with one another. The general distinction between direct and indirect models of lexical representation and processing may be pictured with two easily conceivable metaphors, explains Garman (1990). He portrays direct models as working like a word processor software: to find a word in the computer database we need to type in as many letters of it as it is sufficient to distinguish it from other stored items.

On the other hand, an indirect model may be likened to looking up a word in a dictionary or a book in a library. This process involves more than a single step to find the required item. Representatives of the direct model are Morton's (1982) logogen model and Marslen-Wilson's (1987) cohort model, whereas Forster's (1976) search model exemplifies the indirect type. In the following section I am going to discuss these models briefly, before addressing issues raised by more recent approaches (Fodor, 1983; Jackendoff, 1993; Levelt, 1989; Stubbs, 2001; Wray, 2002).

1.5.1 Morton's logogen model

The logogen model developed and later revised by the British psychologist John Morton in the 1960s and 1970s is based on the observed link between word recognition and word context in sentence completion tasks (Morton, 1982). The model attempts to explain why words are better recognised and retrieved in certain contexts than in others. The word 'logogen' is coined from the Greek 'logos' meaning *word*, and

'genesis' meaning *birth* or *coming to life* from the Latin word 'generare' (Singleton, 1999, p. 85). A logogen in Morton's model is a 'neural unit' in the nervous system where an 'event' takes place when "a lexical response becomes available" (Singleton, 1999, p. 85). The logogen system thus is a set of mechanisms present for each word in an individual's mental lexicon, i.e. there is a separate logogen for every word in the lexicon (Caron, 1992, p. 53).

There is a separate auditory logogen system for the analysis of perceptual acoustic input (what we hear), and a visual logogen system for analysing perceptual visual input (what we see). Both are connected in a two-way link to the cognitive system which, in Morton's view, is a database of various kinds of semantic information including information on contextual probabilities (Singleton, 2000, p. 171). The scattered uni-directional line from the auditory and visual systems refers to possible cases of producing output without consulting the cognitive system for semantic information, which, as Singleton (1999) points out, accounts for being able to pronounce non-word input presented either visually or auditorily. The response buffer, the fourth component uni-directionally connected to the logogen output system, then generates spoken or written production in response to the input. There are two thresholds regulating access to both the cognitive system and the response buffer after which the logogen 'fires' (i.e. a word is available as a response). When the threshold level of activation is reached and once fired, the level of activation decays gradually but never reaches the original state again. This, on the one hand, seems to explain why a frequently used logogen is never completely deactivated again (frequency effect, Caron, 1992, p. 53), as well as the effects of long-term semantic priming (Singleton, 1999, pp. 87-88), i.e. the activation of a logogen also entails the partial activation of other logogens close to it.

Citing evidence both from psycholinguistic experiments with students and data gained from aphasic (brain-impaired) patients, Emmorey and Fromkin (1988) assume a strong connection between the phonological and orthographic representation of words, which the logogen model does not allow. They claim that the so-called Fromkin or Modular

coaddressing model is capable of bridging this gap by proposing separate but interconnected lexicons for orthographic and phonological representation, supplemented by a semantic lexicon linked to both.

Other criticism of the logogen model pointed out that it does not account for picture recognition and naming, which would necessitate the insertion of a "pictogen system" (Singleton, 1999, p. 90), and on the other hand, the output system needs to be more complex, with distinct pathways for written, spoken and graphic output. The difficulty of defining the threshold levels of activation and the lack of evidence on why less frequent words may also be activated instead of more frequent items in the system have both implied a need for new attempts in modelling the mental lexicon to account for these phenomena.

1.5.2 Marslen-Wilson's cohort model

Moving away from the perspective of the dictionary metaphor of previous models, Marslen-Wilson (1999) proposes the Distributed Cohort Model of lexical processing. The word 'cohort' refers to a group of warriors in an ancient Roman legion and serves as a metaphor for Marslen-Wilson' model (1987). It postulates the existence of an auditory word recognition system in which as a reaction to the auditory input all the words in the mental lexicon that posses the same word-initial group of sounds become activated, then restricted step by step to the only word which matches the sound pattern of the input signal by dropping the mismatching candidates progressively. This process assumes the existence of a 'uniqueness point', the precisely identifiable point where the word is recognized as different from other members of the cohort, and identical with the input signal or non-words recognized as not matching any of the words in the mental lexicon.

Singleton (1999, p. 92) illustrates this process with the word *elephant* ['el<u>i</u>f nt]. He explains that the word-initial cohort ['el<u>i</u>] for this word may include words like *elevate* or *element*, but not *elephantine* because of the word-initial stress. The point of

uniqueness is presumably the recognition of the [f] sound as there is no other word in the English language which would begin with the sound [ˈelif]. Experiments on the recognition time of words have shown that recognition time is shorter if the uniqueness point comes early in the word and becomes longer if that point comes late (Wurm, Ernestus, Schreuder, & Baayen, 2006).

Furthermore, Marslen-Wilson (1987) assumes that context plays an important role in word recognition and facilitates semantic processing. He claims that once the word initial sound cohort is established, semantico-pragmatic contextual information determines which members of the cohort are deactivated based on contextual mismatching. Emmorey and Fromkin (1988) cite evidence both for and against the fact that the beginnings of words are responsible for being easily accessible, but they add that their implications for the cohort model are still unclear.

Aitchison (1994a, pp. 217-218) points out that in this model a lot more words are activated than necessary, speakers need to use all kinds of information available to restrict the cohort to one word only, and they have to make the decision very fast, while the words are being uttered. As for the drawbacks of the model she mentions the rigid early version which is unable to cope with distorted word initial acoustic signals and claims that if a wrong decision is made, the wrong cohort may be activated. However, she refers to a more recent version of the model as being more flexible and overlapping with interactive activation models.

1.5.3 Forster's serial search model

The widely discussed serial search model was developed by the psychologist, Kenneth I. Forster (1976); according to it the processes of lexical access in the human mind are best conceivable as similar to looking for a book in a library or a system of files on a bookshelf (Caron, 1992, p. 51). When we need a specific book, usually we only have fragments of information about it, either the author's name or the title, sometimes both, or one with the year of publication or the name of the publisher. We then begin our

search in the catalogue of the library, and after finding the relevant specification data we go to the shelves and browse the shelf-marks until we find the book in question.

According to the search model, lexical access is not different from a library in this respect (Forster, 2006). We begin our search from various starting points according to what kind of information is available to us in the given situation, i.e. the phonological or the orthographic form, the morphological, syntactic or semantic characteristics of the word. This initial search takes place in the peripheral access files organized along one of these characteristics, corresponding to the different library catalogues. These access files contain lists of entries equipped with pointers (shelf-marks) directing to a master file, which can be envisaged as a collection of words containing cross-references among words interrelated by aspects of meaning.

On the effect of word category information on lexical activation, Forster (2006, p. 35) explains that according to a cascaded activation model, activation spreads from "one level to another without waiting for resolution at the initial level" and a word should initially activate the semantic properties of its neighbours. In his discussion he cites evidence from studies on word recognition tasks where a word or non-word similar to the exemplar words proved to take longer to reject (e.g., *turple - turtle*). Singleton (1999, 2000), however, points out that experimental studies have not shown without doubt that the mental lexicon works this way.

Emmorey and Fromkin (1988) explain that both the logogen model and the cohort model are interactive models of lexical processing as word recognition takes place on the basis of an interaction between the sensory input and semantic/syntactic contextual information, whereas Forster's search model assumes autonomous processes where contextual information is dealt with after processing the sensory input. However, neither the logogen nor the cohort model seem to account for the processing of non-linguistic information, while in Forster's search model this aspect is not neglected.

Although the above discussed models may differ significantly in their assumptions

about the structure of the mental lexicon and the processes involved in lexical access, they have one characteristic feature in common: their focus is entirely the mental lexicon. The models to be discussed in the following section seek to cover a wider realm of mental processes, where lexical processing is handled as a sub-dimension of a larger cognitive domain.

1.5.4 Levelt's blueprint model

Levelt's blueprint model of language production, often quoted as 'blueprint for the speaker' differs from the models discussed in that it addresses the aspects of language processing from perception to production. Levelt (1989, p. 181) claims that formulation processes are lexically driven and identifies two major components of the model divisible into several subcomponents: a declarative component refers to the 'knowledge that' (facts we know about the world and language), and the procedural component, the 'knowledge how', responsible for information about how to achieve specific goals with language.

The type of declarative knowledge required for language use includes information about the world (encyclopedia), information about situations (situational knowledge), and information about what style is appropriate in various circumstances (discourse model). Declarative knowledge also involves the lexicon having two major parts: lemmas and lexemes (Levelt, 1989) or forms (Singleton, 1999), i.e. semantico-grammatical and morphological information, respectively. As Singleton (1999) explains, lexical search thus takes place in a two-staged process, making Levelt's model comparable to Forster's search model.

As for procedural knowledge, the model has several subcomponents. The Conceptualizer is responsible for generating messages, micro-planning and monitoring the output. The Formulator gives the pre-verbal message a surface syntactic and phonological shape. The third component is named the Articulator responsible for speech based on the phonetic plan coming from the Formulator (Levelt, 1989), while

the Audition system analyses the input speech sounds. Finally, the Speech comprehension system makes sense of the phonetic information received.

In this model the lexicon has a central role in speech production and is envisaged as being linked to both the formulator and the speech production system. According to Levelt's lexical hypothesis (1989, p. 181), "the lexicon is an essential mediator between conceptualization and grammatical-phonological formulation, supporting the much debated concept of interpenetrating lexis and grammar." It entails, he explains, that nothing in a message triggers a syntactic form; there must be a mediating lexical item equipped with grammatical properties that generates a particular syntactic structure. Singleton (1999), however, highlights some of the problematic assumptions of the model. These problem areas, he points out, are the purely declarative categorization of the lexicon, its separation from encyclopaedic knowledge and the question of the degree of autonomy the various components of the model possess.

1.5.5 Modularity and lexical processing

Prominent researchers propose that the human mind is modular in its functions and postulate the existence of a language module; however, their views about the content of the module overlap and oppose one another at the same time. Noam Chomsky (1957, 1968) is often referred to as the father of investigating language as mental representations and rules, his generative grammar shifting the focus of language study from external language (E-language), i.e. language performance, to internal (I-language), that is the "states of mind/brain that enter into behaviour" (Carston, 1990, p. 38), or the study of I-language "the study of a language as part of human biology, trying to find out what it is that each individual has, that enables that individual to participate in larger social interaction, or to perform discourse" (Andor, 2004, p. 95) and calls E-language everything else, that is not I-language.

I-language comprises theoretical vocabulary and principles that "are defined independently of other cognitive systems" (Carston, 1990, p. 40) and in this

conception the theoretical constructs of linguistic theory denote real mental entities. This view is challenged by instrumentalists who claim that linguistic concepts are not mental primitives; therefore, grammars are not representations of something real and distinct, as Carston (1990) argues. She distinguishes autonomous mental systems and claims that language is one of them.

Chomsky (1968) proposes that there are distinct regions or faculties for deductive reasoning, arithmetics, problem solving, scientific theory formation and language. These input systems (input, as defined by Carston, 1990, p. 42, based on Fodor's (1983) views, is a representation of a proximal stimulus) share certain qualities common to all modular structures, such as sensitivity only to a specific set of stimuli coming from the environment, format of representations specific to the system, a database and a set of principles directing the system, fastness and automaticity, limited access to data. Chomsky interviewed by Andor (2004) claims that language processing is similar in its modules to vision.

1.5.5.1 Fodorian modularity

The American cognitive linguist and philosopher, Jerry Fodor, in his theses on the modularity of language acquisition (1983) maintains that the human brain is modular, having distinct areas for vision and audition, motor functions and language. But whereas Chomsky investigates the modularity of language acquisition in connection with a language acquisition device (LAD), Fodor essentially focuses on processing language. He claims that both input systems and central systems (or the general cognitive system) are computational, the latter being modality and domain neutral and independent. He addresses issues such as the modularity of the modules, the existence of cross-modular connections, the relationship between the input and output functions of the modules, and the relationship between competence and performance. He posits that processes of the language module are domain specific, mandatory, inaccessible to consciousness, rapid, and have their own neural hardwiring, failure or breakdown patterns (as in aphasia) and specific developmental sequences.

As for the structure of the mental lexicon, Fodor (1983) assumes that it resembles a graph, where lexical items are nodes and are interconnected with other nodes in the lexical network, activated by the spread of excitation (stimulus). Excitation thresholds are lowered for the related nodes in the spread of excitation, thus resulting in decreased response times for connected items in lexical decision tasks.

He is particularly concerned with the question that the modules are informationally encapsulated, a cornerstone of his model, implying that general knowledge about the world or contextual information do not play a role in the operation of the module while processing is taking place, they only interact when the operation of a particular module is completed. In order to fend off criticism on the part of psycholinguists referring to results of experiments on sentence completion tasks (cloze procedures), Fodor limited his model to the concept of the language module as being a formal processor only, with no semantic role. However, Singleton (1999, 2000) argues, a more plausible position is necessary to explain context effects in word recognition, and as Asher and Pustejovsky (2000, p. 2) point out, the Fodorian atomistic lexicon fails to explain data about the interaction between pragmatics and semantics.

1.5.5.2 Jackendoff's representational modularity

Challenging Fodor's modularity hypothesis, often referred to as F-modularity, Jackendoff (2000) proposes a variant called the representational modularity of the mind, which rejects Fodor's assumptions of informational encapsulation and domain specificity. He argues that without interface modules mediating between the modules and serving as links making communication between the separate modules possible, the various modules of the mind would be functionally disconnected, thus the perception of the world and behaving accordingly would be impossible.

As it appears from the above, Jackendoff (2000) is not completely opposed to the idea of modularity, he only proposes a distinction between two kinds of modules as

necessary innovations to the original idea of F-modularity: integrative modules (responsible for integrating levels of representation) and interface modules (making communication among representations possible). Jackendoff claims that a step missing in Fodor's model is the possibility of converting one format of information into another, for instance, in speech perception the conversion of the perceived syntactic structure into semantic information. Without such interface modules, he claims, lip reading, reading texts, and understanding signed languages would not be possible.

Jackendoff (2000) describes three types of mental processes. First, processes, when a full representation in a certain format is created from fragments of structures (e.g., a syntactic parser, lexical items organized into a full syntactic structure, a sentence) are called integrative processes. Second, translation, or with a later term, interface processes convert one form of mental representation into another (e.g., the acoustical information of a speech signal into a phonetic representation; or a syntactic structure into propositional structure). Third, inferential processes compare full representation with each other or construct new representations in the same format (e.g., comparing two phonological representations to see if two words rhyme). Jackendoff (2000, p. 13) concludes that "the locus of modularity is not large-scale faculties such as language perception (Fodor's view), but at the scale of individual integrative, interface, and inferential processors".

Jackendoff (2000) abandons the idea of Chomskyan syntax-centred generative grammar and claims that syntax is but one of several generative components, and further clarifies the role of the lexicon. He argues against the Chomskyan view that lexical items are inserted into initial syntactic derivations, and then interpreted semantically and phonologically through processes of derivations. As an alternative he claims that in the process of perception auditory information is processed by the auditory-to-phonology interface module to create a phonological representation. Then, the phonology-to-syntax interface creates a syntactic structure, which is then, aided by the syntax-to-semantics interface module, converted into a propositional structure, i.e. meaning. For that reason, when a lexical item becomes activated, it does not only

activate its phonology, but also its syntax and semantics and thus "establishes partial structures in those domains" (Jackendoff, 2000, p. 25). The same but reversed process takes place in language production.

1.5.6 Connectionism (parallel distributed processing)

Parallel and serial models of processing differ in the number of analyses possible at the same time (Pickering, 1999). While in a serial model (e.g., Forster, 1976) one analysis is selected, a parallel model considers multiple analyses at the same time. An influential model of lexical processing known as connectionism, or in an alternative term parallel distributed processing, unlike the modularity hypothesis, postulates that different pieces of information are processed simultaneously (in parallel), independently from one another and on different levels (distributed). This assumption stands in opposition with Forster's serial search model, where the stages of operations build on and depend on one another. According to this model, the operations of language processing progress independently and envisage a "high degree of interactivity between semantic and formal processing" (Singleton, 2000, p. 179).

Connectionists relish the 'brain' metaphor to language processing and challenge the Chomskyan/Fodorian view of the mental symbol paradigm (Colombo, Stoianov, Pasini, & Zorzi, 2006; MacWhinney & Leinbach, 1991; Plunkett & Marchman, 1993; Rummelhart & McClelland, 1986). They not only claim that mental processes do not involve operations with symbols, but also propose that knowledge is represented in terms of synaptic connections and connection strength determines activation rather than rules or patterns. They propose that signals may not only be excitatory, but also inhibitory, and the spread of activation is bidirectional, moving forward and backwards. Thus, learning is seen as a by-product of information processing as a result of the strengthening of associations among units (Ellis, 1994). Singleton (2000) explains that this view has generated fruitful debates on the issue and connectionism has now become more influential, taking account of semantic aspects of language processing in the brain.

A connectionist model inspired by the Marslen-Wilson cohort model is the interactive Trace Model proposed by Ellman and McClelland (1986). As Caron (1992, p. 57) points out, interestingly this model is able to account for word identification even in cases when the input is distorted or incomplete, as well as for the majority of the data gained from speech production research.

1.6 Towards an integrated model of the lexicon

As it appears from the above discussion, the mental lexicon is complex and its components are closely interrelated and interdependent. Experiments with patients who suffer brain injuries (aphasics) seem to supply evidence for the claims of the above discussed theoretical models by providing and justifying a basis of comparison between 'normal' and 'impaired' language processing. Research into neurolinguistics and psycholinguistics is an especially complex endeavour which far exceeds the focus of the present enterprise into vocabulary testing; thus, only some focal issues are sketched here. The majority of these studies appear to discuss issues of the structure of the language system (Jackendoff, 2000; Garrett, 2000; Pinker, 2007; Turvey & Moreno, 2006; Wray, 2002), the interface of language comprehension and production (Hickok, 2000; Nicole & Love, 2000), lexis and structure (Blumstein & Milberg, 2000; Caramazza, 2000; Feldman, Basnight-Brown & Pastizzo, 2006; Stockall & Marantz, 2006; Swinney, Prather, & Love 2000) and syntax and discourse (Avrutin, 2000; Caplan, 2000; Piñango, 2000). In the next section I am going to touch upon some major issues on how words are stored in the mental lexicon, before moving on towards an integrated model developed by Wray (2002).

1.6.1 The organization of the lexicon

Investigating lexical learning Skehan (1998) assumes that language is more lexical than usually accepted. He claims that the rule-based approach to language "is an imposition of the linguist, and may not always be justified" (Skehan, 1998, p. 31).

This new view challenged the influential views of Chomsky that linguistic competence primarily involves the knowledge of grammatical rules which allow the language user to produce an infinite number of utterances.

However, Skehan (1998) highlights some advantages to the rule-based system: the underlying lexical elements or units need not be represented more than once in the brain, they only need to be well organised according the grammatical rules and "looked up wherever they are kept" (p. 30). It follows that the storage system, which is meant to underlie the processing system, can be as small as possible. However, as he points out, there is no convincing evidence why the human memory system would avoid duplication of storage of lexical items in the brain.

Early models of how meaning is stored and organized in memory were thought of in two ways: as a network or a set of features stored with each word (Quillian, 1967; and Smith, Shoben, & Rips, 1974, respectively). Caron (1992, p. 73) explains that these two models only differ in their formulation and both postulate the principle of economy of storage i.e. "each piece of information appears only once in the network". This was later rejected by Collins and Loftus (1975) who argued for a direct connection between a concept and all the information acquired related to it. Other more complex models of mental activity presume the existence of semantic networks for the storage of meaning (Rumelhart, Lindsay, & Norman, 1972).

Based on corpus evidence referring to connotations, Stubbs (2001) has shown that the meanings of words are often not captured by their dictionary definitions, but defined in terms of logical relations. Earlier Sinclair (1991) expressed similar views and claimed that most combinatory possibilities of grammatical rules are ignored in real-life language production and lexical items occur again and again in different locations in the brain based on these logical relations. He proposes the open-choice principle and the idiom principle to account for multiple storage. The open-choice principle, he explains, is the capacity to use and understand unconstrained numbers of combinations of words. The idiom principle claims that the co-occurrences of words are limited,

many frequent words become delexicalised as they enter into frequent collocations and phrases, and that the idiom principle takes precedence.

Tannen (1989) extended the analysis and focused on the area of conversation, emphasising the preference of language users of formulaic language and repetition to create a frame to new information. Bolinger (cited in Skehan, 1998, p. 34) coined the term 'item-bundles' to refer to multiple representations of lexical items, supporting the view that the same word is stored more than once in the brain, adding that these multiple representations are likely to be the combinations of a base word to form ready-made expressions. This way, he proposes, the memory system is "organised not for efficient compactness, but for ease of use" (Skehan, 1998, p. 34).

1.6.2 Wray's Heteromorphic Distributed Lexicon

As a proposed solution to the vastly debated questions of multiple representation and the storage of words that break the rules in one sense or another, Allison Wray (2002) proposed a single rule-based system for language processing, where regularities in language are more easily explained than irregularities. Within her dual-systems model language is seen to be processed both holistically and analytically, which is more liable to explain problems raised by formulaic sequences (as discussed earlier in section 1.1 of this chapter) and idiomaticity in language.

Her combined model, the Heteromorphic Distributed Lexicon, proposes the existence of five lexicons she named grammatical, referential, interactional, memorized and reflective, each consisting of three holistic units of various distributions and sizes: the morpheme, the formulaic word, and the formulaic word string units (Wray, 2002, pp. 262-265). She emphasises that the unit types are not discrete; units can hover between the levels, and claims that "even if a string is segmented and one or more of its component parts are separately stored, it may also continue to be stored holistically" (Wray, 2002, p. 262). Therefore, an idiom such as *Look out!* may be stored as *look* in the morphemic unit of Lexicon II (referential) and as *out* in the morphemic unit of

Lexicon I (grammatical), while it also permits *Look out!* as a string to be stored as one unit holistically in the formulaic word strings unit of Lexicon III (interactional), meaning *be careful*. This is referred to as 'compositional' versus 'noncompositional' processing in research literature.

This three-layered representation of each lexicon as units of morpheme, formulaic word and formulaic word string is exceptional, Wray (2002) explains, for it displays all the linguistic units that are not subject to further segmentation, thus it should be handled as one holistic unit. Therefore, this model is able to account for those lexical patterns that previous models struggle with. Her model also rejects the theory of economy of storage outlined earlier and proposes that units gain entry to the lexicon "not by virtue, but as a result of pure expediency" (Wray, 2002, p. 267), i.e. we only store things we have a use for, be it morphemes, words, phrases or whole texts. This assumption implies that the nature of the lexicon is determined "not by structural principles which decide whether an item is simple enough to be stored" (Wray, 2002, p. 268), but by what priorities an individual assigns to certain linguistic input, making every lexicon personal and different.

1.7 Conclusion

This chapter has introduced the context, the terminology and the major theoretical issues of vocabulary research. We have seen that the study of vocabulary has grown into a vastly debated field, which can by no means be called neglected today. The place of vocabulary has been discussed in relation to other aspects of communicative language competence and it has been highlighted that although vocabulary is generally seen as part of grammatical competence, several other layers of competence require various aspects of word knowledge in certain respects.

I have covered the main stages of the on-going debate about what a word may be and what layers knowledge of a word may have. The definitions overviewed in these sections differed in how restricted or wide a view of the observed paradigm the

researcher opts to take. The abundance of technical terms used in vocabulary research has been outlined briefly, consciously restricting the choice to the major notions necessary to understand the first chapter, as the following chapters are going to add further colours and relish to the spectrum and refine the scope of analysis.

In this chapter I have discussed some major models of the mental lexicon and lexical processing in the brain in order to cast light on the complexity of the field creating the ground for vocabulary acquisition and testing underlying the assumptions of the empirical chapters. I have touched upon the logogen, the cohort, and the serial search models, as well as some models of higher cognitive processes, such as the blueprint for the speaker, the modularity of the mind and some relevant aspects of connectionism.

An outline of some studies on how words may be stored in the brain has been given in the final section where the use of the conditional is justified by the scarcity of our present understanding of how the human brain works. The next chapter is going to continue this line of thought by discussing the underpinnings of first - and second - language acquisition (L1/L2), as well as how words are used, involving the interface between vocabulary and the four skills and the role of corpus linguistics in the study of vocabulary-related phenomena.

Chapter 2

How is vocabulary learnt and used?

2.1 Introduction

The expansion of the mental lexicon and the learning of new vocabulary is such an all-pervasive process that it would be hard to describe all possible contexts and variety of stimuli. First L1 words are learnt at the mother's knee, later on in the extended family, then the process continues in all educational contexts and in every aspect of life. Similary, an L2 may be acquired in naturalistic or formal instructional settings. As there can hardly be any situations which do not offer possibilities for vocabulary learning, it is not surprising that for the moment it seems there is no unified theory of vocabulary acquisition for all possible contexts.

Four major issues in second language vocabulary acquisition are discussed in the following sections. The first part examines the mechanism of vocabulary learning: the role of cognitive factors, the L1, word form and memory. The second part provides an overview of the scope of vocabulary learning, introducing the reader into research on how much vocabulary is necessary for learners for various purposes (i.e. the breadth of vocabulary knowledge) and how well learners are expected to know the words (i.e. the depth of vocabulary knowledge). The third section explores differences in intentional and incidental learning, considering questions of how much and what type of vocabulary is necessary for it to take place, the role of strategies applied and the amount of exposure required, what types of texts are more conducive to incidental word learning and what the role of input modification may be in the process. Finally, some pedagogical implications are addressed in the last section of the chapter.

2.2 The mechanism of vocabulary learning

Approaches to second language acquisition (SLA) involve three tightly intertwined and interrelated areas: the representation, acquisition and processing of language (for introductions to the field see Aitchison, 1994b; Bardovi-Harlig, 2002; Bialystok, 2005; de Groot & van Hell, 2005; Ellis, 1994; McLaughlin & Robbins, 1999; Meara, 1999; Milroy, 1994; Oxford, 1999; Taylor, 1990). In consequence they can hardly be studied without one another and it is difficult to discuss them successively. The issues raised in the first section of this chapter touch upon such vast fields as neurolinguistics and the study of bilingualism, the thorough discussion of which is beyond the scope, aims and length constraints of the present work with a focus on vocabulary assessment. Therefore, the questions and studies introduced here are necessarily highly selective and restricted, with the aim of distilling a number of issues that arise later in the empirical chapters.

2.2.1 Cognitive factors in the mechanism of vocabulary learning

There are two main reasons why a language teacher or a language scholar can take advantage of knowledge in the field of the neurosciences. First, certain linguistic data can only be accounted for by biological terms of processes of the brain. Second, language acquisition, language perception and production are cognitive skills based on biological principles and mechanisms. Knowing about these processes might help solve language and language teaching related problems. Therefore, in the following section let us briefly overview some aspects of neurolinguistic research that might cast light on how vocabulary, more specifically foreign language vocabulary, is acquired.

In the mid-1990s there was a revived interest among linguists in the neurobiological foundations of language acquisition and recent studies provide evidence for the difference in the mechanisms of acquiring and storing vocabulary and syntactic rules in the human brain. Hebb's law claims that nerve cells, neurons become more strongly associated when they are frequently active at the same time and the strengthening of

the synaptical connections between neurons represent the basis of learning in the neurobiological sense (Ellis, 1994). Thus, argue Pulvermüller and Schumann (1994), language learning also rests on synaptic strengthening and the cerebral cortex of the brain. More specifically, the so called Broca and Wernicke areas are responsible for language acquisition, and in the narrower sense, language production and perception, respectively (Taylor, 1990).

Based on electrophysiological evidence it is assumed that three cortical processes are required for language acquisition: (a) a bubbling process (in which syllable assemblies are formed), (b) content word acquisition (connections between neurons of various cortical areas are strengthened and form assemblies that correspond to various content words), and (c) acquisition of functors and grammar, "during which the co-occurrence of words is stored in the brain by means of connections between cell assemblies" (Pulvermüller & Schumann, 1994, p. 697). Pulvermüller and Schumann (1994) propose that these three processes follow each other in the process of language acquisition, but they considerably overlap. Blumstein and Milberg (2000) go even further and propose that a deficit in lexical activation may be responsible for impaired syntactic processing in Broca's and Wernicke's aphasics (damage of the two brain areas responsible for language perception and production) and assume the existence of a common mechanism for lexical activation, serving the entire vocabulary (see also Berthier, Green, Lara, Higueras, Barbancho, Dávila & Pulvermüller, 2009).

There is no doubt in the research today that the first language influences the acquisition of the second or foreign language to a considerable degree (Ellis, 1994; Schmitt, 2008). With the arising new techniques of neuroimaging in the 1990s, a heated debate in research attempted to find evidence for the difference in the location of regions corresponding to first and second language storage and processing in the brain. Pulvermüller and Schumann (1994) concluded that the perisylvian region is more essential in the acquisition of the L1, while regions further away from the sylvian fissure seem to be responsible for L2 processes. However, their findings exploded a sequence of argumentative studies on the subject (Eubank & Gregg, 1995;

Jacobs, 1995; Paradis, 1994; Pulvermüller, 1995; Schumann, 1995).

A revolutionary new neuroimaging technique, functional magnetic resonance imaging (fMRI) brought new impulse to neurolinguistic research. Studying the spatial distribution of native and second languages in the multilingual brain, Kim, Relkin, Lee and Hirsch (1997) found that late acquired second languages have separate locations in the Broca's area, while the Wernicke's area shows no separation based on the age of acquisition. Nikolov (2002) explains that before this study there was no direct evidence available for the topographic specialization of different languages in multilingual subjects.

Extensive research on bilingualism has contributed to the debate to a considerable degree, although it has to be noted that the interpretation of the term bi- or multilingual may be manifold (for a comprehensive introduction to the field see Kroll & de Groot, 2005; and Kaplan, 2002). The most enduring questions in studies on the bilingual mind are whether there is a single system for the representation and processing of the two (or more) languages, or perhaps bilinguals use separate systems for all the languages (Kroll & Sunderman, 2003; Murre, 2005); whether comprehension and production are served by one and the same or two different systems and how lexical access to the bilingual memory takes place (Heij, 2005; Kroll & Tokowicz, 2005; Myers-Scotton, 2005).

Numerous models have been set up to describe the bilingual memory (de Groot, 1992; Dijkstra & Van Heuven, 1998; Grainger & Dijkstra, 1992; Green, 1998; Kroll & Stewart, 1994; MacWhinney, 2005; Poulisse & Bongaerts, 1994). As Kroll and Tokowicz (2005) point out, contemporary models, as opposed to early models, have become more specialized and focus on one aspect of the linguistic code (for overviews see de Bot, 2002; Kroll & de Groot, 2005): they make assumptions about different levels of representation; address questions of language processing i.e. production or comprehension; and study developmental issues in bilingual language performance.

The age of acquisition has long been a crucial issue in language acquisition research (Ellis, 1994; Nikolov, 2002; Singleton, 1989). The widely discussed and debated Critical Period Hypothesis proposes the existence of a particularly sensitive period in human life before puberty, when language acquisition can take place more easily and faster than beyond this period. This assumption seems to be supported by the evidence found in neuroscience for the maturation of the language cortex in the brain. Pulvermüller and Schumann (1994) attempted to provide a brain-based model to explain the variable success achieved by early and late language learners. They found two basic conditions to be met in order to reach success in acquiring the grammatical rules of a language. One is being equipped with the ability to acquire grammatical knowledge (1994, p. 681), the other is motivation and positive evaluation (for overviews on the role of motivation in SLA see Dörnyei, 1998, 1999, 2001; Dörnyei & Csizér, 1998; Dörnyei, Csizér, & Németh, 2006; Ellis, 1994; Gardner & MacIntyre, 1992, 1993a, 1993b; Nikolov, 2002). As for the former, they explain that around puberty all cortical areas, with the exception of higher-order association cortices, have reached their full level of myelination, and neurons in the perisylvian language cortex suffer reduced plasticity around puberty; therefore, after this period language learning will be impaired.

Varying success in the case of early and late start might be explained by the difference in the rate of maturation of the various parts of the brain. Myelination and loss of plasticity take place early in the lower systems of the perisylvian language cortex, therefore, a late start of language learning causes impaired abilities in storing phonological and syntactic knowledge. There is empirical evidence in applied linguistic research that adults learn faster than children, but their ultimate attainment generally falls short of native speaker standards (DeKeyser & Larson-Hall, 2005). However, Nikolov (2002) found evidence against this strong version of the CPH by identifying successful adult learners who, despite a late start in learning English as a foreign language, were often mistaken for native speakers by native judges in listening tasks.

This might be explained by the fact that age effects are not entirely obvious in SLA as the learners already have a native language and "the language centres in the brain have been activated in the opportune window. Thus, it is more appropriate to consider age-related effects in L2A, not a critical cut-off point, after which it becomes impossible to achieve native-like proficiency", as Slabakova righteously points out in her review (2006, p. 303). Research on the critical period in SLA thus seem to imply that instruction needs to be adapted to the age of the learner and not that instruction should be started as early as possible as the "earlier the better" assumption suggests (DeKeyser & Larson-Hall, 2005).

However, Pulvermüller and Schumann (1994) assume that the maturation of the language cortex does not seem to affect the storage of semantic knowledge. Semantic knowledge is stored throughout the entire cortex, some parts of which mature late, making it possible to acquire semantic knowledge late in adulthood in spite of loss of plasticity of regions corresponding to grammatical information. This assumption seems to be supported by empirical evidence (Slabakova, 2006) showing that "there is no point before death that vocabulary acquisition can be predicted to cease" (Singleton, 1999, p. 42). Pulvermüller and Schumann (1994, p. 715) also point out that "late acquisition of words must lead to synaptic changes and assembly formation involving higher-order association cortices, because the perisylvian cortex is already mature", and this phenomenon affects the processing of a second language in bilinguals. Although a number of studies reported that early bilinguals are more likely to use the same regions of the brain for L1 and L2 processing (e.g. Kim, Relkin, Lee, & Hirsch, 1997), Abutalebi, Cappa and Perani (2005) concluded that the level of proficiency is a more influential factor than the age of acquisition. Highly proficient bilinguals tend to activate the same parts of the brain, whereas less proficient L2 users have distinct neural networks for the two languages. In respect of the lexicon, however, the age of acquisition seems to determine the meaning representations of words in that words learnt earlier in life have processing advantages over later acquired words (Morrison & Ellis, 1995).

2.2.2 The role of L1

There is indeed an abundance of materials available on L1 lexical acquisition (for an introductory survey see Aitchison, 1994a; Bloom, 2000; Sanford, 1999). It has been shown that the average American or British highschool graduate knows and uses about 60,000 words (Aitchison, 1994a; Pinker, 1999) which gets down to the impressive challenge of "learning one word every waking 90-minutes" (Bloom, 2000, p. 25). This is a rough estimate though and there may be considerable individual differences. Goulden, Nation and Read (1990, p. 341), for instance, found that university graduate adults, native speakers of English knew 13,200-20,700 words, while other studies with undergraduate students estimated a figure of 14,000-17,000 English words (D'Anna, Zechmeister, & Hall, 1991; Nusbaum, Pisoni, & Davis, 1984; Zechmeister, D'Anna, Hall, Paus, & Smith, 1993). On the other hand, as Bloom (2000) points out, word learning does not proceed at an even pace, it starts at about 12 months of age, and a three-year-old does not learn more than 10 words a week. The differences in estimations may root in the different interpretations of word knowledge discussed in the previous chapter.

Prior to the onset of word production a child has to face the challenge of discriminating speech sounds and developing concepts. The nativist tradition claims that this is made possible by innate structures (see Chomsky, 1968; Piaget, 1952; Pinker, 1999; Wong-Fillmore, 1991). As for productive lexical development, four developmental milestones are described in the literature on child language (Singleton, 1999). The first stage is (1) cooing, when the child attempts to produce vowel-like vocalizations from about the ages of one-four months. Then (2) babbling combines vowel-like and consonant-like sounds (e.g. *baba, mama*) from about the age of eight months. The third (3), so-called one-word-utterance stage is characterized by producing meaningful one-word utterances from about the end of the first year of a child's life, while in the fourth (4) stage, commencing round about 18-24 months, the infant is able to produce meaningful two-word utterances.

The process is further facilitated by the fact that adults tend to adjust their speech in lexis, syntax and pace: a register often referred to as 'caretaker-talk', 'motherese', 'baby-talk', or 'child-directed language' (Ellis, 1994). In a similar fashion, native speakers adjust their speech in communicating with non-native speakers, this is called 'foreigner talk' (Ellis, 1994). Acquiring an L2 learners develop an interlanguage, a term coined by Selinker (1972) to refer to both the internal system a learner has established at a certain point in time and the interrelated systems that characterize the learner's progress over time, i.e. the 'interlanguage continuum' (Ellis, 1994, p. 350; Tarone, 1999). It can best be conceptualized as "sets of loose lexical networks that are gradually organized into a system or systems" (Ellis, 1999, p. 465). Yet, it is a question to explore how these systems develop.

Studying the lexical aspects of the process, Ellis (1999) proposes that L2 learners extract syntactic categories from items that are implicitly acquired though exposure to input. He distinguishes two subsequent processes taking place. Item learning involves the acquisition of linguistic forms, a word or whole, unanalyzed chunks of language that are stored separately or in very weak connection to each other. System learning on the other hand, means the construction of abstract rules that govern whole sets of items, and their interrelations with other items and the abstract rules. His evidence for item learning comes from free variation observed in the interlanguage development of an L2. Free variation refers to the random use of two or more forms (words) to the same set of functions (Ellis, 1994), with regard to the same situational context, illocutionary meanings, linguistic and discourse context, and planning conditions (Ellis, 1999).

However, as Singleton (1999) explains, the major difference between L1 and L2 lexical acquisition is that in the latter case the learner already has experience of making connections between lexical forms and meanings in his/her L1. There is evidence that in the initial stages of L2 acquisition the learner assigns L1 meanings to the first encountered L2 word form (de Groot & van Hell, 2005; Jiang, 2000; Kroll & Stewart, 1994; Kroll & Tokowicz, 2005; Slabakova, 2006). Furthermore, the entry into

the classification system is enhanced by any existing cross-linguistic or cultural overlap, i.e. the transfer of the L1 conceptual world seems to aid L2 learning (MacWhinney, 2005). This is not all new to research, as we have seen in the previous chapter, connectionism predicts the same. The facilitating effect of L1 seems to be supported by the Parasitic Hypothesis formulated by Hall (2002) who examined English-Spanish cognate words to prove that on initial exposure to a word, learners automatically resort to already existing lexical items in their L1 or L2 to establish an initial mental representation.

Hall's (2002, p. 72) parasitic strategy of vocabulary acquisition postulates a "series of automatic, unconscious cognitive stages" that a lexical item has to undergo from the first encounter onwards. First, a form representation is established, then learners immediately identify an available translation equivalent in the first language by any medium (e.g., direct translation, definition, some icon or contextual cues). He claims that although learners connect novel L2 words to the frame representation (lemma) of a translation equivalent, thus the identification of cognates may lead to errors, the exploitation of the L1 lexicon for L2 learning in such a parasitic fashion is in full accord with general principles of mental representation discussed earlier in the previous chapter.

A similar view on the role of the first language form is introduced in an earlier study on second language lexical representation by Jiang (2000) examining how L2 lexical entries evolve in instructional settings. Based on Levelt's (1989) model of the internal structure of L1 lexical entries, where morphological, phonological and orthographic information constitute a lexeme, and semantic and syntactic information make up a lemma, she proposes a three-stage process. First, in the formal stage, the formal specifications of a lexical entry are established. Then, in the L1 lemma mediation stage the L1 lemma information is copied onto the L2 entry, thus the L1 lemma mediates L2 word use. Third, in the L2 integration stage the L2 semantic, syntactic, morphological specifications are integrated into the L2 lexical entry. Jiang (2000) argues that the majority of L2 words become fossilized in stage two and never reach stage three;

therefore, an L2 lexical entry has three unique features: it consists of an L2 lexeme and a L1 lemma, contains little morphological information, and the L2 word – concept link is very weak.

It seems clear from the above studies that information about lexical form is active in both languages in bilingual word recognition. Kroll and Sunderman (2003) note though, that these findings need to be interpreted with caution, as much of the research has been limited to pictured object naming, restricting the scope of semantics to concrete nouns only. How does lexical selection for production then take place if both languages are activated? Comprehension and production is language non-selective (Kroll & Dijkstra, 2002). Studies on interpreters' bilingual speech production (Christoffels & de Groot, 2005) found that although the activation flow is language non-specific, lexical selection may become language specific in highly proficient bilinguals. However, empirical data are more controversial in this respect and the mechanism of filtering is not clear yet.

In most computational language production models two subsequent selection processes are described. First, a relevant concept is selected; then, in a lexical selection process the word form to be used is selected from the competing activated set of lexical representations, argues Heij (2005). Challenging Green's (1998) inhibitory control model he proposes a model in which the preverbal message, and only that, contains all the necessary cues to arrive at the correct word; therefore, there is no need for selective activation or inhibition mechanisms. This can be seen as the extension of Levelt's (1989) model discussed earlier to the bilingual situation.

As for the form-meaning influence of the L1, others think that in L2 production the L1 translation of the L2 form is activated first and only then is the L2 form retrieved and produced (de Groot & van Hell, 2005), similarly to the process described in word perception. With increasing L2 proficiency, however, the L2 form is gradually "freed", becomes functionally detached from the L1 representation and its meaning is accessed as directly as that of the corresponding L1 word is (de Groot & van Hell, 2005, p. 21;

de Groot & Poot, 1997). In sum, to be able to use an L2 word in a native-like way, the attached L1 meaning has to be narrowed to get rid of unique L1 meanings, and the unique L2 meanings have to be extended and fine-tuned to capture all its possible connotations (de Groot & van Hell, 2005). This is most likely to happen in immersion settings (Erben, 2004, Wode, 1999) or through extensive reading (Krashen, 2004).

2.2.3 The role of lexical features

Besides the above described interlexical factors deriving from the relationships among words in the L2 and L1, intralexical features inherent in the word itself may influence lexical acquisition to a considerable degree. With regard to the formal aspects of word knowledge Nation (1990) outlined three skills that may enhance vocabulary acquisition: 1) being able to break the new word into parts (affixes and roots); 2) knowing the meanings of the parts; and 3) being able to make a connection between the meaning of the part and the meaning of the word (Nation, 1990, pp. 168-174).

Even though mention has already been made in the previous chapter of views of the structure of the mental lexicon, in order to clarify the importance of the above morphological factors, it is necessary to make a short detour here into the morphologist's perception of the mental lexicon. Chomsky's lexicalist hypothesis (1970) claims that the formation of nouns from adjectives and verbs (i.e. nominalization) is a morphological process and nominalized items (e.g., happi-ness) are listed as such in the mental lexicon. An alternative view proposed by Halle (1973) claims that the morpho-syntactic properties of a word are determined by its syntactic context. His mental lexicon therefore accommodates a) the list of morphemes, i.e. bound and free roots and affixes; b) the dictionary: a list of possible words in the language; and c) the filter: a list of possible but for some reason never occurring words, explains Martsa (2007).

According to Martsa (2007), two more recent and widely accepted views challenging Chomsky's and Halle's concepts are represented by Lieber (1980) and Butterworth

(1983, cited in Martsa, 2007). In Lieber's views (1980), the mental lexicon comprises morphemes (indivisible morphological units with grammatical and/or lexical meanings) and rules determining how to create more complex lexemes by linking morphemes. It thus follows that complex words are not listed in the mental lexicon, but are accessed through a procedure called morphological parsing, i.e. decomposing the words into constituting morphemes. The Full Listing Hypothesis proposed by Butterworth (1983, cited in Martsa, 2007) in contrast claims that all word forms, along with complex words, are listed in the mental lexicon, presupposing a holistic representation. However, explains Martsa (2007), this does not exclude the possibility of morphological parsing. If we accept then that a possible way of decoding the meaning of complex words may be morphological parsing, it is obvious that learners can inevitably make good use of the skills named by Nation (1990) enabling them to analyze word structures in the process of L2 word learning.

In a plight to find out what in an L2 word may make it hard or easy to learn, Laufer (1997b) examined the effect of the following factors on vocabulary acquisition: pronounceability (phonemes, combinations of phonemes, stress); orthography; length; morphology: inflectional and derivational complexity; deceptive morphological transparency; synformity; part of speech; abstractness and specificity/register restrictions; idiomaticity and multiplicity of meanings. Table 4 outlines the effect of these on L2 word learning.

It is interesting to note here that in contrast with the general assumption that abstract words are more difficult than concrete words, in her view concreteness/abstractness has no clear influence on L2 word learning. This she explains by the fact that second language learners have already developed abstract concepts in their L1, "why then should an abstract L2 word like *love* be more difficult to understand and remember than a concrete L2 word like b*ook*?" (Laufer, 1997b, p. 150).

Table 4. Intralexical factors affecting L2 vocabulary learning (Laufer, 1997b, p. 154)

Facilitating factors	Difficulty-inducing factors	Factors with no clear effects
familiar phonemes	presence of foreign phonemes	
phonotactic regularity	phonotactic irregularity	
fixed stress	variable stress and vowel change	
consistency of sound-script relationship	incongruency in sound-script relationship	
		word length
inflectional regularity	inflectional complexity	
derivational regularity	derivational complexity	
morphological transparency	deceptive morphological transparency	
	synformy	
		part of speech
		concreteness /abstractness
generality	specificity	
register neutrality	register restrictions	
	idiomaticity	
one form for one meaning	one form for several meanings	

Empirical evidence from translation experiments, on the other hand, shows that besides word form, word concreteness and word frequency are also influential in vocabulary acquisition. Based on empirical results there has been more overlap found in translation speed for concrete than abstract, cognate than non-cognate words, as well as for less ambiguous translation equivalents and high frequency words (de Groot & Keijzer, 2000). Therefore, de Groot and her colleagues proposed a Distributed Feature Model (van Hell & de Groot, 1998) for bilingual semantic representation. In this model concepts are differentiated in the degree to which the languages share types of semantic features.

2.2.4 The role of memory

For processing and storing knowledge the human mind uses short-term (STM) and long-term memory (LTM). Information in STM, a temporary potentiation of neuron connections lasting only for a few seconds, can become LTM by rehearsal and repeated association. It is assumed that different types of knowledge are stored in different parts of the brain. In Anderson's (1984, cited in N. Ellis, 1997, p. 123)

general model of memory, LTM comprises a declarative memory, where knowledge is represented as propositions, schemata and propositional networks, and a procedural memory interpreting and using this knowledge. Declarative memory is further subdivided into episodic memory making it possible to store specific events in time and semantic memory storing knowledge about the external world. Information is stored in chunks (N. Ellis, 1997), i.e. units of memory organization and chunking refers to "the ubiquitous feature of human memory" (N. Ellis, 1997, p. 124) to form larger units from already made chunks leading to a hierarchical organization of the mind.

Summarizing empirical research on lexical acquisition N. Ellis (1997, p. 133) points out that "some people have difficulty acquiring lexis because of their problems in sequencing and chunking in phonological memory", as STM, measured as the longest sequence of digits an individual is able to repeat in correct order, is a reliable predictor of long-term acquisition of vocabulary and syntax. Furthermore, he claims that phonological STM, measured as the longest string of non-words that an individual is able to repeat in correct order, is a reliable predictor of later vocabulary acquisition in both L1 and L2. N. Ellis (1997) argues that unlike the implicit acquisition of word form, the acquisition of lexical meaning is an explicit process requiring a conscious application of strategies of searching for information, hypothesis formation and testing. He claims that "some people have difficulties acquiring L2 lexis because they fail properly to infer the meanings of new lexis" (N. Ellis, 1997, p. 135).

Language is more lexical (i.e. memory-based) than syntactic (i.e. rule-based), argues Skehan (1998) and claims the existence of a dual-representation system, which has recently been further described by Ullman (2001). His model makes a clear distinction between the memorized mental lexicon and the computational mental grammar, and proposes a major shift in second language acquisition research. He states that while grammar is sub-served by the procedural memory in L1, experimental evidence shows that in second language learning grammatical rules are largely memorized by declarative memory, just like lexical items.

Numerous studies have investigated the role of memory in lexical acquisition (Skehan 1998), most of which examine strategy use and methods of retention in vocabulary learning. Several factors emerge in research as influential, the most significant being the difficulty of the task. Skehan (1998) identifies three factors determining task difficulty: 1) code complexity i.e. the syntactic and lexical difficulty of the input; 2) cognitive complexity i.e. the processing demands of the task and the availability of schematic knowledge; 3) communicative stress i.e. factors involving time pressure, number of participants involved, or the modality of the task performance. Schneider, Healy, and Bourne (2002) found that retention and transfer are better when the conditions of training are difficult (referring to task difficulty, not the difficulty of materials), which they explain by the added mental processing requirements that difficult tasks impose on the learner.

This assumption is further strengthened by the findings of Robinson (2001a), who in a study examining the effects of task difficulty on learner production of 44 Japanese undergraduate students found significant correlations between the increased cognitive complexity of a map-direction-giving task and speaker production. More complex task demands induced more lexical variety, while greater fluency was observed on a more simple task version. Similarly, Rott, Williams and Cameron (2002) proposed that multiple-choice L1 glosses requiring increased mental effort on part of the learner significantly contributed to both productive and receptive word gains with 76 learners of German in tests immediately after treatment. Similar task difficulty effects were reported by Taguchi (2007) for oral L2 production with 59 Japanese students of English.

On the other hand, Kuiken, Mos and Vedder (2005) examining the effect of task complexity on written production found increased accuracy in cognitively more demanding tasks, with stronger effects for high-proficiency than low-proficiency learners, but no significant effects on syntactic complexity or lexical variation. To create a framework for the converging empirical evidence in L2 research Laufer and

Hulstijn (2001) proposed the construct of Task-Induced Involvement referring to the observation that "the retention of hitherto unfamiliar words is conditional upon the amount of involvement while processing these words" (2001, p. 1).

The first comprehensive taxonomy of vocabulary learning strategies with indications on their perceived helpfulness by the 600 participants of a large-scale study was proposed by Schmitt (1997, pp. 207-208). He made a distinction between strategies used for the discovery of a new word's meaning and subsequent strategies for consolidating word meaning. It has clearly emerged from the vocabulary research literature that memory strategies (committing form-meaning pairs to memory) and vocabulary learning strategies need to be distinguished. Mnemonics as successful aids to boost memory miraculously has long fascinated psychologists, language teachers and learners alike. The most widely used mnemonics device is the keyword method (KWM), in which a target language word is linked to a sound-alike L1 word through an image or a sentence involving both words, assuming that the foreign word would trigger the keyword facilitating the retrieval of meaning (Hulstijn, 1997). As Saphiro and Waters' (2005) experiment revealed, "the KWM is effective because it provides a meaningful visual image upon which to base memory for a new word's meaning" (2005, p. 129).

How new vocabulary is pre-organized for learners seems to be another key issue in vocabulary acquisition with direct pedagogical implications. Tinkham (1997) reported that semantic clustering of semantically and syntactically similar words hinders (e.g., *eye, mouth, nose, ear*), while thematic clustering fosters vocabulary learning (e.g., *frog, green, hop, pond*), which might cast light on the importance of underlying schemata or frames segmentizing the speaker's background knowledge in the organization of the mental lexicon.

Vocabulary learning, on the other hand, is not simply memorizing items in isolation but a skill to be developed (Gu & Johnson, 1996; Schmitt, 1997); therefore, the aim of vocabulary learning should be learning the skill of recognizing words in their contexts,

guessing their meanings, and using the words correctly and appropriately (Gu & Johnson, 1996). It has been found that contextual guessing/inferencing meaning from context, dictionary use, note-taking, paying attention to word formation, contextual encoding and activation of newly learnt vocabulary all contribute to better retention of words in subsequent tests (Gu & Johnson, 1996; Hulstijn, Hollander, & Greidanus, 1996; Kojic-Sabo & Lighbown, 1999; Laufer & Hill, 2000; Lawson & Hogben, 1996; Schmitt, 1997; 2000). It is reasonable to suppose then that a combination of strategies might lead to better results. Rodriguez and Sadoski (2000) comparing the effects of the rote rehearsal, the context, the keyword and the context/keyword methods of training observed the superiority of the combined context/keyword method in word retention over the other three.

Metacognitive strategy training also has a significantly positive effect on the vocabulary retention of EFL students (Gu & Johnson, 1996; Rasekh & Ranjbary, 2003; Schmitt, 1997, 2000). It has been shown that raising learner awareness of their own learning, the importance of preparation and planning, the selection of vocabulary learning strategies to be applied, the monitoring of strategy selection and use, the use of a combination of vocabulary learning strategies at the same time and the evaluation of their effectiveness either by self-questioning, learning logbooks, check-lists of strategies used or open-ended questionnaires all prove to be efficient aids in the vocabulary learning process. However, it clearly derives from the above discussed empirical research that the choice and use of vocabulary learning strategies depend on the person, the task and the context of learning, as Gu (2003) points out.

2.3 The scope of vocabulary acquisition: breadth and depth

In the first chapter we have seen that lexical competence is a major component of linguistic competence and that researchers do not seem to agree on how lexical competence is built up. In an attempt to further clarify the processes of vocabulary acquisition, the dimensions of vocabulary knowledge and the interrelations between them, Henriksen (1999) proposes a three-dimensional description: (1) partial to precise

knowledge, (2) depth of knowledge, and (3) receptive to productive use ability. By development from partial to precise knowledge of word meaning she defines a "move or progression from rough categorization or vagueness to more precision and mastery of finer shades of meaning" (1999, p. 311). She argues that learners may never achieve the full knowledge of a word, but this is not even necessary for all comprehension purposes and communication may be successful with these gaps in lexical knowledge. She proposes that development along dimension 1 is primarily associated with a mapping process, i.e. narrowing down the field of reference by labeling (creating a link between sign, concept and referent), and packaging (the process of discovering what things can be packed under one label).

Her depth of knowledge dimension is often referred to in other studies as the quality of word knowledge. As has been shown in the previous chapter, knowing a word may involve knowing how it behaves in terms of its orthography, morphology, syntax, and as Singleton (2000) put it, what kind of company it may keep, i.e. its collocation patterns with other words, its paradigmatic and syntagmatic relations (Laufer, 1997a; Nation, 1990; Richards, 1976; Singleton, 2000). In Henriksen's view, in the semantization process the depth of knowledge dimension thus involves network building processes she defines as "developing and handling new sense relations between words" (1999, p. 308). There have been numerous attempts to measure the depth of word knowledge (Bogaards, 2000; Laufer, 1998; Meara & Jones, 1990; Paribakht & Wesche, 1993) which are going to be discussed in detail in the next chapters on vocabulary assessment. It is only important to note at this point that research findings seem to depend crucially on task demands, word class, and on how the concept of word knowledge is perceived.

Henriksen's (1999) third dimension is the receptive-productive continuum. It is generally accepted that only a limited number of words we know receptively will ever become productive and that most words we encounter initially enter the receptive vocabulary and may only subsequently become productive (1999, p. 313). Nation (1990), for instance, proposes an inclusive relation between the two: productive

vocabulary knowledge involves and extends receptive knowledge (for an overview see also Melka, 1997). Similarly, Laufer (2005a) and Webb (2008) claim that the receptive vocabulary is always larger than the productive vocabulary, implying that the active use of vocabulary is more difficult to achieve.

A great number of studies have attempted to estimate the size of receptive and productive vocabulary discussed in Chapter 3 in detail (D'Anna, Zechmeister, & Hall, 1991; Goulden, Nation, & Read, 1990; Hazenberg & Hulstijn, 1996; Laufer, 1987, 1992a, 1992b; Nusbaum, Pisoni, & Davis, 1984; Webb, 2008; Zechmeister, D'Anna, Hall, Paus, & Smith, 1993). Melka (1997) summarizes the critical points responsible for the discrepancies between estimates of reception and production in such studies as the type of words tested, the problem of possible words versus actual words, the role of L1 and the cross linguistic influence of cognates. Meara (2004, 2005), however, criticizes all continua models of vocabulary acquisition for being merely descriptions of states and not of progress, and calls for a more comprehensive approach presupposing a tighter cooperation of applied linguists and psychologists.

All three above proposed dimensions, however, leave open the widely debated question of acquisition and learning. Based on the developmental sequences and acquisition hierarchies of morpheme orders in English as a second language, Zobl (1995) argues for Krashen's (1977) Monitor Model making a distinction between 'learning' and 'acquisition'. He claims that "'acquisition' operates incidentally to processing for comprehension and results in implicit, intuitive knowledge; 'learning' relies on memorization and problem solving and leads to explicit, conscious knowledge about the L2" (Zobl, 1995, p. 35). This distinction is applied in the rest of the studies discussed further on in this chapter, some researchers, however, use them as synonymous terms (Laufer, 1998). In the subsequent chapters, therefore, the use of the terms is going to vary according to the nature of the topic and the arising need for specifying or emphasizing one or the other process.

2.4 Incidental vocabulary acquisition

The Default hypothesis of L1 vocabulary acquisition states that we acquire most words by repeated exposure to written or oral input (Laufer, 2005b), based on the assumption that our vocabulary is simply too vast to result from direct instruction only. This has also found supporters in L2 research. Krashen's Input Hypothesis predicts that more comprehensible input in the form of reading results in greater competence in L2 vocabulary (Krashen, 1989, p. 441). Although the hypothesis has so far been ground for a lot of debate, it is generally accepted now that some lexical learning appears to take place while reading (Ellis, 1994) and this is often referred to as incidental vocabulary learning (Gass, 1999; Huckin & Coady, 1999). Even though a vast body of research has aimed at clarifying the mechanism of this process (Brown, Sagers, & LaPorte 1999; Ellis & He 1999; Fraser, 1999; Hulstij, Hollander, & Greidanus, 1996; Nassaji, 2003; Paribakht & Wesche, 1999; Rott, 1999; Walters, 2004; Wode, 1999), the use of the term is still ambiguous (Huckin & Coady, 1999).

2.4.1 Incidental vs intentional vocabulary learning

Most researchers seem to agree that incidental learning is a "by-product" of something, but definitions of what that "something" might be are diverse. Wode (1999) sees it as a by-product of language use without focusing attention on the linguistic structure itself. Huckin and Coady (1999, p. 182) term secondary learning as a "by-product, not the target of, the main cognitive activity". Hulstijn, Hollander, and Greidanus (1996) employ a negative definition: learning without an intention to learn.

A similar definition is shared by Laufer and Hulstijn (2001) who refer to 'incidental learning' in the experimental psychology literature from the beginning of the twentieth century and argue that in studies of this field learners are usually asked to perform a task without being told that there will be a later test on how they are able to recall the information or the vocabulary involved. Therefore, it is the "absence or presence of a forewarning of an upcoming retention test after the information processing task" that

makes the difference between incidental and intentional learning (Laufer & Hulstijn, 2001. p. 10).

Research has shown, however, that incidental vocabulary acquisition is incremental, with a low pick-up rate (Fraser, 1999; Gardner, 2004; Gu, 2003; Hulstijn, Hollander, & Greidanus, 1996; Laufer, 1997a, 2005a, 2005b; Paribakht & Wesche, 1999; Schmitt, 2000). Schmitt (2008) in his overview explains that early studies on incidental vocabulary acquisition had considerable methodological weaknesses (small amount of reading exposure, inadequate measurement instruments and control of text difficulty, lack of delayed post-tests, etc.), while recent studies report more encouraging results (Lehmann, 2007a; Sankó, 2008). The major issues involved in the immense body of research carried out in the past decade focus on how incidental vocabulary acquisition takes place, how much and what type of vocabulary is necessary for it to take place, the role of strategies applied, the amount of exposure required, and what the role of input modification such as glossing or using hypertextular input may be in the process (see Sankó, 2008).

2.4.2 Vocabulary size and lexical coverage

To understand a text the reader needs to comprehend its constituting words. But how many and what words are necessary for successful comprehension? Corpus research, the study of large compilations of written and spoken language, has considerably contributed to this aspect of vocabulary research. It has shown that the 5,000 most frequent words in the English language cover 90-95% of authentic texts (Hirsch & Nation, 1992). Nation (2001) found that the first 2,000 most frequent words, together with the 570 word families included in the Academic Word List, a list of the most frequent words occurring in a wide range of academic texts compiled by Coxhead (2000), cover 92% of spoken language and 84% of newspaper language. In consequence, estimations on how many words in a text a reader needs to be familiar with to achieve reasonable comprehension amount to 95-98% of all the words in any text (Hirsch & Nation, 1992; Laufer, 2005a, 2005b; Nation & Waring, 1997),

implying that "one in every 50 words is unknown" (Schmitt, 2008, p. 330).

Studying the Wellington Corpus of Spoken English, Nation (2006) found that to achieve the 98% goal, the knowledge of 6,000-7,000 word families would enable the learner to understand spoken discourse, a figure supported by the findings of Adolphs and Schmitt (2003) based on an analysis of the CANCODE corpus. Similarly, using the British National Corpus (BNC) Nation (2006) calculated that learners need to have a vocabulary size of 8,000-9,000 word families to be able to read a wide range of authentic texts. Earlier Laufer (1992b) suggested 5,000 word families.

However, it is important to note here, that as was seen in the first chapter, a word family involves the base word and all its possible inflected and derived forms. Nation's (2006) calculations show that the knowledge of 8,000 word families required for good reading comprehension entails 34,660 individual word forms. It cannot be taken for granted, therefore, that a learner is familiar with all the word forms within a word family. Schmitt (2008), for instance, refers to a study conducted jointly with Zimmerman, where advanced learners of English knew only some of the words included in the 570 word families of the Academic Word List (Coxhead, 2000).

2.4.3 Strategy use and incidental vocabulary learning

The Default-hypothesis is based on five assumptions about the processes involved in vocabulary acquisition from reading: (1) the noticing assumption; (2) the guessing ability assumption; (3) the guessing-retention link assumption; (4) the repeated exposures-retention link assumption; and (5) the extrapolation assumption (Laufer, 2005b). In the following section an overview of research attempts to highlight these aspects of the vocabulary acquisition process.

In the process of reading learners use a variety of strategies to comprehend the text (Alderson, 2000; Grabe, 2002; Koda, 2005; Schmitt, 2000). Among many others, one is guessing or inferring the meaning of unknown words from their context (the

guessing ability assumption). However, to be able to guess, the reader must first notice the word as unknown (the noticing assumption) which is a prerequisite of any learning taking place, argues Laufer (2005b). She describes the following possible cases with L2 learners when a new word occurs in a text as shown in Figure 1:

Word in text

not recognized as unfamiliar (confusion with another word over-evaluation)

recognized as unfamiliar

no attempt to guess (word irrelevant, time pressure)

attempt to guess

too difficult   moderate   easy

no guess

wrong/right right guess
guess   (mostly)

-learning   -learning   - learning   -learning   -learning
                                  or incorrect   or +learning   or +learning
                                  learning
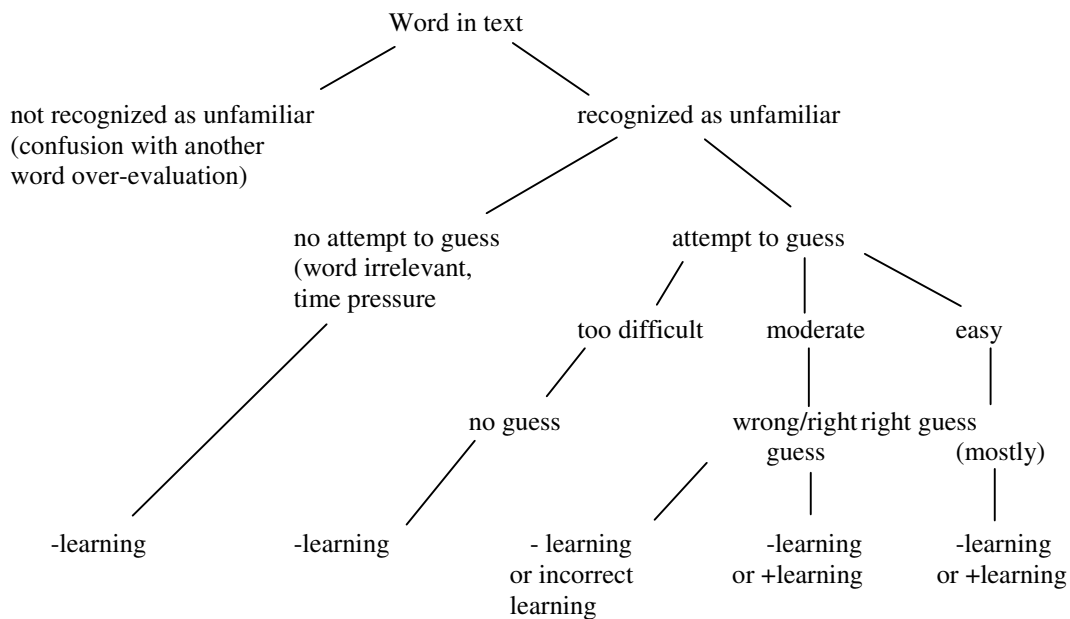
Figure 1. What happens when a learner comes across a new word in a text? (Laufer, 2005b, p. 317).

The new word is either recognized or not recognized as unfamiliar. There is a chance for vocabulary learning only in the former case, provided that the learner makes an attempt to guess the meaning of the word from context. Paribakht and Wesche (1999), for instance, found that ESL university learners from a variety of L1 backgrounds ignored approximately half of the words they recognized as unfamiliar, while Fraser (1999) reported 24% in a study with Francophone ESL learners. Less proficient learners, who possess fewer basic language skills to be able to infer meaning from context, are less efficient guessers than intermediate or advanced learners (Gu, 2003).

On the other hand, if learners attempt to guess, it can still prove to be too difficult or the learner may guess wrong leading to learning an incorrect meaning due to cognate

effects or lexical items that may be misleading (Huckin & Coady, 1999). However, Laufer (2005b) warns that even if the meaning of the word is guessed right, it is very probable that it will not be committed to long term memory, i.e. will not be learnt. This partly contradicts the 'guessing-retention link' assumption claiming that readers acquire vocabulary by contextual inferencing.

Ample empirical evidence has shown that various types of knowledge contribute to efficient contextual guessing (Nagy, 1997; Pulido, 2007; Walters, 2004): morphological and syntactic knowledge, vocabulary knowledge, word schematas, world and strategic knowledge. In an experiment with 21 intermediate-level ESL students Nassaji (2003) found that students resorted to their world knowledge and morphological knowledge most often and most successfully, which supported the earlier findings of Paribakht and Wesche (1999), while Pulido (2004, 2007) highlights the importance of topic familiarity and passage sight vocabulary in the process of text comprehension. Among the inferencing strategies of repeating, verifying, monitoring, self-inquiry, analyzing and analogy, repeating has proved to be used most often, more specifically section repeating most successfully (Nassaji, 2003). This implies the importance of contextual clues and global strategies in successful inferencing.

Besides knowledge sources and strategy use Paribakht and Wesche (1999) further added task and text characteristics, as well as individual differences in strategy use, world knowledge and the learner's perception of the value of the reading task to contribute to successful inferencing. However, Lawson and Hogben (1996) found no significant association between the use of context and the retention of word meaning in contrast to the stress placed on context by the above described studies. In their conclusion they propose the need to distinguish between the use of context for "the generation of meaning of a new word and the use of context for acquisition of the meaning for subsequent recall" (Lawson & Hogben, 1996, p. 102), shifting the issue towards assessment purposes introduced in Chapter 3.

2.4.4 Exposure frequency and incidental vocabulary acquisition

It has been argued that incidental vocabulary learning is incremental and it is reasonable to conclude that little learning takes place on initial exposure to a word ('repeated exposures-retention link' assumption). "First encounters with a word may result in partial or vague understanding. Each additional exposure to the same word may enrich and strengthen the learner's knowledge of it" (Laufer, 2005b, p. 318), therefore a possible way to provide sufficient exposure is by extensive reading. However, there is no agreement on how many exposures are necessary for successful vocabulary acquisition from reading. Rott (1999), for instance, reported that six encounters resulted in more significant gains in vocabulary than two or four encounters both in receptive and productive word knowledge.

In a study with 121 Japanese students learning English Webb (2007) used ten tests to control for vocabulary gains concerning the knowledge of word orthography, association, grammatical functions, syntax, meaning and form after one, three, seven and ten encounters with the words. He found greater gains in at least one aspect of word knowledge by increasing the number of repetitions. He claims that "if learners encounter unknown words ten times in context, sizeable learning gains can occur", but to achieve full level of word knowledge, 10+ exposures are necessary (Webb, 2007, p. 46). Nation (1990) in his review of the research thus concluded that 5-16 exposures are necessary for acquisition. Furthermore, Nation and Wang (1999), after examining the vocabulary of level 5 graded readers, suggested that learners read one-two books a week. This, they explain, means that for 108 words to be met 10 times (if we accept that this is the optimal figure for acquisition), learners need to read 20,000 words.

However, Laufer's (2005b) extrapolation assumption states that there is no evidence to prove that if 5 words are learnt from reading a text of 1,000 words, for example, then 50 words will be learnt from a ten-thousand-word long text, as such a calculation does not take the instability of learning conditions into account e.g., memory fatigue, changing interest, different emphasis on language skills and a changing syllabus in an

instructional setting. Hulstijn, Hollander, and Greidanus (1996) have shown that learners immediately forgot about half of the words they processed (by consulting a dictionary or marginal glosses) for meaning during reading. Hence, Huckin and Coady (1999) warn that there are so many variables that determine the learning of a word as described earlier that it is impossible to determine a threshold for the number of exposures. In other words, all models claiming that vocabulary acquisition is a simple, mechanical, repetition-based process must be wrong. There is more to language learning than repetition.

2.4.5 Input modification and incidental vocabulary acquisition

The figures of exposure frequency recommended above for incidental vocabulary acquisition may seem daunting for second language learners, thus it is sensible to find ways to promote the process. Input can be modified either by making it grammatically less complex and more redundant (premodified input) or in the process of negotiation of meaning (interactionally modified input) and Ellis and He (1999) argue for the superiority of the latter. Hulstijn, Hollander and Greidanus (1996) found that providing marginal glosses and dictionary use fostered incidental vocabulary acquisition with intermediate and advanced L2 learners. They claim that L1 cues appear to have more effect than L2 cues and report that intermediate and advanced L2 learners seldom use a dictionary when reading long texts of more than one page or when there is an overload of unknown words in the text. However, consulting a dictionary is important in verifying inferences (Fraser, 1999). Such modifications of input are ways of directing explicit attention on lexical items. Their efficiency may imply that incidental and intentional learning are complementary processes, as has recently been proposed by a growing number of researchers (Laufer, 2005a, 2005b; Schmitt, 2008). As Read (2000) points out, the position that incidental L2 vocabulary acquisition would take care of itself without any need for pedagogical intervention provided that learners have access to sufficient comprehensible input is no longer tenable.

2.5 Intentional vocabulary learning

In the above sections we have seen what factors foster the incidental acquisition of vocabulary from context. Conditions of learning, however, significantly differ in an instructional setting. Nation (2001) outlined four main trends in vocabulary instruction (Appendix B). Meaning-focused input in his interpretation covers what we have so far discussed under the umbrella of incidental vocabulary acquisition. However, vocabulary knowledge may also be developed in communicative writing tasks (meaning-focused output). Language-focused input involves explicit attention to lexical items together with strategy training, the benefits of which has been discussed earlier in this chapter. The fourth strand, fluency development highlights the need for increasing automaticity in word recognition and production speed for enabling more fluent comprehension and language production.

In another, more recent approach, Laufer (2005b) argues that the 'noticing', the 'guessing-retention link' and the 'repeated exposures-retention assumption' discussed previously in the sections on incidental vocabulary acquisition cannot be taken for granted in an instructed learning context. L2 learners often fail to notice unfamiliar words as unfamiliar (as reported by Fraser [1999] in 3% of all unfamiliar word encounters), or when they do, they fail to infer meaning from context correctly, leading to erroneous learning, especially if they do not know 98% of the words in the text.

Also, inferencing capability is dependent on language proficiency level. Jones (1995), learning Hungarian on his own, for instance, found vocabulary learning the most challenging as Hungarian has few borrowings from other languages; therefore, there is hardly any cognate effect enhancing learning from context. Only after reaching a threshold level was he able to infer meaning from context to read effectively and even correct guessing may not necessarily lead to successful retention of meaning. Coady (1997a) proposed the term 'beginner's paradox' to account for such difficulty: "students must read in order to learn words, but at the

same time they must possess a minimal but critical mass of words to be able to read successfully" (Coady, 1997b, p. 284). Laufer (2005b) furthermore claims that the amount of reading required for repeated exposures is unrealistic in instructional settings and the value of short readings cannot be extrapolated to larger quantities of reading. Paribakht and Wesche (1999, p. 215) also pointed out that

> ...vocabulary learning outcomes from reading will always be unpredictable because different learners attend to different words, invest different levels of effort into figuring out the meanings of unknown words while performing a task, and differ in the availability of previous knowledge and the tendency to use particular strategies that can help them to successfully solve their lexical problems.

For an alternative solution to the problem Laufer proposes the Planned Lexical Instruction (PLI) Hypothesis to compensate for the shortcomings of the incidental learning approach. "In view of the special conditions which obtain in instructed language learning context, the main source of L2 vocabulary knowledge is likely to be word focused classroom instruction" (Laufer, 2005b, p. 321). Her hypothesis is based on three strands of form-focused instruction: (1) the classic 'task-embedded' Focus on Form (FonF); (2) task-related Focus on Forms (FonFs); and (3) 'pure' Focus on Forms (FonFs) not related to any task (Laufer, 2005a, p. 245).

The first approach (FonF) claims that all learning must occur within communicative tasks rather than through explicit teaching and comprehensible input alone is a necessary but insufficient condition for acquisition to take place (Laufer, 2005a, p. 224), thus contradicting Krashen's (1989) Input Hypothesis discussed earlier in this chapter. She argues, on the other hand, that in a Focus on Forms (FonFs) approach learners rehearse and practice words repeatedly in demanding and carefully planned tasks. To ensure repeated encounters that reading only and task-related FonFs is not able to provide, especially with low-frequency items, learners have to be provided with additional word focused instruction and practice ('pure' FonFs), "where words are the objects of learning, and not only tools of communication" (Laufer, 2005a, p. 233).

Meaning Focused Instruction

L2 Vocabulary
Acquisition

| Task embedded FFI Focus on Form | Task related FFI Focus on Forrms | Task unrelated FFI 'pure' FonFs |

Form Focused Instruction
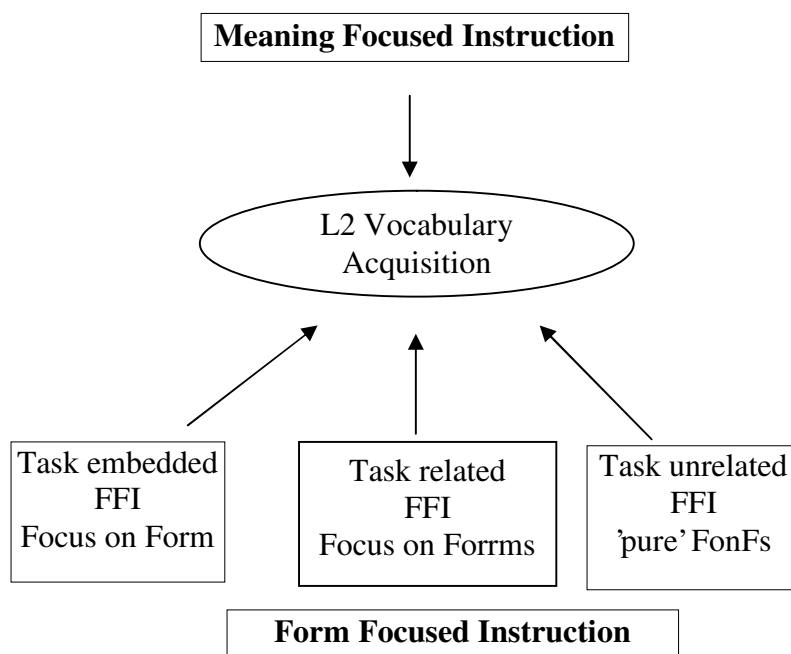
Figure 2. Components of L2 vocabulary instruction (Laufer, 2005a, p. 245)

This rich instruction, defined by Nation (2001) as "giving elaborate attention to a word, going beyond the immediate demands of a particular context of occurrence" (p. 95) makes it possible to progress from passive recognition to more demanding active use, as well as develop along the partial- precise continuum of qualitative word knowledge put forward by Henriksen (1999) by increasing and strengthening the paradigmatic and syntagmatic relationships with other words in the mental lexicon. In comparison of the two approaches Laufer (2005a) reports the results of three studies demonstrating the superiority of FonFs over FonF. However, in conclusion, she proposes that for successful L2 vocabulary acquisition it is most beneficial to combine meaning- and form-focused instruction (Figure 2).

It has emerged from our discussion of these two chapters that vocabulary teaching has benefited from advances in psycholinguistic research on the nature of the mental lexicon; from advances in linguistics on the nature of words and their behaviour in relation to other words; the boom in information technology making computerized

data collection and analysis possible; and the spread of communicative language teaching placing the learners into the centre of the learning and teaching process. The challenge then teachers of an L2 language have to contend with is implementing theory into their everyday classroom practice. With regard to the pedagogical implications of the above discussed empirical research, Hunt and Beglar (1998) proposed the following seven major principles to vocabulary teaching:

Principle 1: Provide opportunities for the incidental learning of vocabulary.

Principle 2: Diagnose which of the 3,000 most common words learners need to study.

Principle 3: Provide opportunities for the intentional learning of vocabulary.

Principle 4: Provide opportunities for elaborating word knowledge.

Principle 5: Provide opportunities for developing fluency with known vocabulary.

Principle 6: Experiment with guessing from context.

Principle 7: Examine different types of dictionaries and teach students how to use them.

Based on his overview of research Schmitt (2008, p. 342) added three further recommendations to the above. First, "use activities that maximize learner engagement with target lexical items", by, to name but a few, using an interactive on-line database, chat programs for negotiating meaning in pairs, receiving L1 translations and using them in a sentence, producing new word forms on their own, recording words in a notebook with multiple aspects of word knowledge marked, etc. Second, "maximize repeated exposures to target lexical items" (p. 343) by consciously incorporating recycling of vocabulary into the syllabus and planning principled vocabulary development in the long-run.

Third, "consider which aspects of vocabulary knowledge to focus upon" (p. 343) by broadening the often followed method of solely introducing the meaning of new words, thus disregarding the many faceted nature of word knowledge discussed in the previous chapter. As Nagy (1997, p. 73) pointed out, "definition-based learning typically involves memorizing (or attempting to memorize) brief definitions representing only a single meaning of the word to be learnt, and hence it leads to only shallow level of word knowledge", thus it does not increase the comprehension of

texts containing the instructed words.

2.6 Conclusions

In this chapter we have seen that a lot remains to be explored about the mental lexicon in terms of neurosciences, applied linguistics and linguistics, and that these fields overlap considerably. Although the scope of the research introduced (especially those related to neurolinguistics, neuronetworks and connectionism) is related to the research I conducted, the development of the majority of the models introduced does not allow their appliciation in applied linguistic research at present. However, I found it necessary to discuss them briefly in this chapter for comprehensiveness.

The first section overviewed research on the mechanism of vocabulary learning, highlighting some major elements in the role of cognitive factors, the first language, lexical features and memory. Research on the bilingual mind has clearly indicated the positive effects of multilingualism on the brain. As Bialystok (2005) addressing issues of the cognitive consequences of bilingual development in children notes, bilingualism facilitates the development of general cognitive functions and results in enhanced mental flexibility across all domains of thought. Thus, bilingualism changes something fundamental about cognitive processes, but to what extent and exactly what and how are still not fully explored.

In the discussion of language representation and processing in a bilingual context it has been argued that late acquired languages are located and processed in different parts of the brain; nevertheless, language proficiency seems to influence where the languages are processed: highly proficient bilinguals tend to activate the same, while less proficient L2 users activate different neural regions for the L1 and L2 languages. There is evidence that although various parts of the brain age differently, the age of acquisition does not seem to result in impaired lexical learning. Words learnt earlier, however, have been reported to have processing advantages over later acquired words.

It has become clear that L2 learners transfer L1 lexical features onto the L2 lexical item, the degree and nature of which is dependent on any existing cultural overlap between the native and the foreign languages, possible occurring cognate effects, number of intralexical features, and the L2 language proficiency of the learner. With increasing proficiency the L2 word is gradually freed from the corresponding L2 word, hence there is reason to believe that it is beneficial to exploit the form-meaning link between the L1 and the L2 and it seems sensible to make use of it in many ways whenever it is to our advantage, especially in an instructional L2 learning context.

It has emerged from our discussion that although memory plays an important role in lexical learning, vocabulary learning is more than merely memorizing words. It has been argued that the degree of involvement in solving a cognitively demanding task is in proportion to vocabulary gains measured in retention tests. It has been emphasized, however, that there is no single best vocabulary learning strategy to commit word knowledge to long-term memory that language learners and teachers could contend with, for the strategies applied and perceived as most useful depend on the person, the task and the context of vocabulary learning, as well as the interpretation of what a word is and what the nature of word knowledge may involve, and some highly idiosyncratic other variables.

The overarching problems of vocabulary breadth and depth, or in other terms vocabulary size and the quality of vocabulary knowledge have been touched upon in an introductory fashion in this chapter, as essential terms and indispensable measurement issues are going to be discussed in the next chapter. It was important at this point to clarify that vocabulary knowledge, and consequently the mental lexicon, are not static in nature. It develops in time along interrelated continua from partial to precise, receptive to productive word knowledge and different words in the mental lexicon may be at different stages of development at a certain point in time.

The remaining sections of the chapter were devoted to highlighting the major differences between incidental and intentional vocabulary learning by focusing on five

widely referred to assumptions about vocabulary learning from reading, the primary source of vocabulary input relevant in the context of the empirical studies of the present thesis. The interpretation of empirical research, however, depends very much on how the concept of acquisition is defined. Krashen is not alone suggesting that vocabulary acquisition takes place by incidental learning from context if sufficient comprehensible input is provided. The value of extensive reading has not been argued, but the incremental nature of incidental learning has urged others to find ways to facilitate the process and adapt to the limitations of instructional settings. If we accept those arguments, it can be maintained that the two processes are complementary in L2 vocabulary acquisition.

Chapter 3

How is vocabulary assessed?

3. 1 Introduction

T he focus of this chapter is the assessment of the multifaceted nature of vocabulary knowledge introduced in the previous chapters, adding further relish to the issues raised so far in four main sections. The first section concerns global questions in vocabulary testing and emphasizes the importance of a multiple tests approach to assessment. The structure of the rest of the chapter is provided by the major distinction of measures and their corresponding tools suitable for exploring either the qualitative or the quantitative aspects of lexical knowledge, as currently there is no available test to measure both in the same test format. Some major issues in corpus linguistics, a closely related field providing sources and tools for vocabulary research, are touched upon with the aim of exploring ways of exploiting corpora and the related computer aids for vocabulary analysis, assessment and test construction. Finally, the fourth section of the chapter introduces four widely referred to vocabulary tests that are applied in the empirical studies in the second part of the present work.

3.2 Assessing global vocabulary knowledge: The multiple tests approach

In the past few decades there has been a thriving development in language testing (for introductions see Alderson & Banerjee, 2001; Alderson, Clapham, & Wall, 1995; Bachman, 1990; 2000; Bachman & Palmer, 1996; Bárdos, 2002; Hughes, 1991; Oller, 1979; Weir, 1993), especially in the field of vocabulary assessment (Carter, 1998; Coady & Huckin, 1997; Henriksen, 1999; Laufer, 2005a, 2005b; Laufer & Hill, 2000; Laufer & Nation, 1999; Nation, 1990; Read, 2000; Read & Chapelle, 2001; Schmitt,

2000, 2008; Schmitt & McCarthy, 1997; Zareva, 2005). Vocabulary tests are being used for a variety of educational and research purposes. To investigate any progress in vocabulary learning, it is necessary to describe vocabulary in quantitative and qualitative measures. Goulden, Nation and Read in a 1990 article claim that the most striking difference between native speakers and second or foreign language learners seems to be in the number of words they know.

However, as we have seen in the previous chapters, the knowledge of a word might vary from superficial to deep, receptive to productive, and can be interpreted in terms of a taxonomy of features from knowing the orthography of the word to being familiar with its possible collocates, which is not merely a question of quantity any more. It has emerged from research that four major issues need to be addressed in testing vocabulary:

1) Why to test vocabulary?
2) What words to test?
3) What aspects of word knowledge to test?
4) How to test the various aspects of word knowledge?

A reply to the first question requires a careful consideration of the purpose of assessment which determines the type of test applied. Is the assessment due at the end of a course to get information about the progress of the test-takers (a progress test or an achievement test) or are we interested in the general level of our test-takers' language proficiency (a proficiency test)? If the focus of testing is determining the strengths and weaknesses of the participants, we need a diagnostic test. To place students into groups at the enrollment to a language course, a placement test is to be applied. To set a level for students to pass, a filter test is to be used.

Once the purpose of assessing vocabulary is identified, obviously not all words are equally useful for all purposes in all contexts. Therefore, a decision has to be made on what words to test. And once the selection of words to be included in the test has been

made, yet there is the question of what aspects of word knowledge are to be tested, depending on what kind of information we want to gain about the knowledge of the words selected. Is it more beneficial for our identified purposes to test qualitative or quantitative aspects of word knowledge, or perhaps both? Are we interested in whether the testees are able to recognize the given words as known or whether they are able to produce them on their own? Finally, a careful consideration has to be given to how to test the above defined aspects of the knowledge of the selected words. The following six dimensions of vocabulary assessment are outlined by Read (2000) as shown in Figure 3:

Discrete
A measure of vocabulary knowledge or use as an independent construct

Embedded
A measure of vocabulary which forms part of the assessment of some other, larger construct

Selective
A measure in which specific vocabulary items are the focus of the assessment

Comprehensive
A measure which takes account of the whole vocabulary content of the input material (reading/listening tasks) or the test-taker's response (writing/ speaking tasks)

Context-independent
A vocabulary measure in which the test-taker can produce the expected response without referring to any context

Context-dependent
A vocabulary measure which assesses the test-taker's abiltity to take account of contextual information in order to produce the expected response

Figure 3. Dimensions of vocabulary assessment (Read, 2000, p. 9)

Discrete vocabulary tests measure vocabulary knowledge as an independent construct, or the vocabulary test can be embedded forming a smaller part of a larger construct, such as a reading comprehension test. In a selective vocabulary test specific words are the focus of assessment, whereas a comprehensive measure of vocabulary knowledge

targets the whole input material such as the test-takers vocabulary use in an essay writing task. A vocabulary test is context-independent when the items are presented in isolation, like in the case of a word list, whereas if the test-taker needs to take into account contextual information to respond to a gap-filling type of test, for instance, the vocabulary test is context-dependent. In order to illustrate the above considerations, Table 5 below shows the design features of some of the major exemplary tests referred to later on in this chapter as well as in the empirical studies discussed in the second part of the present work.

Table 5. Design features of some exemplary tests

Test	Features		
Vocabulary Knowledge Scale	Discrete	Selective	Context independent
Vocabulary Levels Test	Discrete	Selective	Context independent
Test of Controlled Productive Ability	Discrete	Selective	Context dependent
Lexical Frequency Profile	Discrete	Comprehensive	Context dependent

It is clear from the above that the choice of a test format should be preceded by a sequence of carefully considered decisions about the purpose of testing, the type and number of words to test. As Read and Chapelle (2001, p. 10) worded the question in the overall framework for vocabulary assessment they propose, "does the operational test deliver the information required about the learners' vocabulary knowledge or ability, while having the desired impacts on those who use it?" They claim that at present there is a lack of a comprehensive basis for the comparison of current tools of vocabulary measurement or the future design of vocabulary tests and their framework presented in Figure 4 aims to fill this gap.

```
TEST PURPOSE  Inferences              Uses                    Intended impacts
                  │                     │                         │
VALIDITY      Construct             Relevance                  Actual
CONSIDERATIONS validity            and utility                consequences
                  │                     │                         │
MEDIATING     Construct             Performance                Test
FACTORS       definition           summary and               presentation
                                    reporting
                          ╲              │              ╱
TEST DESIGN                  Decisions about the dimensions
                                 Discrete – Embedded
                               Selective – Comprehensive
                         Context-independent – Context-dependent

                                         │

VALIDATION        Arguments based on theory, evidence and consequences
```

Figure 4. A framework for vocabulary testing (Read & Chapelle, 2001, p. 10)

The framework addresses the issue whether vocabulary is to be treated as a discrete construct. Read and Chapelle (2001) explain that L2 vocabulary researchers have long neglected this question central to language testers and describe the following three approaches to construct definition, as shown in Table 6 below:

Table 6. Three approaches to construct design (Read & Chapelle, 2001, p. 7)

	Trait definition	Behaviourist definition	Interactionist definition
Principle underlying construct definition	Person characteristics must be specified independent of context..	Person characteristics cannot be specified. Context must be specified.	Person characteristics must be specified relative to a particular context.
Example of construct definition	Vocabulary size	Vocabulary use in mathematics writing	Vocabulary size for writing in mathematics

According to a trait definition of vocabulary, performance on a test is attributed to the test-taker, eliminating the role of context, therefore, a trait definition entails the use of discrete, selective and context-independent tests, such as the Vocabulary Levels Test (Nation, 1990) or the vocabulary size test developed by Goulden, Nation and Read (1990). In these tests the vocabulary items are presented in isolation and the learners have no possibility to infer meaning from context. The Behaviourist approach, on the other hand, emphasized the role of context as central to test performance. Vocabulary is not a salient issue in test construction and the focus is on macroskills, such as listening or reading comprehension or writing skills. Vocabulary is tested in an embedded, comprehensive and context-dependent way. In an interactionalist approach to construct definition vocabulary is "a context specific underlying trait" (Read & Chapelle, 2001, p. 9); therefore, vocabulary tests may either be discrete or embedded, selective or comprehensive, but context dependent, such as a multiple-choice cloze test or the vocabulary items in section 3 of the Test of English as a Foreign Language (TOEFL).

As it has become clear from the above discussion and is going to be emphasized throughout this chapter, no single test at present suits all purposes and all contexts due to the complexity of the nature of vocabulary knowledge. Laufer (2001), therefore, argues that administering several types of tests to the same learners may be beneficial for two main reasons. First, combining the results of learners on several different types of tests gives a more comprehensive picture of their vocabulary knowledge. Second, a comparison of test results by each learner makes it possible to investigate the relationship of the various aspects of word knowledge and the way progress takes place. For this reason, in the following sections methods of measuring either the qualitative or the quantitative aspects of lexical knowledge are discussed subsequently before introducing four major vocabulary tests applied in the empirical chapters in detail that would provide ground for a multiple tests approach.

3.3 Qualitative measures

For numerous reasons, researchers seem to have focused more on the quantitative than the qualitative aspects of word knowledge which is often referred to as 'depth', 'precision', or 'quality' of vocabulary knowledge. Tests of vocabulary quality are designed to assess the precision of word knowledge to find out whether the test taker has only a vague idea about the meaning of the word or knows a specific meaning of the target item. Some tests attempt to measure degrees of word knowledge on the partial-precise continuum (Henriksen, 1999), while others operate along the multiple aspects of word knowledge described by Nation (1990). Schoonen and Verhallen (2008, p. 212), however, point out that the "depth of lexical knowledge is not a general feature of the lexicon, but a feature of individual words".

3.3.1 The personal interview

The most widely accepted method of qualitative vocabulary assessment is the personal interview which is capable of depicting a precise picture of the learners' lexical knowledge. There are no time constraints imposed on the test taker unlike in a written pen-and-pencil test and the interviewer can assess various levels of word knowledge by asking a series of well planned and structured questions. Verhallen and Schoonen (1993, cited in Read, 2000, p. 94) for instance, asked Dutch bilingual and monolingual children the following set of questions to explore knowledge of the word *book*:

> What does [book] mean?
> What is a [book]?
> How would you explain what a [book] is?
> What do you see if you look at a [book]?
> What kinds of [book] are there?
> What can you do with a [book]?
> Can you make three sentences with the word [book]?
>   (Verhallen & Schoonen, 1993, cited in Read, 2000, p. 94)

The major disadvantages of the personal interview are clearly illustrated by the above

example. On the one hand, the procedure is time consuming and requires a compromise on the part of the researcher about how much information to elicit about each word and how many items to include in the test. On the other hand, interview results are less consistent and less reliable than those of objective tests applied in quantitative assessment. Read (2000) further emphasizes the importance of the good rapport between the interviewer and the interviewee, as adults seem to be embarrassed in a face-to face situation and, consequently, less willing to admit not knowing a word.

A written version of the personal interview was experimented by Read (2000, p. 179) as an alternative to making up for the aforementioned drawbacks. He applied a three-stage procedure for each of the ten tested words. In stage one testees have to write a sentence with two word pairs given including the target word. Stage two is blank filling to test possible collocations of the target word, while in stage three the test taker is to fill in derivational endings of members of the target word family provided in a sentence. However, Read (2000) found the method impractical for time constraints, the limited number of items and the difficulty arising from evaluating the quality of responses.

3.3.2 Word association tests

For these practical reasons it became necessary to find alternative ways to test the depth of vocabulary knowledge. Of numerous attempts, studies on the receptive-productive continuum and research on word associations seem to be most noteworthy. Word association studies presenting the subjects with stimulus words and asking them to say any related words that come into their minds are mainly psycholinguistic in nature and aim to map the organizational complexity of the mental lexicon of fluent speakers. Word associations are generally classified as syntagmatic or paradigmatic. Syntagmatic associations are those which imply a sequential relationship between the stimulus word and the associated words, e.g., *dog, bark, bite, furry*. Paradigmatic associations, on the other hand, are words from the same grammatical class as the stimulus word, e.g., *dog, cat, animal, pet* (Carter, 1998, pp. 198-199).

The most influential word association test is the Birbeck Vocabulary Project conducted at the University of London in the 1980s (Meara, 1982, cited in Carter, 1998, pp. 197-204). The Project set out to explore the differences in the organization of second language learner and monolingual native speaker mental lexicons and reported that word associations produced by native speakers and second language learners systematically differ and learner responses tend to be more varied and less homogeneous (Carter, 1998, p. 199). Recognizing the numerous limitations of the project, Meara and his colleagues concluded that in the early stages of learning there seems to be no clearly established network in the learners' mental lexicons.

3.3.2.1 The Euralex French Tests

In an attempt to overcome these limitations, in the Euralex French Tests format described and validated by Bogaards (2000) and aimed at measuring very high levels of lexical knowledge in French testees have to indicate whether or not there is any obvious connection between pairs of words in a sixty-item list. For instance, with items *pied* and *grue*, the learner has to be familiar with the expression *faire le pied de grue* (to stand about waiting). The test is therefore receptive, measures word recognition and is claimed to be rather precise.

Bogaards (2000) argues that the tests are about fixed expressions, collocates, synonymy, antonymy and hyponymy, i.e. the categories included in Nation's (1990) definition of word knowledge. It is furthermore claimed to be suitable to check a number of aspects of knowledge of lexical units that are composed of more than one word, such as the special verb-subject/object relationships, relevant cultural knowledge, phrasal verbs and different senses of polysemous words, although it cannot test all aspects of word knowledge. Even though in the validation procedure the test did not turn out to be a sufficiently reliable tool due to dummy items, Bogaards (2000) claims that the test format is potentially a useful tool for measuring specific aspects of vocabulary knowledge not addressed by other tests. However, the fact that

yes/no type self-report tests allow fifty percent chance for guessing raises doubts about its usefulness.

3.3.2.2 The Word Associates Test

Based on the results of the Birbeck Vocabulary Project and suggestions from Paul Meara, Read (1993, 2000) decided to develop a word association test where testees are presented with a stimulus word and a set of other words from which they are asked to identify any related words, as in the example below:

sudden

beautiful quick surprising thirsty	change doctor noise school

In the Word Associates Test (Read, 2000) stimulus words are restricted to adjectives, the words on the left are other adjectives some of which are either synonyms of the stimulus word or represent one aspect of its meaning (paradigmatic relationship). Among the words on the right there are nouns that frequently collocate with the stimulus word, thus having a syntagmatic relationship with it. To reduce the chance of guessing the number of associates varies for each item. It focuses on high-frequency words, it is monolingual, involves recognition, and words are presented in isolation from context to allow for a large number of items to be tested. However, Read (2000) admits that the test does not perfectly simulate the way native speakers access their mental lexicons. Drawing on Read's (2000) test Qian (2002) devised the Depth of Vocabulary Knowledge Measure (DVK), a revised version of the Word Associates Test, and used it for measuring the relationship between vocabulary knowledge and reading performance.

As the Word Associates Test (Read, 2000) was originally designed to be used with university students, Schoonen and Verhallen (2008) adapted the test format for young children of primary-school age in their Word Association Test. They simplified the

original test in three ways. First, children are presented only six options instead of eight. Second, drawing on children's visual skills, associated words are arranged in a way that children are asked to draw a line between the target word and the associated words. Third, the number of associations is kept fixed, namely three, for each item. The authors claim that the " intended correct answers represent different kinds of relationships that can be found in a semantic network, such as paradigmatic relations (i.e. superordination, subordination, synonymy), partonomic relations (constituents), decontextualized syntagmatic relations (defining), perceptual features, inherent characteristics and/or means-aim relations" (Schoonen & Verhallen, 2008, p. 219)

As is clear from the above, none of the aforementioned vocabulary tests has proved to be adequate for all purposes of testing the quality of vocabulary knowledge. A pioneering and later widely cited method was developed by Sima Paribakht and Marjorie Wesche (1997) at the University of Ottawa in Canada to measure the depth of L2 vocabulary knowledge. The *Vocabulary Knowledge Scale* (VKS) consists of two scales, one for eliciting answers from the testees and another to evaluate the answers. Words are presented in isolation from context, it is discrete and selective. The five self-report categories vary from no previous encounter with the word through giving a correct synonym or translation to being able to use it in a sentence, with the latter category moving from receptive to productive vocabulary measurement (for a detailed discussion see section 3.5.3).

3.4 Corpora in vocabulary research: Word lists

According to Nation and Waring (1997), not all words are equally useful and a measure of usefulness may be how often a word is used, i.e. the frequency of the word. This aspect of language is explored by corpus linguistics, a dynamically developing field of applied linguistics. Massive corpora are being developed all over the world in order to help explore the nature and process of language use, such as the Brown Corpus, the LOB Corpus, the London-Lund Corpus or the Bank of English originating in the Main COBUILD Corpus, CANCODE, or the British National Corpus (for

introductions to the field see Biber, Conrad & Reppen, 1998; Flowerdew, 2003, 2005; Granger, Hung, & Petch-Tyson, 2002; Hunston, 2002; O'Keefee, McCarthy, & Carter, 2007; Partington, 1998; Sinclair, 1991; Stubbs, 2001; Thomas & Short, 1996; Tribble, 2002). A learner corpus, on the other hand, can be beneficial for linguists, writing specialists and language teachers in three areas: "to collect evidence of language use, to serve as a basis of research, and to serve as a basis of innovative pedagogical application" (Horváth, 2001, p. 2; see also Csomay, 2005a, 2005b, 2006, 2007; and Flowerdew, 2003 for a US context).

The main question of interest for L2 learners and teachers is usually what and how much vocabulary a language learner needs. The selection of target words to be tested may take place on the basis of how frequently the word occurs in the language which is often interpreted as how useful the word is for the learner. Hence, there are four main categories (Read, 2000, p. 159):

- high-frequency vocabulary
- low-frequency vocabulary
- specialized vocabulary
- sub-technical vocabulary

The term high-frequency vocabulary is generally used for words that are the most common in a language, while low-frequency words occur relatively rarely used. Specialized vocabulary (also referred to as technical vocabulary by Nation & Newton, 1997) comprises those low-frequency words that appear frequently in the vocabulary of a specific field, i.e. a register of the language that may be important in English for specific purposes (ESP) courses, such as mathematics, law, and medical sciences. Read (2000) argues that specialized vocabulary is best acquired through content instruction by a subject teacher instead of general language instruction. Sub-technical vocabulary, on the other hand, refers to low-frequency words which occur frequently across a range of fields or registers, such as the vocabulary necessary for pursuing university studies in a range of subject areas. However, it is difficult to draw the boarderlines between these categories and determine where high-frequency vocabulary

ends and low-frequency vocabulary begins. "The division is arbitrary", although most researchers agree that "the distinction can be most usefully made somewhere between the most frequent 1,500 words and the most frequent 7,000 words (Nation & Newton, 1997, p. 239).

An intuitive judgement of the frequency of a word, however, may not be reliable. Studying the word frequency judgement of language teachers McCrostie (2007) emphasises the need for teachers to consult frequency lists rather than trusting their intuitions. To facilitate the process of teaching and learning foreign languages, as well as to provide sources for research and theory, a multitude of research was carried out aided by the newly emerging corpora, that has come to be called the Vocabulary Control Movement (Carter, 1998, p. 206). Various word lists have been produced drawing on large corpora of language. One of the oldest and most widely referred to such list is *The General Service List* (GSL) compiled by Michael West (1953). The GSL contains 2,000 head-words based on frequency patterns drawn from a written corpus of 5 million words, and according to Nation and Waring (1997), it is the classic list of high-frequency words practical for language teachers and curriculum planners. Some sample words from the first 500 most common words in the GSL are *man, face, change,* or *agree*.

The main criterion for the selection of vocabulary for the early stages of learning, as explained by Carter (1998), is the frequency information, i.e. the number of occurrences in the 5 million word corpus given by the GSL after each headword, and "the relative prominence of the various meanings and uses of a word form" (Carter, 1998, p. 206). The author claims that knowing these 2,000 words gives access to the meaning of 80 per cent of the words in any text which may be motivating for learners as "it has a relatively quick return" (Carter, 1998, p. 207), as shown by the percentage of text coverage of the first 6,000 words in the English language in Table 7 below:

Table 7. Frequency band and percentage of text coverage (Laufer & Nation, 1999, p. 36)

Frequency level	Cumulative coverage (%)	Coverage (%)
1st 1,000	72.0	72.0
2nd 1,000	79.7	7.7
3rd 1,000	84.0	4.3
4th 1,000	86.8	2.8
5th 1,000	88.7	1.9
6th 1,000	89.9	1.2

However, Carter (1998) explains that using the GSL for reference is not without disadvantages. First, the GSL may seem to be outdated, as it is based on word counts made in the 1930s; therefore, many common modern words are not included. Second, it is based on written corpora only, spoken language is excluded, as it was originally intended to be an aid in simple readings. Third, there is no information on possible collocations of the headwords and their frequencies, given the unavailability of concordance at the time. Although several attempts have been made to make it more up to date, the GSL is the most widely applied lists of core vocabulary.

Beyond the words in the GSL, another list of high frequency words specialised for general academic purposes, called the University Word List (Xue & Nation, 1984). The UWL as a list of academic words contains 800 headwords frequent in a wide range of texts that appear in secondary or senior high school, university and in newspapers (Nation & Newton, 1997). Based on the GSL and the UWL, Nation and Waring (1997) assume that any word that cannot be found in these two lists is to be considered a low-frequency word in the English language. Classroom implications suggest that for fast results in L2 learning the 2,000 GSL words should be among the words acquired first (Nation, 1990; Nation & Newton, 1997).

A more recent and widely used corpus-based word-list compilation is the Academic Word List (AWL) (Coxhead, 2000) superseding the University Word List. It

comprises 570 word families that occur most frequently across a carefully selected range of academic texts. Some most frequent sample words form the AWL are *analyse, benefit, approach, assume, involve* or *require*. The words in the AWL were supplemented with information on word stress patterns and their frequency by Murphy and Kandil (2004) in an attempt to provide EAP students with an aid to recognize, identify and use polysyllabic words.

Not all researchers share enthusiasm for such lists of sub-technical vocabulary though. To examine the value of Coxhead's (2000) Academic Word List, Hyland and Tse (2007) set out to explore the distribution of the 570 AWL families in a 3.3 million word corpus of texts from different genres and disciplines. They conclude that the AWL words behave differently across disciplines and recommend that teachers and learners should focus more on discipline specific vocabulary. Alderson (2007), comparing the frequency judgements of professional linguists to corpus-based frequency counts, goes even further and questions either the accuracy of highly educated native speakers' frequency estimates of words in their language or large modern corpora being adequate indicators of word frequency and claims that further research is necessary about the nature of intuitions of word frequency.

3.5 Quantitative measures

Quantitative measures of vocabulary operate along the partial-precise or the receptive-productive continua to describe vocabulary size. Laufer (1998) acknowledges that the learning of a word usually progresses from receptive to productive knowledge as suggested by Henriksen (1999) and claims that there is no one single instrument measuring both the size and the depth of vocabulary yet. However, she further refines the stages of the continuum by investigating three components of word knowledge: (1) basic receptive (passive) knowledge, i.e. "understanding the most frequent and core meaning of a word" (Laufer, 1998, p. 257); (2) controlled productive knowledge, which entails producing words prompted by a task such as filling in a sentence with a word the first few letters of which are given; and (3) free productive knowledge, the

use of words on one's own will or choice, like in the case of a writing task.

Another distinction causing a lot of confusion and misunderstanding in research (Read (2000, pp. 154-157) is recall and recognition. In order to overcome this confusion Laufer, Elder, Hill and Congdon (2004) developed a taxonomy of two dichotomous distinctions presented in Table 8. Their taxonomy clearly distinguishes being able to supply the form of a concept vs. supplying the meaning for a given form, and being able to recall a form or meaning vs. the recognition of it if presented.

Table 8. Types of vocabulary knowledge (Laufer *et al*., 2004, p. 206)

	Recall	Recognition
Active (Productive) (retrieval of form)	Active recall	Active recognition
Passive (Receptive) (retrieval of meaning)	Passive recall	Passive recognition

Exploring the relationship of receptive and productive vocabulary Laufer (2005a, p. 232) reports the results of a study with two tests (Table 9) taken by the same learners, where participants had to give an L1 equivalent to L2 words for passive recall, and provide an L2 translation for L1 words for testing active vocabulary use. The results show that the passive vocabulary is notably bigger than the active vocabulary and the distance is growing with the decrease of word frequency, which implies that the learners recognized the less frequent words as known but were not able to use them.

As we have seen in the previous chapter, vocabulary instruction and the proficiency level of the learners have a substantial effect on vocabulary size. Laufer and Paribakht (1998) found that the gap between receptive and productive vocabulary was smaller for EFL than for ESL learners. This may imply that explicit learning might result in deeper knowledge of words and greater gains in productive knowledge than incidental learning of words. In a more recent study Webb's (2008) findings support Laufer's

(2005a) assumptions.

Table 9. Passive and active vocabulary (Laufer, 2005a, p. 232) (scores are out of 30)

Modality level	Active/Passive ratio	Active recall	Passive Recall
2000	35%	M= 6.51	M= 18.61
n=134		SD=5.26	SD= 8.43
3000	30%	M= 5.35	M= 17.58
n=106		SD=3.33	SD= 5.99
Academic	24%	M= 3.18	M= 13.07
Word List		SD=3.04	SD= 8.43
n=82			
5000	16%	M= 2.50	M= 15.60
n=113		SD=2.88	SD=6.88

On the one hand, "participants with higher receptive vocabulary scores had greater ability demonstrating productive knowledge of L2 forms than those with lower receptive scores" (Webb, 2008, p. 92). On the other hand, both receptive and productive scores decreased with the decrease of word frequency and the gap between the two types of knowledge increased. As Webb (2008) concludes, these findings suggests that receptive knowledge of meaning precedes productive knowledge of meaning and learners are likely to have larger receptive than productive vocabularies.

It is doubtful, however, that using translation only for measuring vocabulary knowledge is a suitable tool. Translation is a transfer skill that can be developed and even though in the previous chapter we have seen that the form-meaning link between the L1 and the L2 may facilitate vocabulary learning, a communicative language learning context, where the L1-L2 link is not exploited, might have yielded different results. Similarly, applying a variety of vocabulary tests or a test that combines translation with other ways of demonstrating degrees of word knowledge, such as the Vocabulary Knowledge Scale would perhaps produce more reliable data.

3.5.1 Measures of receptive vocabulary size

It has long been a challenge for research to estimate how many words native speakers know and in the EFL context the figure would inform language teachers about possible targets for language learners. Using a commonly applied procedure, dictionary sampling, Goulden, Nation and Read (1990) estimated that "well-educated adult native speakers of English have a receptive vocabulary of around 17,000 base words" (1990, p. 341), excluding proper nouns, compound words and abbreviations. Based on similar studies by D'Anna, Zechmeister and Hall (1991) and Nusbaum, Pisoni and Davis (1984), Zechmeister, D'Anna, Hall, Paus and Smith (1993) concluded that the vocabulary size of an average university undergraduate is in the range of 14,000 – 17,000 words, thus the receptive knowledge of 14,000 words should be sufficient to pursue academic studies. However, Hazenberg and Hulstijn (1996) warn that fourteen thousand words may not be a minimal or optimal figure for a non-native speaker.

In an EFL or English for Academic Purposes (EAP) context information on how many and what words language learners know has been shown to be predictive of their academic success (Morris & Cobb, 2003) and a measure of academic success may be how well students are able to cope with the reading load. In the previous chapter we have reviewed research on text coverage and the vocabulary size required to be able to read authentic texts. Findings imply that vocabulary of 5,000 words would sufficiently equip a non-native speaker student for university studies, although Laufer (1998) argues that 5,000 is a "bottom line" for reading English at an academic level, whereas in their study with native speakers of Dutch Hazenberg and Hulstijn (1996) concluded that a minimal receptive vocabulary of 10,000 base words is necessary for university studies.

To measure the size of receptive vocabulary, Goulden *et al.* (1990) describe the method of dictionary sampling as a reliable way of determining the size of L1 vocabulary. A representative proportion of words is drawn from a dictionary and

learners are tested on their knowledge of these words. Then, "the proportion of the words they know in the test is taken as the proportion of the words they know in the whole dictionary" (Goulden *et al*. 1990, p. 343). Nation (1993a) provides a detailed description of what steps to follow in dictionary sampling as he argues that most of the earlier estimates of vocabulary size yielded unreliable results for not following the necessary procedures. The essential steps of dictionary sampling are the following (Nation, 1993a):

1) Choose a dictionary that is big enough to cover the known vocabulary of the people being investigated.
2. Use a reliable way of discovering the total number of entries in the dictionary.
3. Use explicit criteria for deciding and stating (a) what items will not be included in the count, and (b) what will be regarded as members of a word family.
4. Use a sampling procedure that is not biased towards items which occupy more space and have more entries.
5. Choose a sample that is large enough to allow an estimate of vocabulary size that can be given with a reasonable degree of confidence.
6. The sampling should be checked for the reliability of the application of the criteria for exclusion and inclusion of items.
7. The sample should be checked against a frequency list to make sure that there is no bias in the sampling towards high-frequency items.
8. In the written report of the study, describe clearly and explicitly how each of the previous seven procedures was followed in sufficient detail to allow replication of any or all of the procedures.

The receptive vocabulary size test Goulden *et al.* (1990) developed is a self-report pen-and-pencil test where the test-takers tick words in a 50 item list either as known or not known. The test does not measure degrees of word knowledge and no proof of word knowledge is required in any form; therefore, a drawback of this test format is the high probability of test-takers overrating their knowledge. As an alternative, the computerized version of the check-list format, the Eurocentres Vocabulary Size Test (EVST) developed by Meara and Jones in 1990 as a placement test contains built-in non-words (or pseudo words) as well to filter those students who tend to overestimate their knowledge in the self-report yes/no response test.

The yes/no test is easy to construct, administer and correct and allows for a lot of items to be tested in a relatively short time. The format has recently been applied in the DIALANG project developed for assessing language proficiency (www.dialang.org). However, the application of pseudo words gave ground to a lot of criticism on the test format. Pseudo words are "words that fulfill the phonological constraints of the language but do not bear meaning" (Huibregtse, Admiraal, & Meara, 2002, p. 227). In order to calculate a meaningful score on a yes/no test a correction for guessing by means of mathematical formula has to be applied that Meara and his colleagues named m. The same formula was used in Meara's (1992) EFL Vocabulary Test later revised in Huibregtse, Admiraal and Meara (2002). Due to several shortcomings of the correction scheme Beeckmans, Eyckmans, Janssens, Dufranne, and Van de Velde (2001) question the validity and reliability of the test, whereas Mochida and Harrington (2006) argue that the yes/no test scores are good predictors of scores on the Vocabulary Levels Test (VLT, Nation, 1990, see in section 3.5.2 in detail) and the yes/no test is a valid measure of the type of receptive vocabulary knowledge measured by the VLT.

3.5.2 Measures of productive vocabulary use

As we have seen in the first chapter, the notion of a word is difficult to define in theoretical terms. Linguists prefer distinguishing between lexical items, lexemes, and word families, but from the second language (L2) teacher's perspective and applying a top-down introspection, a written text consists of shorter units divided by space and these are what we call words in the everyday sense. As the easiest way to approach a text quantitatively is to count the words in it, the total number of words is labelled by the term token. However, there are a lot of recurring words in any text; therefore, another distinctive measure has to be introduced for the purpose of description, the number of different words in the text, i.e. type. Obviously, the number of tokens will always be higher than the number of types, as there are certain grammatical or function words (articles, auxiliaries, prepositions, pronouns, conjunctions, etc.) that are more frequently used than content words (nouns, verbs, adjectives, adverbs, etc.).

Thus, the type-token ratio can serve as a measure of vocabulary use which may also provide information on vocabulary development in longitudinal studies.

The lexical diversity of writing or lexical richness has four major components by which a reader may assess the lexical features of a text: (1) the type-token ratio (or lexical variation), (2) lexical sophistication, (3) lexical density, and (4) number of errors, as defined by Read (2000). The *type-token ratio* (T/t) suggests that a good writer uses "a variety of different words rather than a limited number of words used repetitively" (Read, 2000, p. 200). Statistically, this ratio is calculated by dividing the number of types by the number of tokens. Lexical density is calculated as the total number of lexical words divided by the number of tokens:

Lexical density:

$$LD = \frac{\text{the total number of lexical words}}{\text{the total number of words in the text}}$$

The term *lexical sophistication* suggests that the writer is able to use a variety of relatively low-frequency words rather than words of high frequency, as explained by Read (2000) and is calculated as follows:

Lexical sophistication:

$$LS = \frac{\text{the number of low frequency lexemes}}{\text{the total number of lexemes in the text}}$$

Using the type-token ratio for assessing lexical diversity has been criticised for numerous reasons, most prominently for being dependent on text length. Jarvis, in a 2002 study overviewing measures of lexical diversity and applying curve fitting procedures to compensate for varying text length, argues that certain algebraic transformations of the type-token ration, such as the Herdan, Guiraud or Uber index

prove to be more accurate measures less sensitive to text length:

Herdan index: C= log Type / log Token

Guiraud index: R = Type / Token

Uber index: $U = \dfrac{(\log \text{Token})^2}{(\log \text{Token} - \log \text{Type})}$

3.6 Major vocabulary tests

This section provides an overview of the four major vocabulary tests relevant in the empirical chapters of the present dissertation, each tapping into different levels and aspects of word knowledge. First, the Vocabulary Knowledge Scale (Paribakht & Wesche, 1997) is discussed as a tool for measuring the depth dimension of lexical knowledge. The remaining three tools are used for quantitative measurement at varying levels. The Vocabulary Levels Test (Nation, 1990; Schmitt, Schmitt & Clapham, 2001) estimates receptive vocabulary size at five word frequency levels. The Test of Controlled Productive Ability (Laufer & Nation, 1999) is designed to assess the test-takers capability of producing words at various frequency levels prompted by a gap-filling task. Finally, the Lexical Frequency Profile (Laufer & Nation, 1995) tests free productive vocabulary by calculating the lexical richness of any input text.

3.6.1 The Vocabulary Knowledge Scale

Mention has been made of the Vocabulary Knowledge Scale, a test of vocabulary depth to "distinguish stages in learners' developing knowledge of particular words" (Paribakht & Wesche, 1997, p. 179). Test takers are presented a list of target words in isolation and they are asked to fill in the self-report categories. The Vocabulary Knowledge Scale elicitation categories are as follows (see Table 10):

Table 10. The VKS elicitation scale – self-report categories (Paribakht & Wesche, 1997, p. 180)

I. I don't remember having seen this word before.

II. I have seen this word before, but I don't know what it means.

III. I have seen this word before, and I <u>think</u> it means _____ (synonym or translation)

IV. I <u>know</u> this word. It means _____ (synonym or translation)

V. I can use this word in a sentence: _____ (Write a sentence.)

    (If you do this section, please also do Section IV.)

Testees tick category one or two if they have never seen the word in the list before or the word seems familiar but they do not know its meaning. If they dare to guess, they may put a L1 equivalent or an English definition into column III. If testees feel certain about their knowledge of the word, column IV is to be filled in similarly. To demonstrate an even deeper or more precise knowledge of the word, testees may choose to write a sentence with the word in column V. Scores on the test are then calculated with the help of a scoring scale as shown in Table 11.

The scoring of the test accepts self-report categories I and II for scores 1 and 2, while the other categories require a demonstration of knowledge for higher scores. Score 3 is given for a correct synonym or L1 equivalent. A score of 4 is given when the test taker demonstrates knowledge of the target word used in a sentence in a particular context with incorrect grammar, whereas a score of 5 reflects a semantically appropriate and grammatically accurate use of the target item. Although the researchers proved that the scale was a reliable tool for measuring depth of word knowledge, it raises several questions. One concerns the theory behind the self-report categories. Considering the three dimensions of Henriksen (1999), categories I-IV operate along the partial-precise continuum, while category V shifts to the productive-receptive dimension of word knowledge. Given this, Read (2000, p. 136) doubts that "a learner's developing knowledge can be meaningfully represented by one single scale".

Table 11. VKS scoring categories – meaning of scores (Paribakht & Wesche, 1997, p. 181)

Self-report categories	Possible scores	Meaning of scores
I	1	The word is not familiar at all.
II	2	The word is familiar but its meaning is not known.
III	3	A correct synonym or translation is given.
IV.	4	The word is used with semantic appropriateness in a sentence.
V.	5	The word is used with semantic appropriateness and grammatical accuracy in a sentence.

Qian and Schedl (2004) also point out that the VKS does not assess multiple meanings of a word. Read (2000) refers to a suggestion to the authors to include a sixth category asking the testee to demonstrate all the possible meanings of the word they can think of. On the other hand, sentence writing seems to be problematic as learners may write semantically appropriate and grammatically correct sentences without a thorough understanding of the target word as checked in subsequent interviews. Therefore, Read (2000, p. 138) argues that the VKS is an "interesting effort to measure some aspects of word knowledge in a practical way", but limitations can be expected if we attempt to reduce the complexity of word knowledge into a single scale, and he recommends using multiple scales.

3.6.2 The Vocabulary Levels Test

The Vocabulary Levels Test (VLT) is a discrete point, receptive vocabulary size test aiming to tap into partial lexical knowledge. It was first developed by Paul Nation (1990) at the Victoria University of Wellington as a diagnostic test for teachers and was later revised, expanded and validated by Schmitt, Schmitt, and Clapham (2001).

This widely used test of vocabulary breadth provides an estimate of how many words a candidate knows on five frequency levels of English word knowledge: the first 2,000, 3,000, 5,000 and 10,000 most frequent words on the basis of large corpora of texts. The fifth level, however, is not based on frequency data, it comprises a representative sample of words from the University Word List (Xue & Nation, 1984) in Schmitt *et al.*'s (2000) revised version, i.e. words that are most common in academic texts from a wide range of fields and subjects.

Each level consists of six clusters of six words that need to be matched with three synonyms or short definitions. Below is an example of a noun cluster from the academic word level of the VLT (with key):

1 area
2 contract   __2___ written agreement
3 definition   __5___ way of doing something
4 evidence   __4___ reason for believing
5 method   something is or is not true
6 role

Each level is carefully designed to serve specific language learning objectives. The 2,000-3,000 levels comprise high frequency words that all learners need to know to communicate effectively in everyday situations and begin to read authentic texts. The 5,000 word level is Laufer's (1992b) minimal vocabulary size to comprehend unsimplified authentic texts and signifies the upper limit of general high frequency words, while the 10,000 word level contains the most common low-frequency words. As Hasenberg and Hulstijn (1996) found, this is the vocabulary size necessary for pursuing university studies. The Academic word level, however, is not based on general word frequency and Schmitt *et al.* recommend to place it between the 3,000 and the 5,000 word level sections for balanced difficulty and pedagogical reasons.

Among other crucial considerations of test design, the definitions are short to require minimal reading comprehension. The words in the definitions were deliberately kept simple and always more frequent than the target words, from the first 2,000 words of

the GSL. The target words in the clusters are listed in alphabetical order and the definitions are in order of length to minimize guessing. The words in the clusters were chosen from the same grammatical class for the same reasons and orthographic similarities were avoided.

Comparing the Vocabulary Levels Test and the Yes/No test (Meara, 1992) in an English as an additional language context Cameron (2002) found that the Levels Test was a more useful tool, as the inclusion of non-words in the Yes/No test produced unreliable results. Her study shows that the Levels Test is a useful research and pedagogic tool in that context to "yield an overall picture of receptive vocabulary knowledge across groups" (Cameron, 2002, p. 145).

3.6.3 The Test of Controlled Productive Ability

The Test of Controlled Productive Ability (TCPA) is based on the same considerations about the importance of word frequency as the Vocabulary Levels Test. It was developed and validated by Laufer and Nation (1999) as a diagnostic, discrete point, selective, context dependent test of productive word knowledge. The main difference between controlled productive and free productive knowledge is explained by Laufer and Nation (1999, p. 37) as follows:

> Productive vocabulary ability is not a yes/no phenomenon, but implies degrees of knowledge. For example, a learner may be able to provide a sentence with an infrequent word when required to do so by the teacher, but be reluctant to use it when left to his own devices, as in a composition writing task and choose to use a simpler, more frequent word of a similar meaning.

Items in the tests were chosen to represent a certain frequency level. The test format is similar to a C-test, but uses unrelated sentences and Laufer and Nation (1999) explain that the number of letters given of the target word was decided on to eliminate any other words to fit the sentence but still be minimal to disambiguate the meaning of the word. The overall structure of the test models that of the Vocabulary Levels Test (Nation, 1990). It consists of 18 items at each of the 2,000, 3,000, 5,000, university

word level and the 10,000 word levels; thus, the maximum score is 90. Four parallel versions were devised based on the VLT versions by Schmitt *et al.* (2001); therefore, if the testees are tested on both the receptive VLT and the productive TCPA, or on one of them twice to get a more reliable picture of learner vocabulary knowledge, the same items do not reappear. Below is a sample item for the word *episodes* from the test:

The book covers a series of isolated epis_____ from history.

However, it has to be noted here the TCPA has been criticized on what the blank-filling test is measuring. Read (2000) argues that the test items vary in demands on the test taker due to the differing number of letters provided for each item: some require more word knowledge and more use of contextual cues than other words. He also claims that Laufer and Nation (1999) have little evidence to support their claim that the test scores reflect the number of words available for the learner for productive use at a certain frequency level and argues that the TCPA is simply "an alternative way of assessing receptive knowledge rather than a measure of productive ability" (Read, 2000, p. 126).

3.6.4 The Lexical Frequency Profile

Based on the General Service List (West, 1953) and the University Word List (Xue & Nation, 1984) Laufer and Nation (1995) proposed a new measure of lexical richness in writing, called the Lexical Frequency Profile (LFP). The authors claim that this reliable and objective tool excluding subjective judgements in the assessment of writing quality can be useful in measuring how vocabulary size is reflected in use, as it shows the "relative proportion of words from different frequency levels" (Laufer & Nation, 1995, p. 311) in any samples of writing. Laufer (2001) explains that the profile is topic independent, i.e. it is stable along compositions written by the same learners provided that the topic is general in nature and does not require the use of infrequent jargon.

For instance, the vocabulary of a composition of 300 tokens may distribute as follows. 270 word types belong to the first 1,000 words of the General Service List, twelve come from the second 1,000 words of the GSL, nine types are listed in the University Word List (Xue & Nation, 1984), and the remaining nine words are relatively low-frequency words not found in any of the aforementioned three lists. Converting these figures into percentages, the Lexical Frequency Profile of the text is 90%-4%-3%-3%. Consequently, the ratio of 'beyond-2,000' low-frequency words in this text sample is 3%+3% = 6%, which is found to be a better indicator of progress in vocabulary knowledge (Laufer, 1994, cited in Schmitt, 2000, p. 177). The calculation is done by a computer program named RANGE downloadable free from Paul Nation's website (Lehmann, 2003a).

The software developed for Windows-based PCs, named RANGE, is able to do the analysis based on three pre-set word lists, and it also provides the possibility to develop the users' own new word lists for comparison. The first base word list contains the first 1,000 words of the GSL, the second base word list comprises the second 1,000 words of the GSL, and the third list is the University Word List (in a more recent version the Academic Word List by Coxhead, 2000). Among its multiple possible uses the software matches the input texts, e.g., learner compositions or any type of corpus compiled by the user up to a million tokens, to these base word lists and calculates frequency distributions for tokens, types and word families and lists the input words arranged by order of frequency (Lehmann, 2004).

A fragment of a sample output of an analysis on a student composition discussed in the empirical chapters is shown in Figure 5. For the analysis the input texts need to be typed into the computer and saved into the RANGE folder in text only format. To get an undistorted profile of learner vocabulary the authors recommend that misspelt words should be corrected, proper nouns that cannot be taken as part of the lexis of a language should be omitted together with semantically incorrect words, as they cannot be regarded as known.

```
Processing file: D:\Range\CM14.txt
  Number of lines: 5
  Number of words: 336


  Reading: D:\Range\BASEWRD1.txt
  Reading: D:\Range\BASEWRD2.txt
  Reading: D:\Range\BASEWRD3.txt
```

WORD LIST	TOKENS/%	TYPES/%	FAMILIES
one	303/90.2	125/81.2	100
two	13/ 3.9	13/ 8.4	12
three	9/ 2.7	6/ 3.9	5
not in the lists	11/ 3.3	10/ 6.5	?????
Total	336	154	117

```
 Number of BASEWRD1.txt types: 4119  Number of BASEWRD1.txt families: 998
 Number of BASEWRD2.txt types: 3708  Number of BASEWRD2.txt families: 987
 Number of BASEWRD3.txt types: 3107  Number of BASEWRD3.txt families: 570
```

Types Found In Base List One

TYPE	RANGE	FREQ	F1
THE	1	22	22
IT	1	15	15
IS	1	13	13
AND	1	10	10
THAT	1	10	10
TO	1	10	10
A	1	9	9

Figure 5. Sample RANGE output

A complete website is dedicated to the free dissemination of knowledge and providing all the currently available technical assistance for conducting vocabulary research both for teachers and academics. The site is run by Tom Cobb (2000) and is named The Compleat Lexical Tutor. Besides offering the possibility of running concordancers, building corpora and calculating various statistical measures of vocabulary, it hosts a recent version of RANGE, and features VocabProfile, an updated, further developed, very practical and user friendly version of the original profiler. A sample output of VocabProfile is shown below (Figure 6):

Output text: buck did not read the newspapers or he would have known that trouble was brewing not only for himself but for every tide water dog strong of muscle and with warm long hair from puget sound to san diego

Breakdown:

1k types: [families 27 : types 29 : tokens 31] and_[1] buck_[1] but_[1] did_[1] dog_[1] every_[1] for_[2] from_[1] have_[1] he_[1] himself_[1] known_[1] long_[1] newspapers_[1] not_[2] of_[1] only_[1] or_[1] read_[1] sound_[1] strong_[1] that_[1] the_[1] to_[1] trouble_[1] was_[1] water_[1] with_[1] would_[1]

2k types [3:3:3] hair_[1] tide_[1] warm_[1]

OFF types [?:5:5] brewing_[1] diego_[1] muscle_[1] puget_[1] san_[1]

Figure 6. Sample VocabProfile output

The software is able to do the same calculations as RANGE to sort words in input texts into different frequency bands for a vocabulary profile. However, it is extended with features such as colour coding words from different frequency levels in the input text, giving the number of content words and function words, calculating the type-token ratio and lexical density right away, as well as providing profile information on words of Anglo-Saxon origin and even calculating Greco-Latin/French cognate indices.

3.7 Conclusion

This chapter overviewed major issues in vocabulary assessment centering around the intertwined questions of why to test vocabulary, what words to test, what aspects of word knowledge to test and how to test it. It has been emphasized that the choice of the test format should always depend on the type of information the users of the test desire to gain and the purpose of testing needs to be established before making any choice of test format to be used. We have seen that no test at present is able to assess all possible features of word knowledge. The choice of the teacher or the researcher is often a trade-off between the number of items to be tested, the simplicity of the

procedure and the amount of information about word knowledge gained from the test.

Applying several different tests tapping into different dimensions of word knowledge, however, is argued to provide a more comprehensive picture of vocabulary knowledge. Adopting a multiple tests approach (Laufer & Nation, 1999, p. 45) allows the study of how the different aspects of word knowledge are related to one another, how vocabulary knowledge develops and how the relationships between various dimensions change over time, and how the input and the teaching method influence the development of the different aspects of word knowledge.

Part II

Empirical studies exploring English majors' vocabulary and their
needs based on their assessed vocabulary in the light of
curricular requirements at UP

Chapter 4
Background to the empirical studies

4.1 Introduction

T he previous chapters in Part I introduced major issues relevant to the context of vocabulary research. In Chapter 1 we have seen that the description of lexis and the mental lexicon is a complex endeavour and a widely debated field where issues of defining what a word is, what knowing a word involves and how words are represented, stored and retrieved from the mental lexicon have yet to be further explored. The second chapter provided an overview of research on how L1, and more specifically L2 vocabulary is learnt and used. Recent publications in the field debate questions of intentional or incidental vocabulary acquisition being more efficient or beneficial for the learners.

Empirical evidence converge in indicating that a combination of the two may give a very good return for learner efforts, whereby the facilitating role of attention to individual differences and the effects of strategy training are by no means to be ignored. Chapter 3 has shown that vocabulary assessment has been a major focus of research in the last few decades and an abundance of materials have been developed to

measure either the size or depth of learner vocabulary, constituting the two main directions of approaching the field. It emerges from current EFL/ESL vocabulary studies that language is more lexical than grammatical and that vocabulary is a good predictor of success on reading comprehension tests, and thus indirectly and in a wider context, of academic studies.

This latter issue gave the rationale for further research into vocabulary. The subsequent three chapters report the four developmental stages of English for academic purposes (EAP) vocabulary testing at UP. In Chapter 5 three pilot studies are described which initiated further research into the lexical needs and lexical knowledge of students of English. Chapter 6 addresses issues emerging in testing specialized vocabulary for academic purposes as well as questions related to vocabulary as a construct in assessing English language proficiency. Finally, Chapter 7 reports an attempt to overcome the burdens of widely used tests of vocabulary exposed on the present context by introducing a specialized corpus of compulsory readings in English studies at UP (CORES), as well as a new corpus-based test battery designed to meet the special lexical needs of Hungarian English majors. As this dissertation deals with the lexical needs of a special group of Hungarian learners, university students of English, it seems sensible to introduce the specificities of their context, the curriculum they follow and the proficiency milestone the battery developed by the researcher to assess their vocabulary knowledge is embedded in before moving on in our discussion of the empirical studies.

4.2 The research context

The political and educational context of language teaching in Hungary has undergone dramatic changes in the past thirty years. In the 1980s political change necessitated drastic reforms in the system of education in the new context. This put an end to the strong communist control over the curriculum by giving more autonomy to schools and teachers in the choice of teaching methodology, as well as of materials and contributed to the development of the present three-tier system of education policy.

However, the insecure social and political situation and the next decade of governmental policy with innovations and several versions of the *Hungarian National Core Curriculum* (Version 1: 1990, Version 2: 1992, Version 3: 1995, and Version 4: 2003) caused a feeling of insecurity in both teachers and learners. On the one hand, a nationwide representative survey (Nikolov, 1999) has shown that governmental policy has had little impact on what actually goes on in the classroom. On the other hand, international comparative studies (PISA, 2001) have found a drop in the academic performance of Hungarian learners compared to learners in other European countries since the 1980s and this downward trend seems to be continuing today.

The *National Core Curriculum* (NCC, Nemzeti Alaptanterv, 2003) is the highest level document regulating education in Hungary. It aims to lay down common principles for all levels of education and at the same time to provide autonomy to schools in selecting their own educational content. It describes fundamental knowledge and skills to be developed and provides guidelines on developmental aspects and how to spread content over various phases of education. On the second level, frame curricula for the various subjects provide more detailed outlines and recommendations for schools, teachers, textbook writers and developers of national tests.

Modern foreign languages in the NCC (2003) are seen as key conveyors of change. Russian has ceased to be the primary and compulsory foreign language to be taught and with the accession to the European Union in 2004, all documents concerning language education "have been created to be 'euroconform', have adopted a functional-notional syllabus, and have advocated humanistic and communicative principles of education" (Medgyes & Nikolov, 2002, p. 203).

Although language learners may decide what foreign language they want to learn and what level they wish to achieve by their school-leaving age, this choice is often restricted by local constraints, determined by local education boards and schools, as well as by peer needs. The National Curriculum monitors the process of language acquisition by prescribing the levels of language proficiency to be achieved, and

together with the frame curriculum prescribes language input and output at various stages of the process. The evaluation of what has been acquired is also monitored by the NCC through school-leaving examinations at two levels determining admittance to higher education (Sifakis, Oder, Lehmann & Bluma, 2007).

4.3 Setting and participants

Only a few studies have undertaken research on the language needs of English majors in Hungary (Kormos, Hegybíró-Kontra & Csölle, 2002; Nagy, 2008); even fewer have examined the lexical knowledge of this special group of learners (Doró, 2007, 2008; Lehmann, 2007b, 2007c, 2006a, 2006b, 2003b; Ötvösné Vadnai, 2002). Kormos *et al*. (2002) investigated the language wants of a representative sample of students of English language and literature at six universities in Hungary and as students from the University of Pécs were also involved, their findings are relevant in the present study. Kormos *et al*. (2002) found that one third of the students become teachers after graduation (31.3%), a quarter of them go into business (25%), and one in ten works as a translator (11.3%). Other jobs taken involve post-graduate studies, educational manager, journalist or IT. They conclude that students seem to use their university studies as springboards which implies for curriculum designers that a basic aim of training English majors should be to develop their general language competence to be able to use the language competently in a wide variety of jobs.

Concerning the language learning needs of the students, Kormos *et al*. (2002) reported that English majors listen to lectures and take notes on a daily basis. They listen to presentations by their peers, express their opinions verbally, use dictionaries and read professional books once or twice a week. They read professional journals, texts on the Internet and take notes while reading only once or twice a month, and hold presentations, write essays and read fiction in English with a monthly frequency only. This latter result is unexpected to teachers in the English departments as it suggests that as opposed to the weekly reading requirements students either do not read the compulsory readings regularly or read them in Hungarian. As we have seen in the previous chapters, a source of incidental vocabulary learning is extensive reading,

therefore this finding may imply that students miss or do not use this opportunity of vocabulary development to its full potential.

4.3.1 The curriculum

The University of Pécs (UP) is located in the South-West of Hungary. The city stretching at the foot of the Mecsek Hill is the fifth largest town in the country with its 156,664 inhabitants (http://hu.wikipedia.org/wiki/Pécs). In 2008 over 30,000 students pursued studies in the ten faculties of UP (http://www.pte.hu/files/tiny_mce/File/adatok/2008-oktoberi-statisztika-rovid.pdf). The Department of English Applied Linguistics is incorporated in the Institute of English Studies at the Faculty of Humanities on the Ifjúság street campus.

The curriculum of English studies has undergone considerable changes in the past decade due to the Bologna process. At the time of my initial data collection in 2002 English majors were to earn 140 kredits in eight semesters. The number of contact ours was 1,350 and the ratio of practical courses was 43 percent of all courses. Students may have selected one of two tracks in the course of their studies either to get a single degree in English language and literature, equivalent to an MA degree in the European system according the 1993 Higher Education Act, or take up an additional teacher training module to their English major to become teachers of English as a foreign language in addition to the MA degree. This additional degree allowed them to teach English at any school of primary, secondary or tertiary education. Students were expected to be at a near native level of proficiency (C1 of the CEFR, 2001, levels).

The time frame of university studies was ideally set for eight semesters, however, the majority of the students graduated in 10-12 semesters due to double majoring in other subjects in the natural sciences not following a credit-based curriculum at that time such as biology, geography or physics; or for taking gap years for child bearing, study abroad, or doing part-time jobs during their studies. Students were offered to select courses from three major fields within English studies: applied linguistics, linguistics

106

and English and American literature and culture, provided that they pass four major tests serving as milestones. The first such significant event as a prerequisite to complex exams in linguistics, literature and culture as well as the submission of the thesis was the Proficiency test due at the end of the second semester, on completion of introductory courses to the three major fields and two semesters of language skills development. Therefore, the following section is devoted to the description of the proficiency test.

4.3.2 The Proficiency test

The idea of testing the language production of first-year students at the Department of English Applied Linguistics was first raised in 1993. Various aspects of language testing and curricular requirements had been carefully considered before deciding on the aims and specifications of the initial test battery which went through numerous modifications until it reached its present format in 2002. The rationale for testing language proficiency, explains Szabó (1996), who has been a member of the testing team from the beginnings, was, on the one hand, the top priority of language development in the first two years at the Department. On the other hand, the fact that the first overall measurement of language performance took place only in the fourth semester of English studies at the first complex exam ("szigorlat"), necessitated getting information and giving feedback on the efforts and improvement of the students within this period and filtering students before this consequential examination.

The aim of the original test, which has not changed till the present day, was to select those students whose language level was below a carefully determined level; therefore, the test devised was a filter test. As it aimed to measure overall language achievement, not performance based on or compared to any syllabus, it was a criterion referenced proficiency test at the same time, filtering those who were not "proficient enough" to pursue further studies (Szabó, 1996). In its original form the battery intended to test all four skills (reading, writing, listening and speaking) with an additional grammar

component, however, in later stages of its development the idea of testing speaking was dropped due to its time consuming nature, the high number of candidates and the limited number of interlocutors resulting in unbridgeable reliability issues, as well as the fact that students' speaking skills were developed in the language practice courses where the instructors had a whole semester to continuously evaluate and incorporate speaking performance into the course grade, making the assessment of this skill more reliable (Szabó, 1996). Therefore, the speaking subtest was administered only once as a trial in the form of peer-to-peer interaction in an information gap exercise, further on it was omitted from all later versions of the Proficiency test, to be reintroduced in the BA programme, but not within our present scope of investigation.

4.3.2.1 Testing reading

In the construction of the reading subtest special attention is to be paid to the selection of the texts. Szabó (1996) refers to the richness of literature on the decisive impact of the text on test results and reliability and explains that among test parameters authenticity, familiarity and subject interest were to be primarily borne in mind. Testing techniques have shown a great variety in the history of the test, including multiple choice items for objectivity, ease and speed of administration and assessment, as well as suitability for computerized item analysis; information transfer; short answer; and sequencing (for problem areas see Szabó, 1996) to measure both micro- and macro-skills of reading.

Since I joined the team in 2000, two kinds of multiple matching tasks have been used, where "two sets of stimuli have to be matched against each other" (Alderson, 2000, p. 215); part one (with eight items) being an article where eight paragraphs have to be inserted to fit into the numbered gaps to measure macro skills of text comprehension. Similarly, in part two testees have the same task with missing parts of sentences to test micro skills of reading. It is important to note here the significance of providing more alternatives than the number of items required by the matching task to avoid the possibility that there is only one possible final choice at the end (Alderson, 2000).

4.3.2.2 Testing listening

Although the majority of the above techniques can be utilized in the design of the listening comprehension subtest, certain issues are more complex and thus require special attention on the part of the item writer. Apart from general text selection criteria Szabó (1996, 2008) emphasizes the somewhat different interpretation of authenticity in the case of a listening text, varying considerably in rate of delivery, accent, and clarity of recording. It was agreed that both American and British English native speakers were to be chosen, who speak "standard" English with "average" speed of delivery. Items involved multiple choice and multiple matching with careful spacing of information to avoid testing the memory of the candidates.

4.3.2.3 Testing writing

Writing is tested directly for the sake of validity and reliability. In the early versions of the subtest candidates were presented with a choice of topics to be selected from a three-column grid, where each column contained a part of the possible essay title to be chosen (for details see Horváth, 1996). More recent versions have remained student-friendly in this respect; however, the choice of topics has been restricted to two with four guiding points each, specifying what aspects of the topic to elaborate on in each paragraph of the 300-word essay. Multiple scoring based on analytic scales has been common practice right from the beginning for enhanced objectivity.

4.3.2.4 Grammar as a separate test component

Although grammar is not a skill, all four skills make use of it indirectly, therefore "it is legitimate to construct direct tests of grammar separately" (Szabó, 1996, p. 82). The basic principle of item writing with the grammar subtest was discrete point, rapid mass testing, he explains, with techniques and content selected on the basis of practicality, objectivity and suitability for statistical item analysis. The techniques and item types will be discussed in more detail later on.

4.3.2.5 Assessment and data analysis

The early versions of the Proficiency test used to be pass/fail tests; however, to meet new curricular requirements in 2003, evaluation had to conform to the 1-5 scale of marking. By constant and detailed item analysis the battery has continuously been improved, employing the latest available approaches to language testing and statistics to become as objective as possible. Every semester there are two optional test dates provided for the students, thus two tests need to be devised. Thorough item analysis procedures, relying on the Rasch-model within the Item Response Theory of practical language testing and the use of anchor items from previous tests ensure maintaining the same level of difficulty and making the comparison of test results possible from time to time (see Szabó, 2008). Pass marks have long been set on the basis of score distribution curves and the original idea of item banking has been developed to its pedagogical potential serving as a basis for recent test construction.

4.4 Overview of research methodology

Although the components of the above described proficiency test involve an embedded assessment of lexis, recent developments in vocabulary research discussed in part one of the present thesis inspired new insights and consequently, certain modifications to the old battery. Three empirical studies were designed for a deeper understanding of the lexical needs of English majors. The main research questions focus on the three major issues in vocabulary research introduced in the previous chapters: How many words and what words learners are required to know and how well they are required to know them to be able to comprehend the compulsory readings of their courses and thus pursue successful academic studies.

As has been emphasized in the previous chapter, the choice of appropriate tools for vocabulary assessment in a particular context with a particular purpose has to be a trade-off between the number of items to be tested, the simplicity of the procedure and

the amount of information about word knowledge gained from the test. Therefore, the research methodology applied in this thesis builds on quantitative analysis mainly, as it allows a relatively large number of items to be administered to a relatively large population in a relatively short time, these being the external constraints imposed on the new battery embedded in the frame of the well-established proficiency test. However, adhering to a mixed approach, a qualitative analysis of certain aspects of word knowledge is reported in a small-scale preliminary study. The empirical studies are presented in chronological order as well as in their sequence of development, for the findings of one served as a starting point for the other. The surveys are cross-sectional as data collection took place annually with different tests and participants.

The three pilot studies reported in Chapter 5, constituting initial explorations into the vocabulary size of English majors at UP, employ descriptive statistics and correlational analysis. These small-scale studies investigate how many words Hungarian students know in the English language, how their receptive vocabulary size may be reflected in their productive vocabulary use and whether intentional or incidental vocabulary learning yields better gains in the long-run. Chapter 6 introducing the initial three phases of the four-staged vocabulary test development process gains further insights from linear regression analysis about the initially surveyed relationship between receptive and productive vocabulary measures, as well as their relation to reading, writing and listening skills, general language proficiency and academic performance. Finally, in Stage 4 (Chapter 7) of the design of our vocabulary test battery the scope and methods of data analysis are further widened and refined by corpus analysis, vocabulary profiling and item analysis.

Chapter 5

How many words? Pilot studies exploring the receptive vocabulary size of first-year students

5.1 Introduction

The focus of this chapter is assessing the receptive vocabulary size of first-year students at the Department of English Applied Linguistics. All the three quantitative studies introduced in this chapter apply a vocabulary size test as a starting point but tap into different aspects of vocabulary knowledge discussed in Part I. The first study measured the receptive vocabulary size of 93 participants to estimate whether first-year students have the vocabulary size sufficient for pursuing academic studies. In the subsequent two sections the findings of two small-scale studies are reported, exploring the receptive-productive continuum of vocabulary knowledge and comparing the nature of incidental and intentional vocabulary learning involving qualitative analysis, respectively.

5.2 The receptive vocabulary size of first-year students

5.2.1 Research questions

The aim of this study was to explore the vocabulary size of English major students. Based on complaints voiced by both students and fellow colleagues as well as my own classroom experience I assumed that first- and second-year students find it difficult to understand the compulsory academic reading materials required for their courses. As research has shown a close relationship between vocabulary knowledge and reading comprehension (Alderson, 2000; Gass, 1999; Grabe & Stoller, 1997; Hulstijn, Hollander & Greidanus, 1996; Laufer, 1997a; Nation, 2006, Paribakht & Wesche,

1997; Quian, 2002; Quian & Schedl, 2004; Zareva, 2005), I hypothesized that the lack of sufficient vocabulary size for pursuing academic studies hindered reading comprehension (Hazenberg & Hulstijn, 1996; Hirsh & Nation, 1992; Laufer, 1987, 1992a, 1992b). The study is innovative for two reasons. On the one hand, it fills a gap in our information about the EFL vocabulary size of Hungarian students, as it is the first study reported to measure this aspect of vocabulary knowledge in a tertiary education context. On the other hand, the research instrument applied in the study combines the benefits and potentials of former test formats designed either for quantitative or qualitative measurement.

5.2.2 Participants

The participants were 93 first- and second-year English majors, 18 males and 75 females, all non-native speakers of English, attending the compulsory Language Practice seminars offered by the Department of English Applied Linguistics in the 2002/2003 academic year. At the outset of the study, their estimated level of proficiency was B2 on the six-level scale described by the *Common European Framework of Reference* (Council of Europe, 2001), based on the description of the Hungarian school leaving examination that all students had passed before being admitted to university (OKM, 2002).

5.2.3 Instruments

Test four of the Goulden *et al*. (1990) tests was adapted for the study and was extended with the Vocabulary Knowledge Scale (Paribakht & Wesche, 1997, p. 180) elicitation categories (Table 10 in Chapter 3). Testees tick category one or two if they have never seen the word in the list before or the word seems to be familiar but they do not know its meaning. If they dare to guess, they may put a Hungarian equivalent or an English definition into column III. If the testee feels to be certain about his or her knowledge of the word, column IV is to be filled in similarly. To demonstrate an even deeper or more precise knowledge of the word, testees may choose to write a sentence with the

word in column V (see Appendix C).

The Goulden *et al.* (1990) test is originally designed for self-assessment, which was abandoned in this test for assessment by the researcher for the following reasons: (1) to avoid a tendency of students overrating their knowledge of the words; and (2) the procedure of self-assessment seemed to be too time consuming to fit the time-frame of the seminar available for the present study. Instead, the VKS scoring categories were used, which appeared to provide a more reliable measure of vocabulary knowledge with its five categories measuring the depth of lexical knowledge as well, although data on how well students know the words is not analyzed in this paper.

Thus, in the modified test format each item was assessed either as known (1) or unknown (0), with no respect to the qualitative categories. This way the size of the receptive (passive) vocabulary could be calculated as described by Goulden *et al.* (1990). The categories of scoring are shown in Table 12:

Table 12. The modified VKS scoring categories for vocabulary size measurement

Self-report categories	Possible scores	Meaning of scores
I	0	The word is not familiar at all.
II	0	The word is familiar but its meaning is not known.
III	1	A correct synonym or translation is given.
IV	1	The word is used with semantic appropriateness in a sentence.
V.	1	The word is used with semantic appropriateness and grammatical accuracy in a sentence.

If testees mark category I or II, they get 0 point for the item, as they have never met the word before or have seen the word but do not know what it means. If a correct

Hungarian equivalent or an English definition is given in category III or IV, testees get 1 point for the item and the word is taken as known even if they are not able to use it in a sentence (category V left blank), but score 0 if the synonym or translation is incorrect in category III or IV (see arrows in Table 12). One point is given in every case testees provide a semantically appropriate use of the word in an English sentence, even if category III or IV is not filled in or the provided sentence is grammatically inaccurate but meaning is clear. However, if the word in the sentence is semantically inappropriate, score 0 is given (see arrows). If the word is used with semantic appropriateness and grammatical accuracy in a sentence, again 1 point is awarded. No extra points can be obtained for filling in all three categories (III, IV, V) correctly.

Goulden *et al.* explain their dictionary sampling based scoring as "each item represents 500 words, so the number of items known in a test should be multiplied by 500 to get the total base vocabulary score" (1990, p. 355). Thus, the maximum score of the fifty-item test is 50; therefore, the maximum size of receptive vocabulary it is able to measure is 25,000 English words.

5.2.4 Procedures

The test was administered to all 93 testees at the language practice seminars in the first two weeks of the semester (Appendix C). There was no time limit set, but no testee took longer than twenty minutes to complete the test. The results were then computerized and analysed.

5.2.5 Results and discussion

The Goulden *et al.* (1990) test sequenced the items into five levels of ten words according to word frequency, beginning with the most common (easy) ten to the very uncommon (difficult) ten words. Therefore, I expected that the first ten items would be known by the majority of the students and scores would hyperbolically decrease with the increase of item difficulty. Results are shown in Figure 7 below:
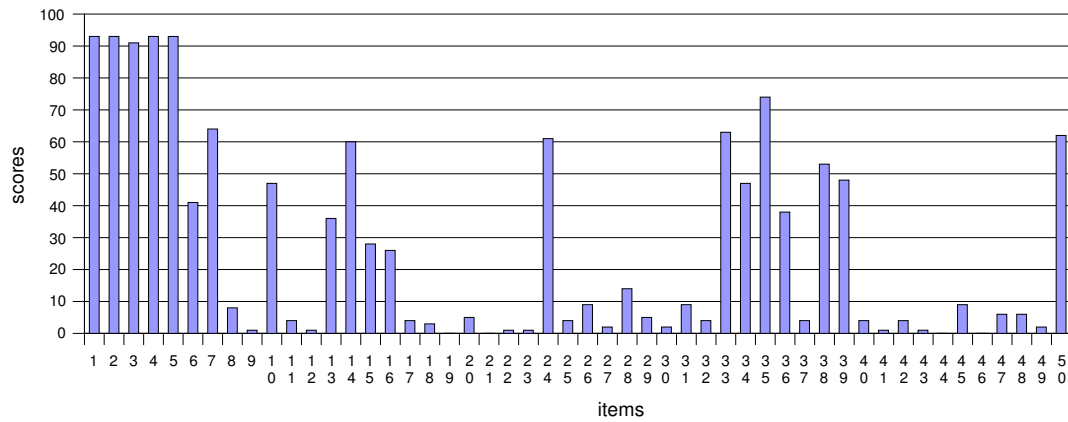
Figure 7. Item difficulty on the vocabulary size test

Only four out of the fifty words are completely unknown to all testees; thus, score zero on the test: item 19 (*untoward*), item 21 (*carpel*), item 44 (*apertometer*), and item 46 (*gusli*). All 93 students know the first five words in level one (items 1-10), as expected (*cool* 93 students, *kitchen* 93, *lead* 91, *cow* 93, and *frog* 93), whereas the second five words have proved to be more difficult and score lower (*scent* 41, *harsh* 67, *ascertain* 8, *sprig* 1, *matron* 47). *Ascertain* (item 8) and *sprig* (item 9) score surprisingly low, only 8 and 1 students out of the 93 have been able to give a correct definition, respectively.

The second group of ten words (level two, items 21-30) generally scores lower than the first ten, as expected (*amorphous* known by 38 students, *bagpipe* 60, *choleric* 28, *cock* 26), while items 11 (*coronet*), 12 (*jut*), 17 (*incumbent*), 18 (*offal*), 19 (*untoward*) and 20 (*amphitrite*) are hardly familiar to the testees. Level three (items 31-40) seem to be as difficult as the last ten words of the test (level five, items 41-50) with scores below ten, although there is one item in each group that juts out: *nominative* (item 24) correctly defined by 61 students and *slipper* (item 50) known by 62 students out of 93. As the majority of the testees were not only majoring in English language and literature but other languages as well, knowing *nominative* as a term in the field of

linguistics was of high probability, whereas the word *slipper* has a Hungarian culture specific connotation presumably contributing to the high score.

Level four (items 31-40), however, shows some unexpected results. The number of known words is much higher here than on levels three and five, contradicting my assumption that less frequent words are more difficult for the students than more common words. Items 31 (*directrix*), 32 (*footage*), 37 (*tandoor*) and 40 (*chanterelle*) are known by fewer than 10 students each, while items 33 (*horseshit*), 34 (*nighthawk*), 35 (*ravioli*), 36 (*aeroplankton*), 38 (*cogito*), and 39 (*corvette*) outscore items in level two. This might be explained by L1 specific cultural background knowledge again, e.g., in case of *nighthawk* the Hungarian word "*éjjelibagoly*" is used to refer to a person active at night, or the word *ravioli* as a popular international food widely known in Hungary, but any claim on this would be hypothesizing only, as no data was collected on this aspect of word knowledge.

Test scores are in the range of 8-28 words out of 50. The vocabulary size of the testees (score x 500), therefore, falls between 4,000 and 14,000 words in the English language, with an average of 7123,66, standard deviation from the average being 1777,97, median 8,000, as shown in Figure 8.
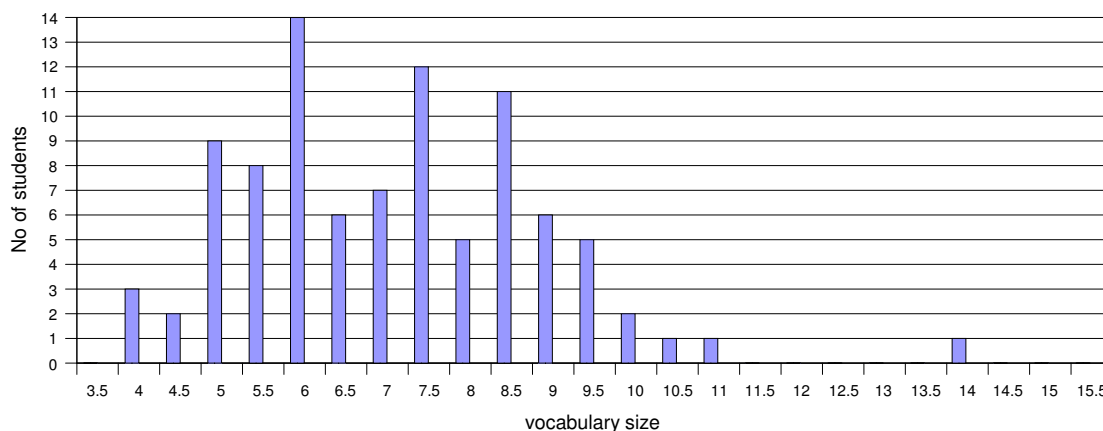
Figure 8. The distribution of receptive vocabulary size (1,000)

Only one testee reaches the fourteen thousand word level defined by Zechmeister *et al.*

(1993) as sufficient for academic studies in case of native speaker university undergraduates, standing far above the results of the rest of the students, with a considerable 3,000 word gap between the best and the second best score. The rest of the testees score below 11,000. If we take it as an outstanding individual performance, without the best score the average of the group falls to 7,049 words (SD= 1,634.21).

The most frequent score (mode) in the sample is 6,000 words (14 testees) which is below the average. Fourteen students scored 5,000 or lower, representing 15.5 percent of the testees. If we accept Laufer's (1987, 1992a, 1992b) claim that 5,000 is a "bottom line" for reading English at an academic level, it implies that 15 percent of these English majors will presumably face difficulty in reading and understanding academic texts in the course of their studies. In the light of the Hazenberg and Hulstijn (1996) study defining the minimal vocabulary size for academic studies in the receptive knowledge of 10,000 base words for non-native speakers of English, this figure mounts to a devastating 94 percent.

5.2.6 Summary of findings

Researchers of vocabulary distinguish between the breadth and depth of vocabulary knowledge. Proponents of qualitative (depth) research investigate how a word is represented in the language learner's mental lexicon, while research on vocabulary breadth provides an estimate of how many and what kind of words a language learner knows. There is converging evidence in recent vocabulary research and computerized corpus-based studies on the lexical coverage of texts that implies a link between receptive vocabulary size and reading comprehension. Research has shown that 95 percent of all tokens in an average text come from the 5,000 most frequent words in the English language and that for text comprehension readers need to be familiar with 95-98 percent of the words in a text. It is not clear how many words a non-native speaker of English should know receptively to be able to pursue academic studies, but the estimates range from a receptive knowledge of 5,000 to 14,000 base words.

This research has shown that taking the 5,000 base word estimate as sufficient for successful academic reading performance, 15 percent of the 93 English majors examined at the University of Pécs would face difficulty in the course of their studies, while 94 percent of the students fall below the 10,000 word level and only one student reaches the upper estimate of a minimal base vocabulary size. It was not the aim of this study to explore how well learners know the words of the test, as it only attempted to give an estimate of the receptive vocabulary size of the testees. Therefore, these results necessitate further research into the field that would provide deeper insights into both the quantitative and the qualitative aspects of student vocabulary by preparing vocabulary profiles of student essays, exploring performance on reading comprehension tests and investigating the link between these and general language proficiency, as well as academic success at various courses.

5.3 Receptive vocabulary size and productive vocabulary use

5.3.1 Research questions

The following small-scale study was inspired by the findings of the receptive vocabulary size test reported above to explore the relationship between receptive and productive word knowledge. Read (2000, p. 200) claims that "it is reasonable to expect that more proficient writers have a larger vocabulary knowledge that allows them to avoid repetition by using synonyms, superordinates and other kind of related words" and this "range of expression" is shown by the type-token ratio. The assumption is examined from a new perspective inquiring whether learners with a larger vocabulary show a wider range of lexical expression in their writings than their fellow students with smaller vocabulary. It is hypothesized that those L2 learners who have a relatively smaller receptive vocabulary would show a less varied use of words in their writings.

This study intends to explore whether lexical density in the writings of students with a large receptive vocabulary is higher than lexical density in the writings of those

learners whose receptive vocabulary is significantly smaller. In other words, a higher percentage of lexical words compared to grammatical words is expected in high vocabulary size test (VST) scoring texts (HVST) than in low VST scoring writings (LVST). Furthermore, it is hypothesised that students with higher VST scores use a wider selection of low frequency words in their writings than students who scored lower on the vocabulary size test.

5.3.2 Participants

All the 33 participants were first-year English majors at the University of Pécs, attending a Language Practice seminar run by the researcher. The students were randomly divided into two groups.

5.3.3 Instruments

Two instruments were applied to investigate the relationship of vocabulary size and the lexis in writing. To measure the size of the students' receptive vocabulary an adopted version of the 50-item test developed by Goulden, Nation and Read (1990) was used the same way as in the previous study. However, this time each item was assessed either as known (1 point) or unknown (0 point), with no respect to the qualitative categories.

In the second phase all 33 students taking the vocabulary size test were asked to write a 400 word take-home essay. Based on the comprehensive measures of lexical richness defined by Read (2000), the following calculations were done on the texts. The type-token ratio (T/t), lexical density and lexical variation were measured as discussed in Chapter 3.

Besides, the type-token ratio of the lexical words (as distinct from grammatical or function words) in the text was also calculated. In Laufer's study (1991, cited in Read, 2000, p. 203) a lexeme is a "single lexical item which may consist of more than one

form". She identified inflected forms of the verbs as one lexeme, homonyms were taken as separate words and derivatives of base words were counted as separate entries. A high figure means that the text contains a wide range of different words, a low one indicates that the writer has relied on a small stock of lexical words that are frequently repeated.

$$LV_B = \frac{\text{the number of different lexemes in the text}}{\text{the total number of lexemes in the text}}$$

A shortcoming of the statistics is, however, that although there usually is a word limit set on how long an essay should be, student writings are rarely exactly the same length. Read (2000) draws attention to the fact that the ratio changes as the number of tokens increases, making the comparison of data less reliable.

5.3.4 Procedures

The vocabulary size test was administered in the first language practice session of the autumn semester. The students had 90 minutes to complete the paper-and-pen test. In October, a month after the vocabulary size test both language practice groups were given the task to write a 400 word take-home essay titled 'The Student Loan', which I considered to be a hot topic of common interest to all participants at the time, as this new loan was introduced by the Hungarian government at the beginning of the semester. Since all the students were involved in the issue, no background research or knowledge was necessary on their part to form an opinion.

Based on the results of the vocabulary size test ten students were put into two categories (five each) and their essays were examined on the measures of vocabulary. The essays of the five lowest VST scoring students (named group LVST) were compared with those of the five highest VST scoring students (named group HVST) with the help of the Longman Mini Concordance software. The coded essays were

stored in the computer as text files stA-stJ with the consent of the students. To investigate lexical richness in the essays the software RANGE was used. As has been mentioned earlier in Part I of this thesis, the computer program applies three ready made base lists for comparative analysis: the first (BASEWRD1.DAT) includes the first most frequent 1,000 words in the English language from the General Service List by Michael West (1953); the second (BASEWRD2.DAT) involves the second most frequent 1,000 words from the same (GSL) word list; and the third (BASEWRD3.DAT) contains 570 word families of the Academic Word List developed by Coxhead (2000). The words of the Academic Word List are not among the first 2,000 words of the GSL, but are the most frequent in secondary school and university texts from a large selection of topics.

5.3.5 Results and discussion

The LVST group included the five essays of students with the lowest scores on the vocabulary size test. These files were stored in the computer as text files stA-E. The first search of the software was directed at how many words each text consisted of (token) and how many words out of the total were different (type). Table 13 presents the results of the first analysis of text information on each text:

Table 13. LVST Individual text information

text code	VST score	vocab size	type	token	T/t
stA	6	3,000	132	269	0.490
stB	6	3,000	201	422	0.476
stC	7	3,500	204	436	0.467
stD	7	3,500	160	358	0.446
stE	9	4,500	175	370	0.472

In the LVST group the average length of the essays is 371 words, which is exceeded by two essays. The type-token ration ranges between 0.44 – 0.49 and stA has the highest type-token ratio (0.4907), although this student scored the lowest on the vocabulary size test. The analysis was performed on each of the five texts of the HVST group as well, the results are as follows:

Table 14. HVST individual text information

text code	VST score	vocab size	type	token	T/t
stF	17	8,500	220	558	0.394
stG	17	8,500	199	405	0.491
stH	18	9,000	205	455	0.450
stI	18	9,000	248	624	0.397
stJ	21	11,000	210	345	0.608

In this group the average length of the essays is 477 words. The range of the type-token ratio is wider in this group, varying from 0.39 to 0.60. StJ has the highest type-token ratio (0.60) which seems to support the hypothesis that a student with the largest receptive vocabulary (11.000) uses more varied language in essay writing than others with smaller receptive vocabulary. However, comparing the scores of stA, stF and stI it can be observed that although stA possesses the smallest receptive vocabulary (3,000), its type-token ratio (0.49) outscores those of stF and stI (0.39) in spite of their larger vocabulary size (8,500/9,000), which thus contradicts the expectations of my first research question.

The above tables provide detailed textual information on each student's writing performance, however, as the aim of this study is not to compare individual output, but to obtain more group specific data on low and high VST scoring student writings, the 5-5 texts were joined to form two long text files labelled as "highvst" text and "lowvst" text, which were then made subject to concordance repeatedly:

Table 15. Comparative joint-text information

text code	VST score range	vocab size range	type	token	type/token ratio
highvst	17-21	8,500-11,000	643	2,387	0.2694
lowvst	6-9	3,000-4,500	511	1,855	0.2754

"Highvst" text is longer than "lowvst" text, as seen from token rates, which might suggest that students with a larger receptive vocabulary tend to express their opinion

more fluently, although this assumption is contradicted by text stJ, which is the shortest of all, although its author is the highest VST scoring student (21) with the largest receptive vocabulary (11,000). Running an independent samples t-test on VST scores a statistically significant difference has been found in vocabulary size between HVST and LVST group (t=5.5451; p< .001).

Surprisingly, although "highvst" text is longer and it contains more types than "lowvst" text, there is no significant difference in the type-token ratio of the two corpora according to the independent sample t-test (t=0.945; p< .05). Thus, my hypothesis is not supported by the type-token ratio: higher VST scoring students did not use more varied vocabulary in their essays. A possible explanation to this could be one of the shortcomings of the present study, that measuring vocabulary size by only one test might not give an unquestionably reliable result. Besides, the vocabulary size test did not measure the depth of vocabulary knowledge, no context was provided with the items, which might also have influenced the test results.

Calculating lexical variation as the ratio of content words in the text (LVB) might reveal more of the process as in the previous calculation the high percentage of function words veiled lexical variation. Table 16 shows lexical variation. First, the frequency lists of both texts were displayed with the help of the Longman Mini Concordance software, then the calculation was done partly manually and partly by the program Microsoft Excel.

Table 16. Lexical variation in ten essays

Lexemes	Highvst text	Lowvst text
Total lexemes	1106	851
Different lexemes	457	362
LVB1	0.413	0.425
LVB2	0.404	0.417

A higher LVB figure was expected for highvst.txt, as the higher this score is, the more varied the text is lexically. In this study, as Table 16 indicates, "lowvst" text shows a

little higher lexical variation (0.425) than "highvst" text (0.413). LVB2 is the new lexical variation rate if numbers are excluded from the calculation as numbers are not taken as informative content words. However, this minor exclusion does not change the rate either, thus the first hypothesis, that learners who have a relatively smaller receptive vocabulary would show a less varied use of words in their writings is not verified in the study.

As for lexical density (Table 17), the second hypothesis is not justified by the results. The rate of lexical words per total number of tokens in the text is not higher in "highvst" text than in "lowvst" text. It means that the essays of the students with a larger vocabulary did not contain more lexical words vs. grammatical words than the essays of the students with a lower vocabulary test score.

Table 17. Lexical density in the two joint texts

	Highvst.txt	Lowvst.txt
Total lexical words	1144	870
Total tokens	2387	1855
LD	0.479	0.469

However, Laufer and Nation (1995) argue that the validity of the lexical density index is questionable, as it depends on the number of function words involved in tokens and the fewer grammatical words can reflect a more sophisticated use of syntax, which is not a lexical but a structural aspect of construction. In order to investigate lexical sophistication by the lexical frequency profiles of the texts, RANGE was run on both files. Results are as follows:

Table 18. RANGE lexical profile of "highvst" text

```
WORD LIST              TOKENS/%           TYPES/%

one                    2054/87.0          448/71.5
two                     137/ 5.8           59/ 9.4
three                    73/ 3.1           50/ 8.0
not in the lists         96/ 4.1           70/11.2
Total                  2360                627
```

In this text 71.5 % of the types of words in the student essays occurred among the first 1,000 most frequent English words, and 11.2% of the words were not represented in any of the three base word lists, which suggests that these are low frequency words in the English language. The same profile was prepared for the low VST scoring texts:

Table 19. RANGE lexical profile of "lowvst" text

WORD LIST	TOKENS/%	TYPES/%
one	1623/89.1	374/76.8
two	86/ 4.7	29/ 6.0
three	50/ 2.7	38/ 7.8
not in the lists	62/ 3.4	46/ 9.4
Total	1821	487

Students used more words from base word list 1 (76.8%), and less low-frequency words not included in any list (9.4%). The two profiles now show a more significant difference between the two groups of essays, although the analysis included function words. To be able to examine the difference in lexical sophistication, though, I restricted the range of vocabulary to lexical (content) words and ran a second analysis on the texts:

Table 20. RANGE lexical profile of content words

WORD LIST	"highvst" types/%	"lowvst" types/%
one	332/67.3	262/73.4
two	57/11.6	29/ 8.1
three	50/10.1	37/10.4
not in the lists	54/11.0	29/ 8.1
Total	493	357

Students with a larger vocabulary seem to have used a lower percentage of words from base word list 1 than those with a smaller vocabulary (67% - 73%), and the former essays display a higher percentage (11%) of low-frequency words than the "lowvst" group (8.1%). This result is consistent with the findings of Laufer and Nation (1995), suggesting that second language learners' vocabulary size is reflected in the productive

use of the language.

5.3.6 Summary of findings

In this study lexical density and lexical variation did not reveal any significant difference between the writings of students with larger and smaller receptive vocabulary, although simply by reading the essays the difference is easily noticed: the essays constituting "highvst" text "read a lot better" than the essays of those students who scored lower on the vocabulary size test. On the other hand, the Lexical Frequency Profile of the essays showed that students with a larger receptive vocabulary displayed more rare (low-frequency) words in their writing than others whose receptive vocabulary was significantly smaller. These contradicting results might suggest that the difference is not dependent on vocabulary size only, and it is to be further investigated.

A factor of reliability concerns whether the sample is representative enough. As the present study examines only ten students' writings and vocabulary size test scores, extending the study to all 33 participants or involving even more testees in the procedure might provide more reliable results. The results of lexical variation might also suggest that the students who scored the lowest at the vocabulary size test perhaps did not take the test very seriously, so their scores may not reflect their real vocabulary size. Therefore, student attitudes and writing strategies may also be investigated in a follow-up study.

This study used receptive vocabulary as a measure of vocabulary size. It would be interesting to see whether measuring productive vocabulary only, or joining receptive and productive vocabulary test scores would yield similar results. Vocabulary size test data, on the other hand, were compared with vocabulary measures of take-home essay samples. The circumstances of writing a take-home essay presumably allow slower pace in the process of production and more chance to look up or check vocabulary items in a dictionary than in case of an in-class essay. It would be wise therefore to

replicate the study and examine in-class essays instead of essays written at home, or compare the two types of essays of the same authors. Read (2000) argues that only a few studies have concentrated on the reliability of lexical statistics; therefore, it would be valuable to continue the study with the same participants for a longer period of time. A period of three or four semesters could better highlight the amount of vocabulary learning taking place over time.

5.4 Intentional versus incidental vocabulary learning

5.4.1 Research questions

This exploratory study aims to investigate and compare the nature of incidental and intentional vocabulary learning. Krashen's Input Hypothesis predicts that more comprehensible input in the form of reading results in greater competence in vocabulary (Krashen, 1989, p. 441). The study based on two groups of EFL learners intends to explore whether intentional vocabulary learning and conscious preparation for regular vocabulary tests are more effective in the long-term retention of words than incidental vocabulary learning as a by-product of reading only. The research reported below is innovative in the double scoring procedure applied to assess qualitative word knowledge besides providing a quantitative estimate of lexical acquisition in one test format. On the other hand, the majority of studies on incidental vocabulary acquisition lack delayed post-tests or operate along a relatively short time span between the pretest and the post-test (Schmitt, 2008). This study measures incidental and intentional vocabulary acquisition during a 14-week semester.

5.4.2 Participants
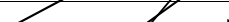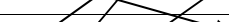
The 33 participants were all native speakers of Hungarian and English majors in their first semester at the English department of the University of Pécs. They were randomly divided into two groups (A and B), each attending fourteen 90-minute weekly seminars of Language Practice during the semester.

5.4.3 Instruments

Both groups took two tests during the semester. The pretest intended to find out whether the receptive vocabulary size of the students differed significantly before the treatment. Test 2 of the 50-item Goulden *et al.* (1990) tests was applied in the same fashion as discussed in the first study of the chapter.

The 50-item post-test (Appendix D) contained 42 words taken from the short texts read during the semester in both groups. These words were tested in the weekly word tests administered to group A only during the semester. Eight words were selected based on their intuitively estimated frequency in the novel read by both groups, which was unexpected for all participants.

Table 21. Scoring categories of the post-test for in-depth analysis

Categories	Score	Meaning of score
I	0	The word is not familiar at all.
II	1	The word is familiar but its meaning is not known.
	2	The word is familiar, context is identified, but its meaning is not known.
III.	3	A correct synonym or translation is given. (III. or V. correctly filled in).
IV	4	A correct synonym or translation is given and context is identified. (III./V. and IV. correctly filled in).
V	5	A correct synonym or translation is given and the word is used with semantic appropriateness and/or grammatical accuracy in a sentence. (III./V., and VI. correctly filled in).
VI	6	A correct synonym or translation is given, the word is used with semantic appropriateness and grammatical accuracy in a sentence and context is identified. (IV., III./V., and VI. correctly filled in).

As a modification of the answer categories used in the previous studies an extra category (IV) was inserted to inquire about the context of the words to find out whether participants remembered where they had come across the particular word. The

assessment followed a double-scoring procedure to explore how many of the words students remembered and how well they knew these words. Therefore, the post-test results were first scored with the above described method giving 1 point for each known item and 0 if the item was not known. Then, the tests were scored as shown in Table 21 to get an insight into the depth of lexical knowledge. Besides the pre- and the post-tests, both groups filled in a course evaluation form in the last session of the semester, thus providing valuable information on students' beliefs (Appendix D).

5.4.4 Procedures

The two groups were pretested in the first week of the semester in order to find out about their initial vocabulary size as well as to prove that there is no significant difference between the receptive vocabaulry size of the two groups. Both groups received the same readings during the fall semester of the 2001 academic year covering a wide selection of topics. Texts to be read at home ranged from authentic *Newsweek* articles to short texts from proficiency course books as well as a contemporary novel, Janice Galloway's (1989) *The Trick Is to Keep Breathing*. In class readings were only discussed in terms of content.

Group A (henceforth intentional learning group) was instructed to look up all the unknown words in the texts at home and wrote weekly retention tests on the vocabulary of the texts excluding the novel. These regular tests consisted of a list of 20 English words and students were asked to provide a definition and write a sentence illustrating a possible meaning of the word in English.

Group B (henceforth incidental learning group) read the same articles at home but did not write any word tests and dictionary use was not encouraged either. Although this study does not investigate strategy use, the possibility that the participants of this group also looked up a few words on their own for clarification of meaning cannot be excluded. Paribakht and Wesche (1999) warn that learners may have their own conceptions about learning and strategy use in spite of instructions. However,

dictionary use does not disturb the comparison of vocabulary learning processes in the two groups, as in Laufer's (1998) definition of the term incidental learning does not exclude conscious learning: what makes a difference is the presence or absence of warning for a later upcoming vocabulary retention test. Neither of the groups was forewarned of the post-test.

5.4.5 Results and discussion

The test scores on the pretest did not indicate any significant difference between the two groups. The mean of group A and group B was 11.71 and 11.93 out of 50, respectively, thus the average receptive vocabulary size is 5,855 English words in group A and 5,965 words in group B. A two-tailed independent sample t-test (t= .754066, df=31) further strengthened the assumption that the two groups were similar in the size of their receptive vocabulary, thus setting an adequate starting point for further investigation into intentional and incidental vocabulary learning.

The post-test was taken by 30 students and data loss occurred with item 47, so only 49 out of the original 50 items were analysed. Results of the first scoring are shown in Figure 9 where group A is represented by the first column and group B by the second scoring on each item. Scores range from 6 to 33 in group A and from 3 to 29 in group B with the averages of 15.8 and 13.3 points, respectively. The words to investigate incidental learning from the novel are items 5, 25, 30, 32, 36, 38, 48 and 49 (*comeuppance, numb, rafting, recurring, surrender, drown, ultimate, wee*). There was one word only that nobody knew: item 16 (*hookah*) which appeared in a tale. The next two most difficult items were from the same text (item 4: *caitiff* and item 19: *kaboosh*), which were absolutely unknown to group B, but known by 2 and 4 members of group A. Other equally difficult items from the articles were item 1 (*affidavit*), item 11 (*immense*), 23 (*miffed*), 31 (*rave*), 44 (*thwart*) and 50 (*yearn*). Group A highly outperfomed Group B on items 14 (*efficacious*), 20 (*limelight),* 40 (*muse on*), 35 (*spooks)*, 34 (*smack*), and 8 (*dime-store*). These words are relatively low-frequency words in the English language representing vocabulary specific to *Newsweek*. The fact

that these items were better recalled by group A seems to support the hypothesis that intentional learning is more efficient than incidental learning in the retention of words.
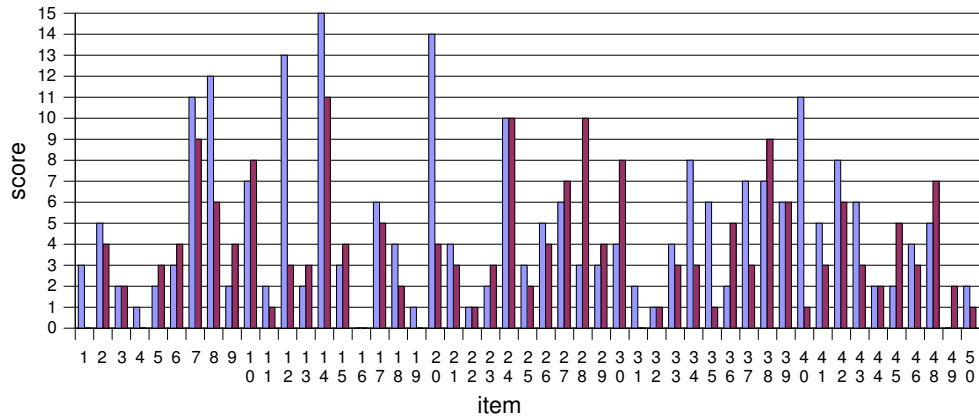
Figure 9. Item difficulty on the post-test with the first scoring

The most well-known item turned out to be item 14 *(go bankrupt)* in both groups. Another high-scoring item was 24 *(nightingale)* from a tale. However, on some items group B outperformed group A: items 28 (*preview)* and 45 (*trigger*). Interestingly, the other six such items were all words from the novel: item 5 (*comeuppance*), 30 (*rafting*), 36 (*surrender*), 38 (*drown*), 48 (*ultimate*) and 49 (*wee*). These represent 6 out of the 8 words taken from the novel (item 33 was equally unfamiliar and item 25 scored higher in group A).

In group A, 30 percent of the students knew the words from the articles on average per item, and 20 percent knew the words taken from the novel. It means that those students who wrote weekly word tests were able to retain more of those words they consciously focused on than those words they did not write tests on. In group B, this rate is inverse: 24% of the students knew the items from the articles and 30 percent identified the items from the novel correctly.

However, it is necessary to examine whether the difference between the retention rates of the two groups is statistically significant to be able to conclude that this difference

is clearly the result of the difference in treatment. In this respect the post-test brought unexpected results. A two-tailed independent sample t-test showed no significant difference between the two groups in the retention of words (t= .37009478; df= 28). This result seems to contradict my hypothesis, and suggests that the treatment did not affect retention rates significantly.

As Henriksen (1999) points out, the concept of successful acquisition may vary as word learning outcomes range over continua of lexical knowledge and use of competencies from partial recognition of a word to precise knowledge and productive use. Quantitative scoring provided information only about whether an item was known or unknown, but it did not reveal the nature of this knowledge. In the second part of the study I attempt to approach word knowledge according to the partial - precise, receptive - productive continua (Henriksen, 1999). Results of the second scoring based on Table 21 are shown in Figure 10 below.
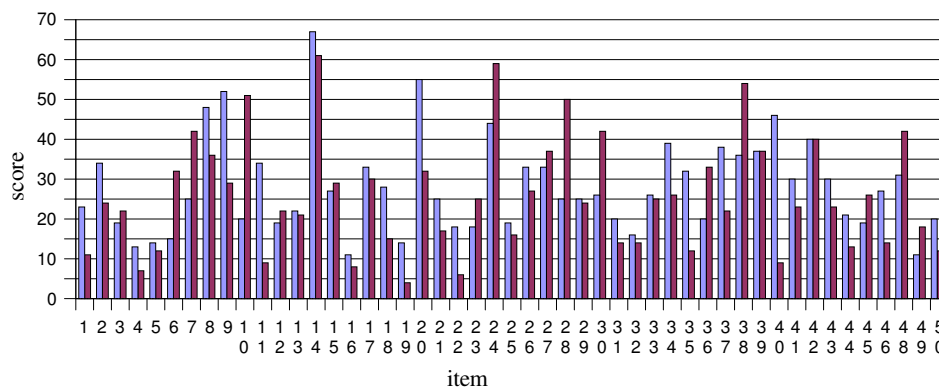
Figure 10. Item difficulty on the post-test with the second scoring

High item scores mean that a lot of testees managed to define meanings correctly, wrote a semantically and grammatically correct sentence with the word and identified context as well. As there were 30 testees the highest possible item score is 180, however, the highest item score on the post-test is 67 in group A and 61 in group B on

item 14 (*go bankrupt*), while the lowest scores can be observed on items 16 (*hookah*) and 49 (*wee*) in group A and item 19 (*kaboosh*) in group B.

The results of the two groups are compared in Figure 11. The diagram shows the difference in the scores on each item. Positive scores represent cases when group A outperformed group B on a particular item, whereas negative scores mark better results of group B. For instance, on item 1 group A outperformed group B by 12 points; on item 10 group B scored 31 points higher than group A.
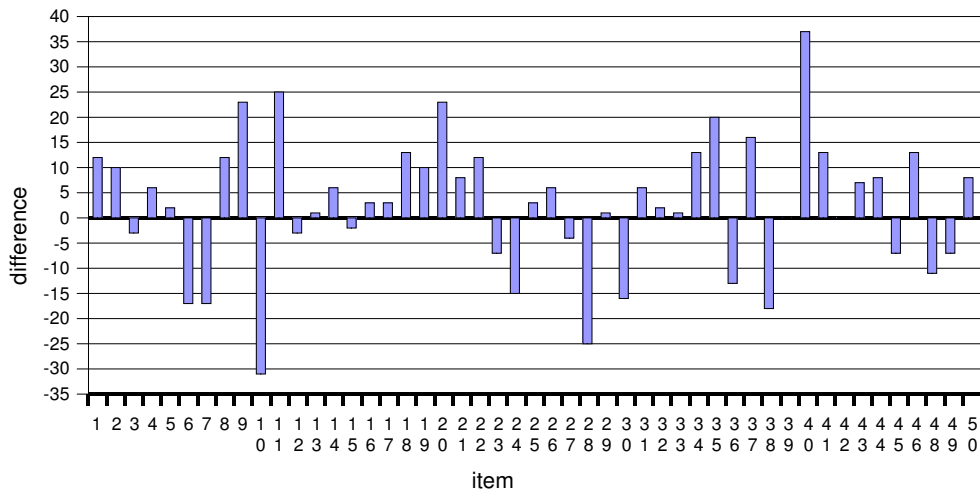
Figure 11. Difference in the depth of word knowledge on the post-test

Sixty-seven percent of the scores is above zero (positive) representing a better overall performance of group A. The highest positive differences are on item 40 (+37; *to muse on*), item 11 (+25; *achelons*), item 20 (+23; *limelight*), item 9 (+23; *dimly*), and item 35 (+20; *spooks*). On the other hand, group B outperformed group A on the following items: item 10 (-30; *dungeon*), item 28 (-25; *preview*), item 38 (-18; *to drown*), item 6 and 7 (-17, -17; *delicacy* and *desperation*). Concerning the 8 words taken from the novel the results of the two groups are similar, group A outperformed group B on four items only (items 5, 25, 32 and 49), while group B did better on the other four words (items 30, 36, 38 and 48). However, an independent sample t-test run on the scores of

the second scoring did not show any statistically significant difference between the two groups (t= .417; df=28).

In this in-depth analysis higher scores represent more precise, deeper knowledge, while lower scores imply partial, superficial knowledge of a lexical item. The frequency of scores from 0 (no previous encounter with the word) to 6 (correct synonym or translation given, correctly used in a sentence and context is identified) is shown in Table 22:

Table 22. Frequency of scores with the second scoring on the post-test

Score	0	1	2	3	4	5	6
Group A	10.20%	55.60%	2%	14.50%	0.70%	15.80%	1.22%
Group B	24.60%	45%	2.80%	8.70%	1.50%	14.10%	3.10%

The first row of the table shows the possible item scores based on the scoring categories described in Table 21, while the second and third rows present the percentage of the occurrence of the score in group A and group B. In group A, 10.2 percent of the items were identified in category I of the answer sheet as no previous encounter with the word, although all of the words appeared in one of the weekly word tests they consciously prepared for. This rate is twice as high (24.6%) in group B.

It is interesting to examine the rate of items scoring 1 on the post-test (the word is familiar but its meaning is not known). In group A 55.6 percent of the students reported previous encounter with a word (noticing), but was unable to provide a Hungarian equivalent or define it in English, write a semantically and/or grammatically correct sentence to illustrate meaning or identify the context where the word was encountered. An item also scored 1 if the testee filled in any of categories III, IV, V or VI incorrectly, which, in my opinion, implies noticing or previous encounter with the word but its meaning is not known, although the chance of simple guessing cannot be excluded either. This rate is somewhat lower in group B (45%).

These results might be explained by Gass' (1999) assumption that in the process of inferring meaning from context some words are unnecessary for general comprehension and thus are ignored or neglected in the text. Fraser (1999) notes in her study that the rate of ignoring words decreased with instruction on strategy use. Lawson and Hogben (1996) also point out that if a particular word is unknown but its presence does not disturb the general comprehension of the whole sentence or passage, or the meaning of the word is clear from a contextually rich text, then the noticing of the word does not automatically lead to acquisition and better long-term recall.

Group A was able to give a correct definition or Hungarian equivalent to 14.5 percent of the items without identifying the context, while group B scored 3 only on 8.7 percent of the items. The difference in the percentage of correctly defined words without context (score 3) might be explained by the presence or absence of dictionary use, however, this study did not examine vocabulary learning strategies. The frequency of score 5 representing productive use of a word in a sentence without the identification of context is similar in the two groups; 15.8 percent in group A and 14.1 percent in group B.

The role of context in vocabulary learning is highlighted by items scoring 2, 4 or 6. An item scored 2 when context was identified by naming the text the word occurred in but the meaning of the word was not known, an incorrect definition or Hungarian equivalent was provided or the use of the word was semantically inappropriate in the sentence. The score was 4 when a correct synonym or translation was given and context was also identified: category II. or V and IV were correctly filled in, but there was no proof of productive vocabulary use. Score 6 was given in cases when a correct synonym or translation was provided, the word was used with both semantic appropriateness and grammatical accuracy in a sentence and even context was identified. However, I also gave 6 points when a testee quoted a sentence from a text in category VI, but failed to fill in category IV on context, as in my opinion context was clearly defined by the quoted sentence.

Results show that group B seems to have made better use of context in their reading and vocabulary acquisition: 2.8 percent of the items scored 2, 1.5 percent scored 4, twice as many as items in group A (0.7%) and three times as many scored 6 as in group A (3.1% vs. 1.22%). This result appears to be supported by the findings of Lawson and Hogben (1996) who proposed that it is the nature of processing that governs long-term retention, thus a reader whose intention is to comprehend a text may apply processing that is successful in the long-term retention of words. Conversely, they argue, a reader whose intention is deliberate vocabulary acquisition may use memory procedures that are more effective for short-term recall.

Paribakht and Wesche (1999) warn that learners may have their own conceptions about learning and strategy use in spite of instructions. This assumption is supported by the results of the end-term course evaluation form. All of the students in group B would have preferred weekly word tests, as in their opinion, this way they would have learnt more words, because otherwise they are lazy to look up meanings of unknown words on their own. This implies their assumption that looking up and intentional learning of words is more efficient than reading only. This assumption is further strengthened by the feedback received from group A, where all students thought that the weekly word tests helped them develop their vocabulary knowledge. However, twelve out of 15 students noted that they forgot most of the words a few weeks after the weekly tests.

Any correlation between the scores of the pre-test and the post-test would reveal possible explanations; therefore, some further statistical investigation was attempted. Running a Pearson correlation test it was found that the correlation in group A is not significant (r= .349235; df=13), but a strong correlation of the two tests was shown in group B (r= .878546; df=13; p< .001). It means that in the 'incidental learning' group (B) those students who did well on the vocabulary size test attained higher scores on the post-test as well, and those who scored lower at the beginning of the semester also scored lower on the post–test.

5.4.6 Summary of findings

This study attempted to explore and compare the nature of intentional and incidental vocabulary learning and aimed to find out whether conscious preparation for regular vocabulary retention tests proves to be more efficient in a fourteen-week study than incidental vocabulary learning as a by-product of reading only. A pre-test was administered to the two participating groups before treatment to define their initial vocabulary size. Then, the post-test measured the rate of vocabulary learning after treatment in both groups. A new scoring method was proposed to measure both vocabulary size and depth in one test format and the post-test was scored with a double scoring procedure. The results show no statistically significant difference between the retention rates of the two groups. Thus, it cannot be claimed that in the long-run the intentional learning of words is more efficient than incidental learning defined as vocabulary learning from reading with the absence of warning for an upcoming retention test.

5.5 Conclusion

This chapter discussed the nature of receptive vocabulary knowledge from various perspectives and the three pilot studies reported here served as starting points for further research into the lexical knowledge of first-year students of English. The first study reported the results of a vocabulary size test administered with the aim of estimating whether their receptive vocabulary size enables the students to cope with academic reading texts. The interpretation of the findings is manifold; however, one is tempted to argue that the majority of the UP students lack the sufficient vocabulary size and may face difficulty in higher grades of their studies. These findings inspired further research into the assessment of academic vocabulary discussed in chapter seven. Special consideration is given to low-frequency sub-technical vocabulary common in texts used in the departments of English studies, therefore, as a follow-up,

a corpus-based study reported in chapter six was designed to examine the nature of lexis in the required compulsory readings.

A small-scale study presented in the second section of the chapter investigated the relationship between receptive vocabulary size and productive vocabulary use. It has been argued that the vocabulary profile of student writings is a better predictor of receptive vocabulary size than the type-token ratio only. This line of argument is further pursued in a larger scale study of a student corpus in chapter seven.

The third study examined whether intentional or incidental vocabulary learning proved to be more effective in the long run. The idea of the incremental nature of incidental vocabulary learning has been introduced early in chapter 2 of this thesis and arguments for the direct teaching of vocabulary, especially low-frequency lexis, have been discussed in detail. Although this preliminary study did not reveal a significant difference in the post-test between the two types of treatment, it is argued in the subsequent chapters that special attention has to be given to the sub-technical vocabulary of compulsory readings required for English studies.

Chapter 6

How well? Explorations into the vocabulary of first-year English majors

6.1 Introduction

Based on the results and experiences gained from the studies reported in the previous chapter, Chapter 6 introduces the developmental stages of the vocabulary component of the first-year proficiency test that lead the researcher to the design of the corpus-based test battery discussed in Chapter 7. The three stages of test development represent three years of testing different batteries as the proficiency test aiming to filter out students below a set level is administered on an annual basis according to curricular requirements.

Assuming that the knowledge of words common in a wide range of academic texts enables students to cope with the reading load at university, stage one intended to explore how well the students are familiar with academic vocabulary. Stage two reached a little further into the domain of lexis by exploring the knowledge of less frequent words based on the assumption that the 10K word level better differentiates among students with larger and smaller vocabularies, the main aim of the present test development process.

As neither of the two widely used vocabulary tests adapted in stages one and two seemed to match the vocabulary needs of our students perfectly, the findings of these two studies forged the development of a new battery of lexis perceived as useful in English studies. The study reported as stage 3 in this chapter seeks to explore among others the relationships between vocabulary and listening, reading and writing skills, general language proficiency, and academic success.

6.2 Developing the vocabulary component of the Proficiency test

Mention has been made of grammar not being one of the four skills, but a component of linguistic competence in the widely accepted communicative competence model (e.g., Hymes, 1971; Swain, 1985; Bachman, 1990). So is vocabulary. A sufficient knowledge of both is a prerequisite of successful text comprehension, written or aural, as well as using language productively in an essay writing task. The wealth of research overviewed in the first part of this thesis supports the prevailance of word knowledge over knowledge of syntax in text comprehension (Widdowson, 1978, in Boyd-Zimmerman, 1997, p. 6; Laufer, 1997a, 1997b; Nation, 1993a, 1993b), while other studies (Doró, 2008; Morris & Cobb, 2003) describe vocabulary as a good predictor of academic success determining how students are able to cope with the reading load at university.

On the basis of the predictive value of vocabulary knowledge on later academic success and the results of the vocabulary size test introduced in the previous chapter, the testing team responsible for writing items for the Filter/Proficiency test decided to insert a separate vocabulary section into the original test battery in 2002. While Zareva (2005, p. 560) concluded that at higher levels of language proficiency vocabulary size and a measure of productive-receptive control are "more revealing of the overall state of learners' vocabulary than the dimension of quality", the general guiding principle remained the same as the one applying to the rest of the test, i.e. to filter out those candidates who were below a certain level, in this case lacking in vocabulary knowledge. These, together with other criteria of practicality, objectivity, suitability for computer item analysis and ease of both scoring and administration lead us to the decision of using a discrete point, receptive vocabulary size test first developed by Nation (1990), later revised and validated by Schmitt (2000), the Vocabulary Levels Test (VLT) introduced in detail in Chapter 3. The fifth level of the test is not based on frequency data, it comprises a representative sample of words from the Academic Word List (Coxhead, 2000), i.e. words that are most common in academic texts from a wide range of fields and subjects. Each level consists of ten clusters of six words that

need to be matched with three synonyms or short definitions in order to reduce the need for reading and knowledge of syntax, and at the same time to minimize the chances of guessing correctly. Below is an example of a noun cluster from the academic word level of the VLT (with key):

```
1 area
2 contract              __2___ written agreement
3 definition            __5___ way of doing something
4 evidence              __4___ reason for believing
5 method                       something is or is not true
6 role
```

Considering the place of such a long, 150-item test in the construct of our original test battery restricted by limitations of time allotment, scoring and item analysis procedures, the testing team agreed on using the academic level clusters of VLT only. It was decided to be included in the grammar and usage component of the test as a third section, to follow a four-option, 30-item multiple choice test of grammar, and a 10-item multiple choice section of three-option corpus-based sentences, where only one of the options fill in all three sentences correctly. Thus, with the 30 items of the vocabulary test the possible maximum score on the grammar and usage section is seventy. To avoid over-emphasizing the importance of the vocabulary test, no pass marks are set for the separate sections, the three sections are treated as one unit. Based on facility values and Rasch item difficulty figures Schmitt (2000) found that the words in the academic level of VLT fit between the 2,000 and the 10,000 word levels, thus it covers a broad range of word knowledge. He recommends the use of the levels test not as a fixed battery, but as flexible on the basis of the specific demands in each testing situation, allowing for picking a certain level only to be administered, or mingling the two versions for a given level to create a longer test.

6.3 Stage 1: Academic vocabulary

Considering the findings of research reviewed in the previous chapters as well as the above outlined issues, we decided to apply the academic level only, as the knowledge of these words has proved to be essential in understanding authentic academic texts. Besides filtering students lacking in vocabulary knowledge, this study aims to find out how familiar English majors are with academic vocabulary essential for pursuing university studies and explores how well the academic section of the Vocabulary Levels Test (Schmitt, Schmitt & Clapham, 2001) discriminates between candidates with larger and smaller vocabularies.

In the adaptation of the Schmitt *et al.* (2001) test, the format of the clusters was tailored to suit the ITEMAN software better, i.e. the clusters were numbered 1-10 and the items within the clusters were assigned capital letters A-F (Appendix F). The first tests involving the vocabulary component were then administered in the Spring semester of 2002, 154 students taking Version 1, and 66 students taking Version 2 of the VLT academic level test. Data were analyzed with ITEMAN (TM) Version 3.50, the results shown below first represent the overall results gained from the grammar and usage component supplemented with the academic word level test of the VLT. It is important to note here that the examinees of the two test versions were not the same students.

Table 23. Descriptive statistics on the academic word level tests

GAU + Academic word level	Version 1	Version 2
N of Items	70	70
N of Examinees	154	66
Mean	46.39	41.14
Variance	56.37	29.15
SD	7.51	5.4
Skew	-0.25	-0.02
Kurtosis	0.84	-0.41
Minimum score	23	30
Maximum score	67	53
Median	47	41

The mean on the grammar and usage component was 46.39 and 41.14 on the two versions (SD: 7.51 and 5.4, respectively). A difference in variance can be observed between the two versions indicating a wider scale of scores on version 1. Taking a closer look at the vocabulary sub-tests (Table 24), the relatively high means (26.14 and 25.1) and the narrow range of scores (min1= 12, min2= 17; max1= 27, max2= 30) suggest that the academic word level test proved to be manageable for the majority of candidates on both versions.

Table 24. Descriptive statistics on the subtests of the grammar and usage component

GAU + academic word level	Version 1			Version 2		
Part	1	2	3	1	2	3
N of Items	30	10	30	30	10	30
N of Examinees	154	154	154	66	66	66
Mean	14.9	5.38	26.12	13.46	2.61	25.08
Variance	16.12	3.04	11.27	12.49	1.39	7.56
SD	4.02	1.74	3.36	3.53	1.18	2.75
Skew	0.39	0.02	-2.69	0.71	0.74	-0.6
Kurtosis	0.41	0.14	12.61	1.63	0.28	0.2
Minimum score	6	0	4	6	1	17
Maximum score	28	10	30	25	6	30
Median	14	5	27	13	2	25

Score distribution curves illustrate the above data in a more practical way. Figure 12 present score distributions on the two test versions applied. Both curves are shifted to the right showing that the number of correct answers is high. The first curve (Version 1) is more elongated which means that the range of scores is wider (4-30) than on Version 2 displaying a stubby curve (17-30). On the second test version the distribution of scores is closer to the normal distribution, although shifted to the right as no candidate scored lower than 17 out of 30.
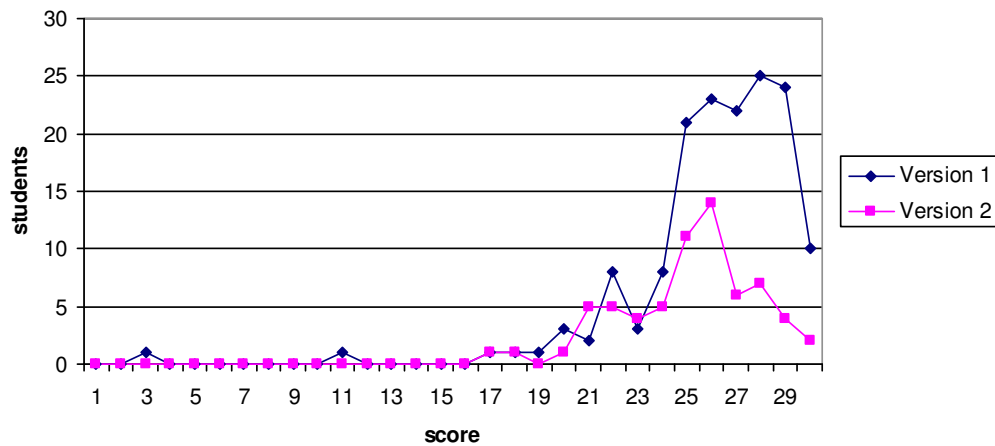
Figure 12. Score distribution on the Academic word level of the VLT

Item discrimination indices range from .00 to .59 and .61 on Version 1 and Version 2, respectively, showing a high number of non-discriminating items and a very few highly differentiating ones on both sub-tests. It implies that for the majority of the testees the words involved in the academic word level of the VLT is manageable.

A moderate correlation was found between the vocabulary sub-tests and the other two sub-tests of the GAU component on both test occasions (Version1: r_{13}= .502, p< 0.001; r_{23}= .400, p< 0.001; Version 2: r_{13}= .33, p< 0.01; r_{23}= .0103, p>0.05) indicating that those who did well on the vocabulary test scored higher on the grammar test as well. This significant correlation might allow us to assume that performance on the vocabulary test is related to performance on the grammar test with these two groups of examinees in this particular testing situation, however, further investigation is necessary to be able to arrive at more generalizable conclusions in this field, keeping in mind that correlation indicates relatedness but does not mean causality. Furthermore, this moderate correlation suggests that the two constructs are related but the two components of the test measure different things, as intended.

6.4 Stage 2: The 10K word level

The four VLT levels were sampled on the basis of word frequency, thus it seems reasonable to hypothesise that the higher the level, the better the test differentiates among candidates with larger and smaller vocabularies, and consequently among more and less proficient examinees. On the other hand, although there is no consensus on the minimal vocabulary size necessary for pursuing university studies (Zechmeister *et al.* 1993; Hazenberg & Hulstijn, 1996; Laufer 1992b), these findings are by no means to be ignored.

From previous research outlined earlier, the question arises whether academic- or low-frequency vocabulary is a better indicator of student vocabulary for the purposes of filtering first-year English majors lacking in lexis necessary for pursuing academic studies. Is it academic or low-frequency vocabulary that differentiates among candidates better? To find out, in 2003 the 10,000 word level (10K) of the Levels Test was administered. Two clusters (i.e. six items) were kept to serve as anchors linking the test to previous year's test for purposes of statistical analysis, which meant dropping two 10K level clusters of both Version 1 and Version 2 and substituting them with 2-2 clusters from the academic level. I hypothesized that these two clusters would prove to be less difficult and thus discriminate less than higher level clusters. The following academic level clusters were used as anchors from the previous year :

Version 1:

Cluster 5.
A colleague
B erosion   __F___ action against the law
C format   __B___ wearing away gradually
D inclination   __C___ shape or size of something
E panel
F violation

Cluster 8.
A anticipate
B compile   __E___ control something skillfully
C convince   __A___ expect something will happen
D denote   __F___ produce books and newspapers
E manipulate
F publish

Version 2:

Cluster 2.
A debate
B exposure   __E___ plan
C integration   __D___ choice
D option   __C___ joining something into a
E scheme   whole
F stability

Cluster 3.
A access
B gender   __B___ male or female
C implementation   __F___ study of the mind
D license   __A___ entrance or way in
E orientation
F psychology

The use of anchor items necessitated some practical modifications in the cluster number and sequence of the original Schmitt *et al.* (2001) test battery. To stick to the original idea of inserting 30 vocabulary items (ten clusters) only into the grammar and usage component of our proficiency test, clusters 1 and 2, and 1 and 7 were dropped from the 10K word level (Version 1 and Version 2, respectively) due to limitations of space and time available. In the new battery then, the above academic clusters were inserted to constitute the first and third clusters (Appendix G). Thus, it follows that in Version 1 items 1-3 and 7-9 (items 41-43 and 47-49 on the whole GAU test component) represent academic level vocabulary, similarly to items 1-3 and 7-9 in Version 2 (items 41-43 and 47-49 on the GAU test).

Version 1 was administered to 94 participants in May 2003, whereas Version 2 was taken by 117 students in June as part 3 of the grammar and usage component of the Proficiency test. Table 25 presents the results of all three component parts separately for the two versions, part 3 being the 30-item vocabulary test. The means of the two vocabulary tests exceed the 50 percent of the maximum score (30). The scores are spread out in a wide range (0-23 and 6-26), as shown by the scale of variance values. Correlations among the three subtests were again examined and significant correlations were found between the vocabulary and the grammar components (Version 1: $r_{13}= .446$, $p< 0.01$; $r_{23}= .448$, $p< 0.01$; Version 2: $r_{13}= .341$, $p< 0.001$; $r_{23}=$

.31, p< 0.01)

Table 25. Descriptive statistics on the 10K word level tests

GAU + 10K level	Version 1			Version 2		
Part	1	2	3 (10K)	1	2	3 (10K)
N of Items	30	10	30	30	10	30
N of Examinees	94	94	94	117	117	117
Mean	17.63	4.78	15.83	15.97	4.39	17.21
Variance	13.45	2.6	13.42	12.29	2.68	11.12
Std. Dev	3.67	1.61	3.66	3.51	1.64	3.34
Skew	-0.15	0.27	-0.87	-0.07	-0.12	-0.5
Kurtosis	-0.02	0.23	2.48	-0.41	-0.18	1.01
Minimum score	9	1	0	7	0	6
Maximum score	27	9	23	23	8	26
Median	18	5	16	16	4	17

The distribution of scores is illustrated below for both versions. On Version 1 the curve resembles a platykurtic distribution in 13, but instead of the plateau there are several small peaks at scores 13, 15, 17 and 19. The polygon is a bit negatively skewed as the mean is a little lower than it would be in a normal distribution, indicating that some students performed much better than we might have expected.
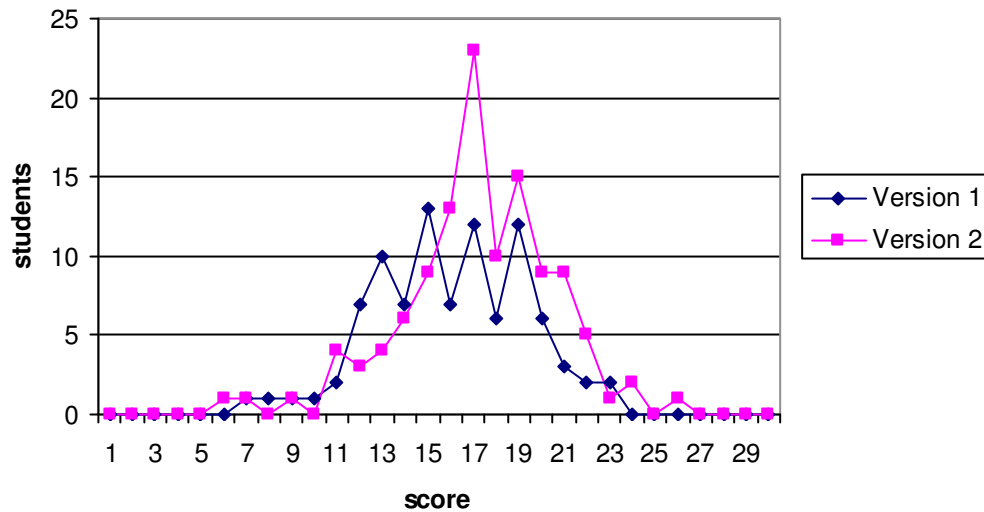
Figure 13. Score distribution on the 10K level tests

Running the same analysis on the scores of Version 2, we get a distinctly peaked, leptokurtic distribution, where skewness is somewhat higher, but still in the negative (-0.5), as shown in Figure 13. The mean (17.21) of this version was higher than that of Version 1 (15.83), with 56% of the examinees scoring below this score. The positive extreme scores of 23-26 out of 30 were achieved by four students only, shifting the mean upwards.

As the goal of assessing vocabulary in this test remained the same as in the previous year, i.e. to differentiate among candidates with larger and smaller vocabularies, therefore, it is again wise to examine the discrimination indices of the items. Mean discrimination indices on Version 1 and Version 2 are $M_1 = 0.259$ ($SD_1 = 0.158$) and $M_2 = 0.170$ ($SD_2 = 0.222$), respectively. Discrimination indices range from -0.03 to 0.55 on Version 1, and from 0.0 to 0.52 on Version 2 (Figure 14). A negative value indicates that the item works the opposite way it should, i.e. more low scorers than high scorers get the item right. There is only one such item, 30 on Version 2, whereas items 2, 4, 5 on Version 1 and items 26 and 27 seem to function quite satifactorily based on the guidelines provided by Ebel (1972, cited in Szabó, 2008, p. 32).
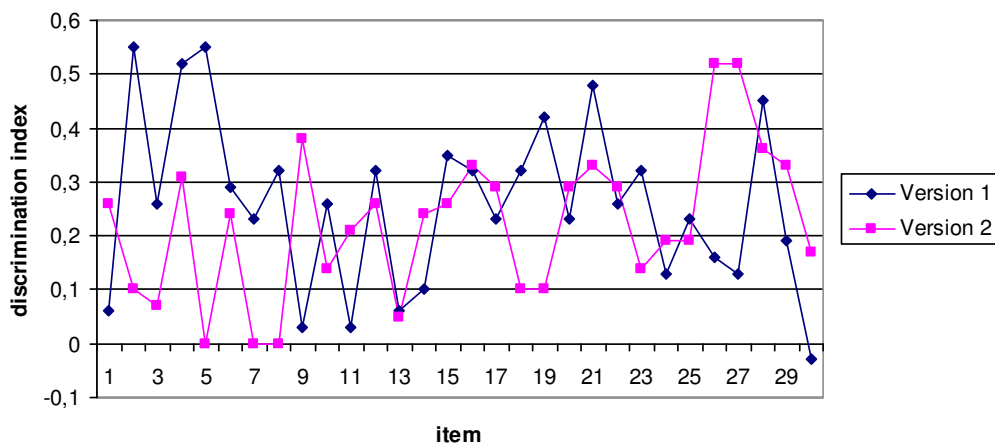
Figure 14. Item discrimination indices on the 10K-level tests

Comparing the discrimination indices and facility values of the six academic level anchor items (items 1,2,3,7,8,9) and the 24 10K word level items within this test (Table 26), it can clearly be seen that the academic level facility values exceed those of the 10K items. It means that these anchor items proved to be easier for the candidates than the low-frequency-level items. However, on Version 1 the gap in discrimination indices between the two item types is less striking.

Table 26. Mean facility values and discrimination indices of AWL and 10K items on the 10K-level tests

		Version 1 (N = 97)	Version 2 (N = 117)
AWL % correct	M	69.333	85.833
	SD	27.354	16.302
AWL DI	M	0.242	0.135
	SD	0.189	0.153
10K % correct	M	48.666	50.208
	SD	22.054	26.673
10K DI	M	0.263	0.244
	SD	0.153	0.127

Comparing the figures in Table 26 and the results presented in the previous section to those of the original VLT, Schmitt *et al.* (2001) reported higher mean facility values both on the academic word levels ($M_{AWL1} = 0.754$, $SD_{AWL1} = 0.094$ on Version 1 and $M_{AWL2} = 0.756$, $SD_{AWL2} = 0.108$) and the 10K word levels ($M_{10K1} = 0.289$, $SD_{10K1} = 0.176$; $M_{10K2} = 0.290$, $SD_{10K2} = 0.165$), as well as higher discrimination indicies on both levels ($M_{AWL1} = 0.519$, $SD_{AWL1} = 0.087$; $M_{10K1} = 0.509$, $SD_{10K1} = 0.233$ on Version 1; and $M_{AWL2} = 0.519$, $SD_{AWL1} = 0.074$; $M_{10K1} = 0.546$, $SD_{10K1} = 0.223$). A possible explanation to this may be the difference in the number of candidates taking the tests (94 and 117 versus 754 in Schmitt *et al.*, 2001), as both facility values and discrimination indices are sample dependent.

The original aim of introducing a vocabulary component in the first-year filter test was to filter out students lacking in vocabulary size necessary for academic studies. The findings of these two studies imply that although the two trials brought similar results, the academic word level tests yielded higher means and lower item discrimination

index values than the 10K word level tests, indicating that these words may be considered 'easier' for this population.

Even though the 10K level words of the VLT differentiated among high and low scorers better, thus fulfilling the original goal of testing better, these words do not seem to be useful in the context of our students, righteously raising the question why these words are tested. Therefore, in the subsequent year a new vocabulary test battery was developed partly made up of academic words and words intuitively perceived as useful in English studies by the researcher in order to tailor the battery to meet the special lexical needs of English majors. This constitutes stage three of the test development process.

6.5 Stage 3: A test of lexis perceived as useful in English studies

Learning the lessons from the findings of the previous two studies on academic and low-frequency vocabulary, the next year further modifications to the vocabulary component were implemented. As an abundance of studies focus on exploring the relationship between vocabulary knowledge and language proficiency (e.g., Morris & Cobb, 2003; Muncie, 2002; Qian, 2002; Zareva, 2005), the following sections report a study comparing performance on a vocabulary test to performance on reading comprehension, listening comprehension and essay writing tasks, and explore the relation of these to general language proficiency and academic performance of English majors.

6.5.1 Research questions

The study operates with ten variables (scores on a listening and reading comprehension, grammar and an essay writing test; receptive vocabulary test score; three productive vocabulary measures: type-token ratio, Uber index, "Beyond 2,000" words in the vocabulary profile of the essay writing test component; overall scores on the proficiency test; and the mean end-term grades of the participants). In an attempt to

explore how these are related to one another it seeks to answer the following four main research questions:

1) Is receptive or productive vocabulary a more reliable predictor of scores on a reading comprehension, a grammar and usage, a listening comprehension and a writing test?
2) How do the vocabulary measures relate to general language proficiency?
3) What is the relationship between receptive and productive vocabulary measures?
4) Which one of the measures investigated in the study is a good predictor of overall academic performance?

Findings are reported in the sequence of the above questions.

6.5.2 Participants

The study was conducted in May and June 2004, with the participation of 135 first-year English majors studying at the English Department, University of Pécs. 61 students (16 males, 45 females) took the May version, 74 students (16 males, 58 females) took the June version of the proficiency test, which is a compulsory milestone in the course of studies at the department (see Szabó, 1996, 2008). All the students were native speakers of Hungarian, studying as English majors.

6.5.3 Instruments

Three main research instruments were applied in the study to measure general language proficiency, receptive and productive vocabulary knowledge, all constituting a component of the milestone proficiency test developed by the test-writing team of the Department of English Applied Linguistics. The first component of the proficiency test was a listening comprehension task where students listened to two texts twice and filled in two 8-8-item multiple-choice type of tests. The second component of the

proficiency exam was a 40-item multiple choice grammar and usage test designed to measure the grammatical knowledge of the students. Similar to the listening comprehension tests, the third component was a two-text, 8-8 item multiple matching reading comprehension test. The May and June versions of the exam were carefully designed to be parallel versions (see Szabó, 1996, 2008).

Similarly to the previous two studies, in order to measure the receptive vocabulary knowledge of the students, the grammar and usage test was supplemented with a 30-item receptive vocabulary test devised for this purpose by the researcher. The test was identical in structure to the widely applied Vocabulary Levels Test (Laufer & Nation, 1999) which was pretested in the preceding two years. For the purposes of the present investigation, however, a new test was designed based on similar principles: the majority of the words to test were chosen from the Academic Word List (Coxhead, 2000), while some were selected on the basis of their intuitively perceived frequency and usefulness for students of English. The definitions attempted to utilize words from among the first 2,000 most frequent English words of the General Service List (West, 1953) and were kept as simple as possible. Students need to match three items of a cluster of six to three short definitions. There are ten clusters in the test, making it a total of 30 items to match (Appendix H).

Productive vocabulary use was studied in a guided essay writing task, the fourth component of the proficiency test. In the May version students were asked to write a 300-word essay on one of the following two topics: (1) The No-Curse Course; (2) Home Work. The June version similarly offered two topics to chose from with the titles of (1) The Ideal Job; or (2) Under the Weather? The essay tasks were scored first by two readers and the final score students received was the higher one given by the scorers. Two points could be earned for task completion, structures, vocabulary, and organization each, adding up to a maximum score of eight. The essays were typed into the computer by the researcher and in order to measure the lexical diversity of the student writings, the earlier discussed type-token ratio and the Uber-index were calculated for each essay, as shown in Chapter 3.

6.5.4 Procedures

The two test versions were administered on 25 May (Version 1) and 4 June 2004 (Version 2), between 9 and 12.30. Both exams followed the same procedure and the sequence of test components was as follows: (1) listening comprehension test (approx. 30 mins); (2) reading comprehension test (40 mins); (3) grammar and usage test (40 mins); a short break of 10 minutes, followed by (4) writing test (60 mins).

After the administration of the tests the essays of the candidates were scored manually by two scorers. The answers on the listening and reading comprehension, the grammar and the receptive vocabulary tests were typed into a computer and analyzed by ITEMAN version 3.5 (1993). Productive vocabulary measures were calculated by the researcher with the help of RANGE and an Excel file. Scores on the receptive vocabulary test and the productive vocabulary measures calculated from the writing component were then compared to scores on the listening comprehension, the reading comprehension and the grammar and usage sub-tests of the exam, as well as to overall performance on the proficiency test and the end-term mean grade of the students.

6.5.5 Results and discussion

In what follows first some descriptive statistics will be presented about student performance on each proficiency test component, proceeding to the analysis of the student writings and the calculated measures of lexical diversity. Receptive and productive vocabulary measures are then compared to results on each sub-test to find out whether receptive or productive vocabulary is a better predictor of success on reading comprehension, listening comprehension and grammar tests, general language proficiency and the overall academic performance of the participating students.

Before moving on in our discussion of the vocabulary component of the proficiency test it is worth devoting some attention to overall test performance. General descriptive statistics on Version 1 are presented below in Table 27. Scores on the 16-

item listening comprehension test range from 3 to 16 points, with a mean of 9.77 points. The 16-item reading comprehension test brought even better results, the mean being 12.39 points. The average score on the 8-item essay writing task was 4.34, a little higher than 50 percent, similar to the 40-item grammar test with a mean score of 23.8 points. The results on the receptive vocabulary test display a wide range of scores from 2 to 28 points, mean 17.43, standard deviation from the mean 5.521.

Table 27. General descriptive statistics on Test 3 Version 1

Test (max score)	N	Range	Minimum	Maximum	Mean	SD
Listening (16)	61	13	3	16	9.77	2.842
Reading (16)	61	8	7	15	12.39	1.855
Essay writing (8)	61	7	0	7	4.34	1.702
Grammar (40)	61	25	12	37	23.80	5.594
Receptive vocab (30)	61	26	2	28	17.43	5.52
Filter overall (100%)	61	57	35	92	62.93	13.19
Valid N (listwise)	61					

Version 2 brought similar results (Table 28). Means on the listening and the reading comprehension tests were 8.53 and 8.15 points, respectively. Candidates earned 4.18 points on average on the writing test and 18.23 points on the grammar test, while the mean on the receptive vocabulary test was 15.16 points, with ranges and standard deviations shown in Table 28 below.

Table 28. General descriptive statistics on Test 3 Version 2

Test (max score)	N	Range	Minimum	Maximum	Mean	SD
Listening (16)	74	10	3	13	8.53	2.506
Reading (16)	74	12	2	14	8.15	2.380
Writing (8)	74	7	1	8	4.18	1.175
Grammar (40)	74	27	7	34	18.23	4.870
Receptive vocab (30)	74	23	0	23	15.16	4.76
Filter overall (100%)	74	43	27	70	51.10	9.07
Valid N (listwise)	74					

6.5.5.1 Receptive vocabulary

Let us now examine the lexical profile of the vocabulary test components in detail. As McCrostie (2007) cited earlier points out, intuitive judgements of word frequency are not reliable. He emphasises the need for teachers and researchers to consult frequency lists rather than trusting their intuitions; therefore, the vocabulary profile of both test versions was drawn. Twelve Version 1 items come from the AWL: *abandon, advocate, arbitrary, conceive, derive, eliminate, facilitate, identical, persistent, prospective, substitution, sufficient.* Another 17 words (56.67% of all items) were chosen on the basis of usefulness in English studies as perceived by the researcher. These words are all low-frequency words, although not technical terms in any specialized discipline, that the Compleat Lexical Tutor developed by Tom Cobb (2000) calls "off-list words": *animosity, capacious, collage, conjunction, depict, digest, discourse, dismay, harness, illicit, indolent, infamous, plummet, proficient, salient, scaffold, vendetta.* One item, *scarce*, however, was a high-frequency word appearing among the first 1,000 most frequent words in the English language according to the General Service List (West, 1953).

Half (15 words) of the remaining distractors in the clusters came from the AWL (*adjacent, consistency, consume, diminishing, diverse, enhance, impact, initiate, intrinsic, marginal, obtainable, perception, pursue, supplement, valid*), whereas the other 50 percent (15 words) were again off-list, low-frequency words (*connotation, dub, ferocious, intrigue, mammoth, narrative, nugget, scan, semicolon, thrive, varnish, vast, verve, vindictive, wily).* Among the words in the definitions 51 types (85%) came from among the first 1,000 most frequent words of the GSL, five words (7.46%) from among the second 1,000 most frequent words of the GSL (*harm, lazy, lot, quickly, skill*), four words (5.97%) from the AWL (*legal, publication, remove, scheme*), and one word was off-list (*concise*). Table 29 presents the distribution of words in the two test versions.

Table 29. Vocabulary profile of the Stage 3 receptive vocabulary tests

		GSL 1	GSL 2	AWL	Off-list
Version 1	*items*	1 (3%)	0	12 (40%)	17 (57%)
	distractors	0	0	15 (50%)	15 (50%)
	definitions	51 (85%)	5 (7.46%)	4 (5.97%)	1 (1.49%)
Version 2	*items*	1 (3.33%)	1 (3.33%)	15 (50%)	13 (43.33%)
	distractors	0	3 (10.34%)	12 (41.38%)	14 (48.28%)
	definitions	50 (81.69%)	11 (15.49%)	0	2 (2.82%)

In test Version 2 the 15 AWL items constituted half of all the items tested (*access, accumulate, adapt, confer, crucial, empirical, grant, mature, mutual, obtain, orient, promote, sequence, sustainable, trace*). The other half included one word from among the first 1,000 most frequent words of the GSL (*reduce*), one word from among the second 1,000 most frequent words of the GSL (*lump*), and 13 off-list, low-frequency words (*adroit, annals, chase, contiguous, devour, hodgepodge, inaugurate, indigenous, liable, plunge, skim, tinge, unruly*). In a similar fashion, twelve of the distractor words originated from the AWL (*appropriate, commence, construct, contemporary, declining, dispose, distinct, expose, ignorant, persistent, reject, retain),* 14 words were off-list items (*abolish, allot, apprehension, clue, colossal, eternal, glaze, junction, overhaul, script, stratagem, target, clause, unveil, yoke*), and three words came from among the second 1,000 most frequent words of the GSL (*cultivate, improve, persuasive*).

Out of the 71 token words constituting the definitions to match with the items, 50 types were chosen from among the first 1,000 words of the GSL (*a, after, amount, based, bring, by, common, control, declare, developed, difficult, direction, each, eat, find, following, for, fully, get, give, important, in, into, make, money, native, next, observation, of, on, open, order, other, piece, put, reach, read, run, shared, signs, small, smaller, something, studies, support, through, to, together, turn, very*). Eleven words came from among the second 1,000 most frequent GSL words (*behavior, confused, decrease, eagerly, encourage, film, mixture, quick, quickly, responsible,*

solid), while two words (*chronicles, tolerable*) were off-list words.

Table 30. Descriptive statistics on Test 3

GAU + Test 3	Version 1			Version 2		
Part	1	2	3	1	2	3
N of Items	30	10	30	30	10	30
N of Examinees	61	61	61	74	74	74
Mean	16.89	6.92	17.43	13.15	4.93	15.16
Variance	18.27	2.99	29.98	13.59	3.6	22.62
Std. Dev	4.27	1.73	5.48	3.69	1.9	4.76
Skew	0.06	-0.2	-0.36	0.5	0.11	-0.95
Kurtosis	-0.41	-0.5	-0.28	0.36	-0.25	1.49
Minimum score	9	3	2	6	1	0
Maximum score	27	10	28	24	10	23
Median	17	7	18	13	5	16

Version 1 of the vocabulary test was administered to 61 students in May, whereas Version 2 was taken by 74 students in June 2004 as part three of the grammar and usage test component. Hence the vocabulary test results are shown in column 3 of Table 30. The means on the two vocabulary tests exceed fifty percent of the maximum score (30). Measures of dispersion show that the scores are spread out in a wide range (2-28 and 0-23). Correlations among the three subtests were again examined and significant correlations were found between the vocabulary and the grammar components on Version 1 (r_{12}= .580, p<0.01; r_{23}=.481, p<0.01); but no significant correlation was shown on Version 2.

The distribution of scores on the receptive vocabulary test is shown in Figure 15 on the two test versions. The distribution of scores is rarely symmetric like the bell-shaped Gauss-curve. The skewness of a dataset indicates whether the deviations from the mean are going to be positive or negative; therefore, skewness is a measure of asymmetry in the probability distribution. It is in the negative on both tests implying a longer left tail of the distribution curve with relatively few low values. However, the kurtosis figure as a measure of the peakedness of the distribution differs on the two test occasions. The negative kurtosis (-0.28) on Version 1 as opposed to the positive kurtosis (1.49) on Version 2 indicates that the variance in the distribution on Version 1 is due to frequent modestly sized deviations rather than infrequent extreme deviations

from the mean.
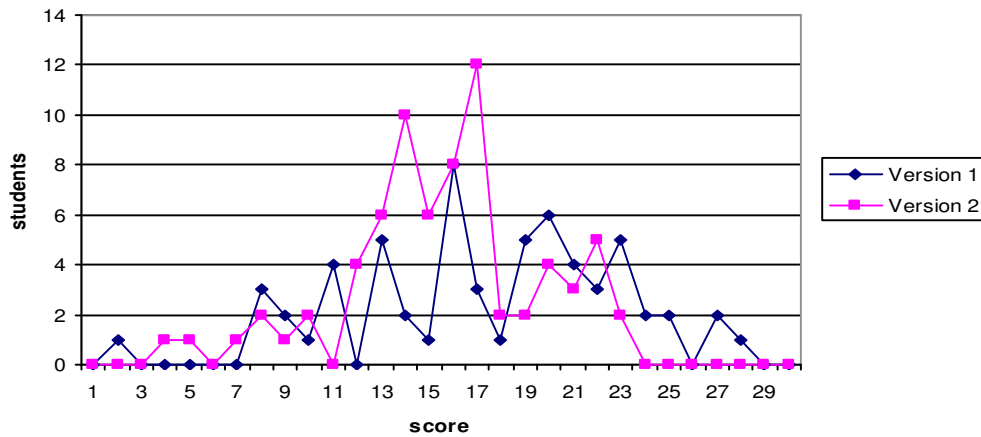
Figure 15. Score distribution on Test 3

There was one item that all students got right on Version 1: *discourse* (Figure 16). Although it is an off-list, i.e. a relatively low-frequency word in the English language, first-year English majors must have encountered it often in academic discourse. Further three words were known to over 80% of the participants: the AWL items *abandon* (84%) and s*ufficient* (80%), and the low-frequency word *proficient* (84%). The most difficult items known by fewer than one third of the test-takers all proved to be low-frequency words as expected: *harness* (16%), *plummet* (21%), *scaffold* (23%), *indolent* (28%), *salient* (31%), *digest* (34%), and *animosity* (34%).
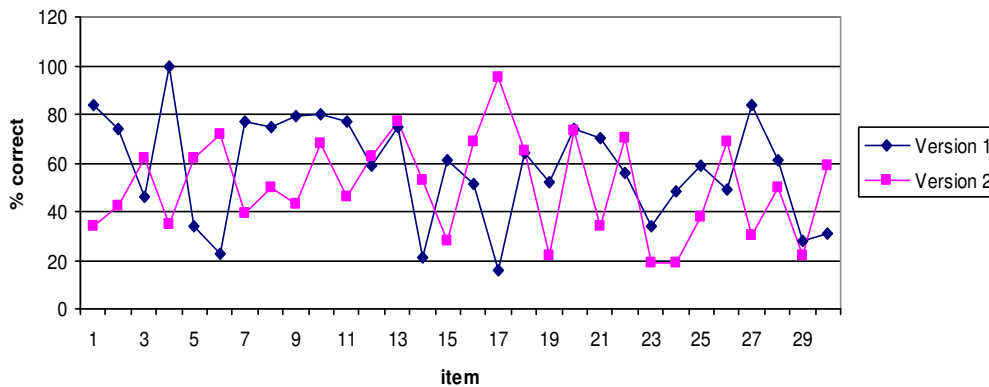
Figure 16. Facility values on Test 3

On Version 2 only one item was known to over 80 % of the testees: *reduce* from among the first 1,000 most frequent words in the English language (GSL1). Interestingly, the other GSL item, *lump* (30%), was familiar to only one third of the students, similarly to the off-list *devour* (19%), *inaugurate* (19%), *adroit* (22%), *plunge* (22%), *skim* (34%), or the AWL items *accumulate* (28%) and *confer* (34%). The relatively easy words seem to be the academic words in the test: *orient* (77%), *mature* (72%), *obtain* (69%). On the whole, the academic words seem to be more familiar to the testees on both test versions, as shown by the number of students getting an item right. The facility values of the individual items ('% correct' in Table 31) indicated that on average 69.67% (SD_1: 11.94) and 56.93% (stdev2: 14.97) of the students got the AWL words right on Version 1 and 2, respectively, while the figures are somewhat lower for the low-frequency words (M_1: 50.41%, SD_1: 24.17 and M_2: 40.69% SD_2: 18.44). A two-tailed independent samples t-test, however, did not reveal any significant difference in the facility values between the two test versions (t= 0.152, p> .05).

The difficulty of an item, however, is indicated by its discrimination index. The higher the discrimination index of an item the better it differentiates among high-scorers and low-scorers on a test, therefore, the better the item works in the test. In this test some items do not discriminate at all, thus need to be discarded and replaced in possible future applications of the test, whereas the majority seem to work satisfactorily. Discrimination indices range from 0 (*discourse)* to 0.75 (*depict*) on Version 1, and from 0.1 (*orient* and *inaugurate*) to 0.58 (*adapt, obtain* and *chase*) on Version 2.

Table 31. Vocabulary item statistics on Test 3

Version 1	List	% correct	DI	Version 2	List	% correct	DI
1.abandon	AWL	84	0.44	1.confer	AWL	34	0.3
2.conceive	AWL	74	0.45	2.trace	AWL	42	0.28
3.facilitate	AWL	46	0.38	3.access	AWL	62	0.4
4.discourse	OFF	100	0	4.annals	OFF	35	0.25
5.animosity	OFF	34	0.5	5.grant	AWL	62	0.13
6.scaffold	OFF	23	0.09	6. mature	AWL	72	0.33
7.identical	AWL	77	0.38	7.indigenous	OFF	39	0.42
8.persistent	AWL	75	0.57	8.liable	OFF	50	0.45
9.capacious	OFF	79	0.33	9.mutual	AWL	43	0.52
10.sufficient	AWL	80	0.56	10.empirical	AWL	68	0.48
11.prospective	AWL	77	0.38	11.sustainable	AWL	46	0.42
12.arbitrary	AWL	59	0.58	12.crucial	AWL	63	0.48
13.derive	AWL	75	0.16	13.orient	AWL	77	0.1
14.plummet	OFF	21	0.47	14.sequence	AWL	53	0.22
15.depict	OFF	61	0.75	15.accumulate	AWL	28	0.3
16.advocate	AWL	51	0.72	16.obtain	AWL	69	0.58
17.harness	OFF	16	0.31	17.reduce	GSL1	95	0.2
18.eliminate	AWL	64	0.59	18.adapt	AWL	65	0.58
19.vendetta	OFF	52	0.6	19.plunge	OFF	22	0.32
20.substitution	AWL	74	0.46	20.chase	OFF	73	0.58
21.conjunction	OFF	70	0.53	21.skim	OFF	34	0.52
22.dismay	OFF	56	0.53	22.promote	AWL	70	0.3
23.digest	OFF	34	0.52	23.devour	OFF	19	0.13
24.collage	OFF	48	0.65	24.inaugurate	OFF	19	0.10
25.infamous	OFF	59	0.41	25.tinge	OFF	38	0.42
26.scarce	GSL1	49	0.56	26.hodgepodge	OFF	69	0.22
27.proficient	OFF	84	0.32	27.lump	GSL2	30	0.38
28.illicit	OFF	61	0.47	28.contiguous	OFF	50	0.12
29.indolent	OFF	28	0.23	29.adroit	OFF	22	0.22
30.salient	OFF	31	0.28	30.unruly	OFF	59	0.13

It also becomes clear from a closer examination of item characteristics that the discriminating capacities of AWL and Off-list items are similar on both test versions (AWL M_1: 0.47, SD_1: 0.14; M_2: 0.36, SD_2:0.15, Off-list M_1: 0.41, SD_1: 0.2; M_2: 0.30, SD_2: 0.16), as opposed to the findings previously reported in Stage 2. Item statistics, however, not only indicated the strengths, but also revealed weaknesses of the tests. Item 6 on Version 1 (*scaffold*) and items 23 and 24 (*devour* and *inaugurate*) on Version 2 did not prove to work well and thus should be replaced or rewritten for any possible future application of the battery.

6.5.5.2 Free productive vocabulary

Let us now move on to the analysis of student writings to explore their productive vocabulary. As discussed earlier in chapter 3 and shown by the findings of the preliminary study reported in chapter 5, numerical measures of several kinds may be applied to describe the lexis of texts. Some of these include the type-token ratio (or lexical variation), lexical sophistication, lexical density, and the number of errors. The recent involvement of mathematicians in applied linguistic research has made it possible to describe vocabulary by algebraic transformations, leading to formulae such as Herdan's C, Guiraud's R, Uber's U, and Zipf Z index. Based on the recognition that the much criticized type-token ratio is sensitive to text length, Jarvis (2002) comparing and assessing these measures found that the Uber-index seems to be the best available descriptor of the lexical diversity of short texts, as it eliminates the distorting effects of varying text length that the type-token ratio is not able to handle. Therefore, both measures were applied in the present analysis to enhance the reliability of the investigation.

The careful handling of a corpus is crucial in data analysis. Subsequent to the administration of the tests, the student essays were typed into the computer, each was saved as 'raw text' in a separate Word document file and named with the indication of the code number of the essay, the initial of the topic and the initial of the month of administration, e.g., '01CM' for an essay written about courses in May with a code number 01. Several sub-corpora were created. First, the essays of the two test

occasions were saved in two separate Word files named 'rawmay2004.doc' and 'rawjune2004.doc'. Then, four further sub-corpora were saved on the basis of the four essay topics, named 'homemay.doc', 'coursemay.doc', jobjune.doc' and 'weatherjune.doc'. This categorization later made it possible to examine both individual and group-specific features of the student writings. Finally, all texts were saved as 'text only' to make them suitable for computerized data analysis.

On the whole, the student essays varied vastly in length on both test versions, even though the instructions asked the candidates to write approximately 300 words. The number of tokens, however, ranged from 215 words to 586 words (M: 382,21; SD.: 74,36) on Version 1, and from 243 to 655 (M: 385.67; SD: 81.57) on Version 2, which makes it difficult to compare the writings based on the type/token ratio only, as that would require texts of similar length (Jarvis, 2002). Type-token ratios, however, did not differ significantly on the two test versions; they ranged from 0.36 to 0.6 on Test 1 (M: 0.48; SD: 0.5) and from 0.34 to 0.61 on Test 2 (M: 0.47; SD: 0.57), neither did the Uber-indices show significant differences in lexical diversity on the two occasions, U values ranging from 15.74 to 26.79 on Version 1 (M: 20.92; SD: 2.93) and from 15.0 to 28.79 on Version 2 (M: 20.46 ; SD: 2.67), as shown in Tables 32 and 33.

Table 32. Descriptive statistics of productive vocabulary on Version 1 of Test 3

	N	Range	Minimum	Maximum	Mean	Std. Deviation
Token	61	371.00	215.00	586.00	382.2131	74.36220
Type	61	118.00	128.00	246.00	180.8361	26.76327
Type/token ratio	61	0.24	0.36	0.60	0.4797	0.05099
Uber index	61	11.05	15.74	26.79	20.9213	2.32772

Table 33. Descriptive statistics of productive vocabulary on Version 2 of Test 3

	N	Range	Minimum	Maximum	Mean	Std. Deviation
Token	74	412.00	243.00	655.00	385.6757	81.56955
Type	74	150.00	127.00	277.00	178.2973	28.07950
Type/token	74	0.27	0.34	0.61	0.4697	0.05707
Uber index	74	13.79	15.00	28.79	20.4604	2.66730

It is legitimate to assume that the topic of the essays may influence the length as well as the lexis of the texts; therefore, it is worth devoting some attention to this possible effect here. The distribution of the essays in this respect was uneven in May, as only twelve students chose to write about courses, while 49 opted to elaborate on their homes. On the June test occasion, however, the topics seemed to be equally appealing to the participants; 35 essays were written about the ideal job and 39 about weather. Again, examining the same descriptive statistics of productive vocabulary, this time re-categorized for essay topic (Table 34), it can be seen that the topic of the essay does not seem to be an influential factor on the length (Token) or the lexical diversity of the texts (T/t and U) with this population.

Table 34. Case summaries for essay topic

Essay topic		Text length (Token)	T/t	Uber index	Academic words %	Off-list words%	˝Beyond 2000" %
Job	N	35	35	35	35	35	35
	Mean	381.171	.460	19.981	5.248	5.080	10.328
	SD	13.614	.0089	.346	.392	.390	.624
Weather	N	39	39	39	39	39	39
	Mean	385.026	.475	20.890	3.795	7.123	10.917
	SD	13.370	.010	.494	.287	.576	.632
Course	N	12	12	12	12	12	12
	Mean	382.083	.452	19.383	6.742	5.258	10.908
	SD	26.380	.018	.73850	.725	.769	1.487
Home	N	49	49	49	49	49	49
	Mean	382.408	.484	21.189	3.715	7.877	11.355
	SD	10.216	.006	.298	.296	.335	.477
Total	N	135	135	135	135	135	135
	Mean	382.815	.473	20.629	4.410	6.693	10.923
	SD	6.745	.005	.215	.198	.259	.324

As Laufer (2001) pointed out, beyond the use of the most frequent 2,000 words in the English language, often referred to as K1 and K2, the ratio of academic words and infrequent (off-list) words in any text is a more adequate descriptor of its lexical diversity. In the present context, therefore, it would seem reasonable to expect a high ratio of academic and infrequent words in texts produced in a test of essay writing. Before moving on in our analysis, however, an important technical step undertaken in

data handling needs to be mentioned here. Research literature recommends the exclusion of proper nouns, numbers and misspellt words from vocabulary profiling (Laufer, 2001, p. 244), as the frequent use of these words appearing as off-list items may draw a destorted picture of the lexis in the texts. This proved to be an important factor to consider in the present context, as after the first RANGE analysis run on the corpora it became clear that the essays on home contained an irrationally higher percentage of off-list words than the other essays due to frequent references to Hungarian towns and place names. With the above consideration in mind, therefore, a 426-word stoplist was created in RANGE, so that the software should count these words as part of all tokens but not list them as infrequent words in the English language in the off-list section. The titles of the essays were also removed from the corpus.

Function words appearing a thousand times or more in the four sub-corpora were *the* (3,408), *and* (1,852), *to* (1,764), *I* (1,533), *a* (1,523), *is* (1,434), *of* (1,433), *it* (1,165), *in* (1,142), their rank order being very similar to the findings of Horváth (2001, p. 112) in the JPU student corpus. However, RANGE makes it possible to focus on content words only to explore the lexical density of the texts. Table 35 presents the content words appearing over a hundred times in the corpus, with the indication of their range and frequencies in the four sub-corpora. Examining the data more closely, it does not strike as unexpected that the topic of courses and jobs seem to have evoked the use of more academic words (6.7 and 5.2 percent of all tokens, respectively), while students elaborating on the weather and their homes seem to have resorted to such words less frequently, enriching their writings on the other hand with a higher percentage of unique, infrequent words not appearing in the other two sub-corpora. In essays about courses the most frequently used academic word families were *topic* (31), *goal* (22) and *task* (19), students writing about their homes mentioned *area* (33), *culture* (27), *job* (20) and *environment* (19) most often. The lexis of essays about the ideal job centred around the words *job* (189), *create* (21), *culture* (18) and *task* (12), while the topic of the weather activated words such as *affect* (62), *energy* (24), *positive* (19) and *period* (13), as shown below in Table 35.

Table 35. Word forms appearing 100 times or more in the essay corpus

TYPE	RANGE	FREQ	CM	HM	JJ	WJ
PEOPLE	4	610	13	254	120	223
GOOD	4	310	28	48	54	180
ALSO	4	220	32	100	31	57
VERY	4	219	31	78	61	49
SO	4	214	24	81	40	69
THINK	4	205	17	63	55	70
COURSE	4	199	149	20	12	18
LIFE	4	190	14	40	41	95
MAKE	4	186	29	77	35	45
ONLY	4	164	15	81	34	34
BAD	4	160	1	12	7	140
GO	4	147	3	68	8	68
IMPORTANT	4	143	37	44	45	17
FEEL	4	142	7	19	15	101
WORK	4	136	7	25	83	21
ALWAYS	4	117	5	29	42	41
PLACE	4	116	1	101	8	6
TOO	4	113	11	44	14	44
TIME	4	112	6	36	23	47
TEACHER	3	110	51	0	58	1
JUST	4	106	18	26	12	50
LIVE	4	105	1	80	6	18
THINGS	4	105	14	29	33	29
WELL	4	105	19	40	31	15
DAY	4	101	1	9	18	73
STUDENTS	4	100	77	5	15	3

Table 36. The most frequent academic word families in the essay corpus

AWL Families	RANGE	TYFREQ	FAFREQ	CM	HM	JJ	WJ
JOB	4	203	221	3	20	189	9
TOPIC	4	34	44	31	3	9	1
ENERGY	3	24	30	0	3	3	24
ENVIRONMENT	2	24	25	0	19	0	6
POSITIVE	3	23	24	0	4	1	19
TASK	4	23	38	19	4	12	3
AREA	4	21	38	1	33	3	1
GOAL	3	16	28	22	3	3	0
PRIMARY	2	16	17	0	12	5	0
ROLE	4	14	14	3	5	2	4
AFFECT	3	13	64	0	1	1	62
CULTURE	3	13	49	4	27	18	0
PERIOD	2	13	17	0	4	0	13

'Weather' was most often associated with the words *mood* (94), *rainy* (28), *climate* (16), *nervous* (12), *beach* (12) and *headache* (12), while 'home' with *inhabitants* (44), *tourism* (25), *atmosphere* (21) and *county* (19). As mentioned earlier, the sub-corpora of 'course' and 'job' contained less low-frequency words (5,2 and 5,08 % of all tokens,

respectively), some of these were *boring* (10) and *seminar* (11) for 'course' and *reporter* (16) and *huge* (6) for 'job'.

Table 37. The most frequent off-list word forms in the essay corpus

TYPE	RANGE	FREQ	*CM*	*HM*	*JJ*	*WJ*
MOOD	3	96	0	1	1	94
INHABITANTS	2	45	0	44	1	0
RAINY	1	28	0	0	0	28
ETC	4	26	4	12	5	5
TOURISM	2	26	0	25	1	0
ATMOSPHERE	4	25	1	21	2	1
BORING	4	25	10	8	4	3
CLIMATE	2	20	0	4	0	16
COUNTY	1	19	0	19	0	0
HUGE	3	19	0	7	6	6
CINEMA	1	16	0	16	0	0
RIPORTER	1	16	0	0	16	0
MUSEUMS	1	14	0	14	0	0
NERVOUS	2	14	0	0	2	12
BEACH	2	13	0	1	0	12
FESTIVALS	1	13	0	13	0	0
HOMETOWN	2	13	0	11	2	0
MUSEUM	2	13	0	12	1	0
OPTIMISTIC	3	13	1	2	0	10
PUBS	1	13	0	13	0	0
BORED	4	12	1	5	1	5
CONCERTS	1	12	0	12	0	0
HEADACHE	1	12	0	0	0	12
SEMINAR	1	11	11	0	0	0
GERMAN	2	10	0	9	1	0
MEDITERRANEAN	2	10	0	6	0	4
MONUMENTS	2	10	0	9	1	0

6.5.5.3 Vocabulary and the four proficiency subtests

Having calculated both the receptive and the productive measures of vocabulary, in the next stage of the study the sub-test scores were compared to receptive and productive vocabulary measures in order to answer the research questions outlined at the beginning of this section. The Pearson correlation matrixes are shown in Tables 38 and 39 for the two test versions.

Correlations on Version 1 presented in Table 38 show that all the four subtests (listening, reading writing, grammar) modestly correlate among each other. The figures also revealed that receptive vocabulary seems to be a good predictor of success

on all four sub-tests of the proficiency test. Significant correlations were found between receptive vocabulary and scores on the listening comprehension test (r= .476; p < .01), receptive vocabulary and the reading comprehension test (r= .332; p < .01), receptive vocabulary and the essay writing component (r= .609; p < .01), and receptive vocabulary and the grammar component (r= .597; p < .01) of the proficiency test. Examining the same correlations on Version 2 (see Table 39), although receptive vocabulary correlates significantly but weakly with scores on the reading comprehension sub-test (r= .326; p < .01), as expected from results on Version 1, no significant correlation was found with the other three sub-tests (listening comprehension, essay writing and grammar).

As for productive vocabulary measures, none of the three indices (T/t, U, Beyond 2000) correlate with any of the sub-test scores on Version 1, and all the three correlate significantly with scores on the grammar subtest on Version 2. Interestingly, grammar scores correlated significantly with all the other subtests and receptive vocabulary on Version 1 as well, supporting the assumption mentioned earlier in the previous chapters that knowledge of grammar is closely related to all four skills. Concerning, therefore, the first research question, these findings seem to imply a stronger relationship between receptive vocabulary and performance on a listening, reading, writing and grammar test than free productive vocabulary described by the above three indices.

The interpretation of these findings may be multi-fold. The correlation between receptive vocabulary scores and scores on the sub-tests measuring receptive skills, such as reading and listening comprehension, may not be an unexpected result as it has long been shown that a substantial percentage of words constituting a text must be familiar to the reader or listener for successful comprehension to take place, as has been discussed in the introductory chapters.

Table 38. Correlations on Version 1 of Test 3 (N= 61)

		Listening	Reading	Essay writing	Grammar score	Receptive vocab	Type/token ratio	Uber index	Beyond 2000
Listening	Pearson Correlation	1	,482**	,427**	,601**	,476**	,177	,220	,095
	Sig. (2-tailed)		,000	,001	,000	,000	,172	,089	,467
Reading	Pearson Correlation	,482**	1	,532**	,642**	,332**	,138	,203	,095
	Sig. (2-tailed)	,000		,000	,000	,009	,288	,116	,466
Essay writing	Pearson Correlation	,427**	,532**	1	,620**	,609**	-,092	,111	,175
	Sig. (2-tailed)	,001	,000		,000	,000	,482	,394	,178
Grammar score	Pearson Correlation	,601**	,642**	,620**	1	,597**	,121	,197	,107
	Sig. (2-tailed)	,000	,000	,000		,000	,354	,128	,410
Receptive vocab	Pearson Correlation	,476**	,332**	,609**	,597**	1	,053	,199	,277*
	Sig. (2-tailed)	,000	,009	,000	,000		,684	,125	,031
Type/token ratio	Pearson Correlation	,177	,138	-,092	,121	,053	1	,896**	,145
	Sig. (2-tailed)	,172	,288	,482	,354	,684		,000	,266
Uber index	Pearson Correlation	,220	,203	,111	,197	,199	,896**	1	,240
	Sig. (2-tailed)	,089	,116	,394	,128	,125	,000		,063
Beyond 2000	Pearson Correlation	,095	,095	,175	,107	,277*	,145	,240	1
	Sig. (2-tailed)	,467	,466	,178	,410	,031	,266	,063	

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

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

Table 39. Correlations on Version 2 of Test 3 (N = 74)

		Listening	Reading	Writing	Grammar	Receptive vocab	Type/token	Uber index	Beyond 2000
Listening	Pearson Correlation	1	,274[*]	,140	,225	,085	,124	,091	,042
	Sig. (2-tailed)		,018	,233	,054	,474	,294	,442	,720
Reading	Pearson Correlation	,274[*]	1	,250[*]	,115	,326[**]	-,086	-,068	,122
	Sig. (2-tailed)	,018		,031	,328	,005	,464	,565	,299
Writing	Pearson Correlation	,140	,250[*]	1	,263[*]	,020	,029	,190	,092
	Sig. (2-tailed)	,233	,031		,023	,869	,804	,104	,437
Grammar	Pearson Correlation	,225	,115	,263[*]	1	,146	,301[**]	,399[**]	,261[*]
	Sig. (2-tailed)	,054	,328	,023		,215	,009	,000	,025
Receptive vocab	Pearson Correlation	,085	,326[**]	,020	,146	1	,103	,096	,239[*]
	Sig. (2-tailed)	,474	,005	,869	,215		,385	,417	,040
Type/token	Pearson Correlation	,124	-,086	,029	,301[**]	,103	1	,901[**]	,300[**]
	Sig. (2-tailed)	,294	,464	,804	,009	,385		,000	,009
Uber index	Pearson Correlation	,091	-,068	,190	,399[**]	,096	,901[**]	1	,414[**]
	Sig. (2-tailed)	,442	,565	,104	,000	,417	,000		,000
Beyond 2000	Pearson Correlation	,042	,122	,092	,261[*]	,239[*]	,300[**]	,414[**]	1
	Sig. (2-tailed)	,720	,299	,437	,025	,040	,009	,000	

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

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

Neither is the correlation of grammar scores with both receptive and productive vocabulary measures surprising in the light of lexico-grammar, a recent model of the mental representation of the lexicon and syntax proposed by Halliday (1985) within the framework of his Systemic Functional theory of language combining syntax, lexicon and morphology. This assumption is further strengthened by Ullman (2001), who argues that while grammar is subserved by the procedural memory in L1, experimental evidence shows that in second language learning grammatical rules are largely memorized by declarative memory, just like lexical items. Allison Wray's (2002) Heteromorphic Distributed Lexicon also postulates that lexicon and syntax are inseparable, suggesting that in language testing they may be treated as one construct.

6.5.5.4 Receptive versus free productive vocabulary

The second research question concerned the relationship between receptive and productive vocabulary. It has been shown in the previous chapters that receptive vocabulary is larger than free productive vocabulary, and a strong correlation was expected between these two measures of student lexis. From data in Tables 38 and 39 it can be concluded that of the three indices of productive vocabulary only the ratio of the 'Beyond 2,000' words was in a weak but significant relationship with receptive vocabulary on both test versions (r_1= .277; p < .05 on Version 1 and r_2= .239; p < .05 on Version 2).

This finding seems to support Laufer and her colleagues' (2004) assumption that the T/t and the various indices derived from it are not the most adequate descriptors of the lexical richness of short texts and suggests that vocabulary profiling gives a better indication of lexical diversity. On the other hand, it partly contradicts the findings of other studies (e.g.,Vermeer, 2000), where the widely applied T/t was also found to be a sadly inadequate measure, but the logarithmic Uber-index had a significant correlation with receptive vocabulary. Similarly, Jarvis (2002) found a modest correlation between scores on a discrete-point self-report vocabulary test and lexical diversity in written narratives with Finnish learners of English suggesting that writing proficiency is

influenced by vocabulary knowledge.

Another explanation to the fact that no significant correlation was found between receptive and productive vocabulary measures may be that their development is not parallel (Laufer, 1998) and the size of the receptive vocabulary of these candidates far exceeded that of their productive vocabulary, which is often the case with second language learners. The conversion of passive vocabulary into active is only the final stage of lexical development (Brown & Payne, 1984, cited in Muncie, 2002, p. 226) which is supported by Laufer's findings (2001) that over a one-year period passive vocabulary increased the most while free active vocabulary developed the least. Furthermore, Jarvis (2002, p. 80) suggests that lexical diversity is "most affected by vocabulary knowledge in the relatively early stages of L2 acquisition", while the students in the present study were advanced learners of English.

It is more likely to assume, though, that the lack of significant relationship lies in the inadequacy of describing productive vocabulary with numerical measures only. Writing quality is more complex than the number of words used and the frequency distribution of vocabulary only, as demonstrated by a wealth of research into writing processes (Chiang, 1999; Hedgecock & Lefkowitz, 1999; Henry, 1999, Henry & Roseberry, 2007, Horváth, 2001; Lindgren & Sullivan, 2002; Muncie, 2002; New, 1999). Consequently, it is not by surprise that the T/t-derived and word frequency-based indices in this study are not related to overall essay scores taking task achievement, cohesion, coherence, accuracy and organization into consideration. Another caveat to bear in bind about lexical frequency profiling is its incapability of handling lexical phrases or idioms generally considered to be evidences of sophisticated language use (Muncie, 2002).

6.5.5.5 Vocabulary and general language proficiency

The third research question focused on exploring the relationship between receptive and productive vocabulary measures and general language proficiency. A measure of language proficiency in this study is the overall filter test score expressed in percentages gained by adding up the percentage points on the four subtests and dividing it by four. As this calculation excludes receptive vocabulary scores - in contrast with the original scoring system of the filter test, where receptive vocabulary is part of the grammar and usage component of the test - it is legitimate to perform correlation and regression analysis on the data. A Pearson correlation coefficient matrix revealed that only receptive vocabulary of the four independent variables entered in the analysis showed a high correlation with general language proficiency (r_1= .718; p < .01 on Version 1 and r_2= .400; p < .01 on Version 2) confirming the hypothesis that overall language proficiency and receptive vocabulary are not independent of each other. Correlations, however, do not mean causal relationships between factors.

While correlation allows us to establish the strength of a relationship between variables, regression makes it possible to predict scores on a test based on scores on another test (Hatch & Lazaraton, 1991). In simple regression the performance on a dependent variable is predicted by performance on an independent variable. In multiple regression it is possible to see which independent variable or which combination of independent variables allow us to make the most accurate predictions (Hatch & Lazaraton, 1991). In our case, this way we can determine which one of the four variables describing student lexis or which combination of these variables best explain the largest ratio of variance in overall language proficiency, and which ones do not add any effect.

Therefore, overall language proficiency (dependent variable) was regressed against receptive and productive vocabulary (independent variables) on both test versions. Two descriptors of free productive vocabulary, the T/t and the Uber-index, were

dropped from the analysis for multiple reasons. First, the intercorrelation of these two variables is higher than .80 (Table 38 and 39), which is likely to cause multicolinearity in multiple regression. Second, the number of participants on each test version does not allow entering a larger number of independent variables, as according to Hatch and Lazaraton (1991, p. 481), it is a general rule of thumb that we need to have 30 students for each variable. Thus, the more variables entered into regression, the larger the necessary N size. Third, as we have seen from previous correlation figures, the ratio of infrequent words is a better predictor of performance on listening and reading comprehension, essay writing and grammar tests than the T/t derived measures.

Hence, overall language proficiency was regressed against receptive vocabulary and 'Beyond 2,000', $F(2, 58)=30.908$, $p<.001$ on Version 1 (Appendix I). The square of R, the coefficient of determination, was 0.516, and significant at the $p<.001$ level, which means that the two independent variables together explained 51.6 percent of all the variance of the overall language proficiency score. As the two independent variables weakly but significantly correlate with each other ($r=.277$; $p<.05$), it is necessary to examine both the total and their unique contributions to the dependent variable.

Regressing the two variables separately against overall language proficiency, it became clear that the F value for receptive vocabulary was significant ($F=62.793$; $p<.001$), indicating that it explained a significant proportion of variance in overall language proficiency scores ($R^2=0.516$, SEE=0.092). Interestingly, however, the F value for 'Beyond 2,000' was not significant ($F=1.996$), the $R^2=0.033$ explaining only 0.33%, a very little portion of the variance in proficiency scores. Next, in order to get the individual contribution of each independent variable, the squared semipartial correlations were calculated between overall language proficiency and each independent variable. The values indicate that receptive vocabulary alone explained 69.5 percent of all variance in overall language proficiency scores, while 'Beyond 2,000' showed a negative correlation (-0.019).

The same analysis was performed on test Version 2 (Appendix J). The correlations

174

between receptive and "Beyond 2000" vocabulary and overall language proficiency were both significant (r_R= .400; p < .01 and r_{B200}= .204; p < .05). $F_{(2, 71)}$=7,391 was significant at the p< .001 level. The R squared was 0.172, and significant at the p<.001 level, hence the two independent variables together explained 17.2% of all the variance in the overall language proficiency score, a much smaller proportion than on Version 1. The individual contribution of receptive vocabulary to this variance was found to be 16% (R^2=0.160, SEE=0.084), while the contribution of the ratio of 'Beyond 2,000' words representing lexical diversity in free productive vocabulary can be neglected (R^2=0.041, SEE=0.089).

6.5.5.6 Vocabulary and academic success

The fourth main research question addressed in this section concerns how well each variable investigated in the study predicts overall academic performance or success. Therefore, Pearson correlation matrixes were repeatedly drawn on both datasets, this time to find out about the strength of the relationship between the dependent variable (endterm grade) and the independent variables (subtest scores and the two vocabulary measures). The endterm grade was calculated as the mean of the students' two mean grades at the end of the first and the second semester.

As indicated in Table 40 below, on Version 1 the listening and essay writing scores of the students are moderately correlated with their endterm grades (r_L= .593 and r_E= .351; p < .01), implying a relationship between these factors, and there is a weaker but also significant correlation shown with reading and grammar scores (r_R= .324 and r_G= .305; p < .05). No relationship was found here with the receptive and productive vocabulary measures, while on test Version 2, none of the measures correlated with the endterm grade, and receptive vocabulary even showed a reverse relationship (r_{RV}= - .123). Consequently, as the correlations found are weak, there is no ground for regression analysis to explore the predictive values of the factors in this case.

Table 40. Correlations of measures with endterm grade

		Endterm grade N_1=61	Endterm grade N_2=74
Listening	Pearson Correlation	.593**	.088
	Sig. (2-tailed)	.000	.458
Reading	Pearson Correlation	.324*	.168
	Sig. (2-tailed)	.011	.153
Essay writing	Pearson Correlation	.351**	.292
	Sig. (2-tailed)	.006	.012
Grammar score	Pearson Correlation	.305*	.057
	Sig. (2-tailed)	.017	.628
Receptive vocab	Pearson Correlation	.237	-.123
	Sig. (2-tailed)	.066	.295
Beyond 2000	Pearson Correlation	.099	.204
	Sig. (2-tailed)	.449	.081

Requirements students get grades for in the majority of the courses involve a lot of reading and essay writing tasks, together with listening to the lecturer or taking an active part in the discussions in the seminars. Therefore, it is not surprising to find that those who scored higher on the listening, reading and writing subtests of the filter test were more successful in their courses, although the findings on Version 2 do not support this assumption, clearly showing a very different population.

In the interpretation of the results it cannot be ignored that during their studies students in Hungary are evaluated along a five-grade scale (1-5) which may not be sensitive enough to differentiate between high and low achievers for the purposes of the present research. On the other hand, there is no unified grading scheme at UP, which means that every tutor and lecturer gives grades based on their own grading systems, resulting in a great diversity in ways of assessment in courses. The fact that the credit system allows students to take courses of their own choice and that in addition to the compulsory courses in their major they are required to take other courses at other departments to widen their scope of knowledge further enriches the diversity of courses, lecturers and grading schemes the students face in the first year of their studies. It may have been wiser to enter a course grade into the analysis that all participating students take but it was impossible to find such a course in our context.

6.6 Conclusions

This study attempted to explore the relationship between receptive and productive lexical knowledge by analysing the results of 135 first-year students on the proficiency exam compulsory for English majors in the first year of their academic studies. Receptive word knowledge was measured by a thirty-item vocabulary test developed by the researcher based on a list of the most frequent words occurring in academic texts as well as lexis intuitively perceived as useful in English studies. Productive vocabulary use was examined in the essay writing component of the proficiency test and the lexis of the texts was described by three measures of lexical diversity. These vocabulary measures were then compared to scores on the four sub-tests of the exam (reading, listening, grammar and usage, essay writing) as well as to general language proficiency and academic performance. The findings of the study did not reveal any relationship between productive and receptive word knowledge, but significant relationships were shown between word knowledge and scores on the four sub-tests, implying that receptive vocabulary is a significant factor in language proficiency.

The results of the three stages of vocabulary assessment discussed in this chapter imply that it is legitimate and plausible to test academic vocabulary, as these words appear frequently in academic texts and thus need to be specifically focused on and trained. Some argue that less frequent words are not likely to occur in every text and constitute a very low percentage of words in any text, therefore, do not need specialized training. Even more so, as reading requires a lot more than mere vocabulary knowledge: strategy training on how to deal with unknown lexis might be beneficial in the long-run. However, the analysis of a corpus compiled from representative samples of compulsory readings at the three English departments at the University of Pécs may serve as a basis of developing new vocabulary tests adjusted to the special lexical needs of English majors concerning both academic and low-frequency vocabulary. This attempt is reported in the next chapter.

Chapter 7

What words? Corpus-based vocabulary testing at UP

7.1 Introduction

Having estimated the vocabulary size of English majors at UP and explored how well they know the words previous studies reported to be essential for pursuing successful academic studies, the last chapter of this thesis reports an endeavor into corpus linguistics and an attempt of incorporating the findings of corpus analysis in test design. As has been shown in the previous chapters, neither the well established vocabulary tests adapted nor the test of lexis intuitively perceived as useful in English studies seemed to be fully adequate for our assessment purposes. Therefore, in the fourth stage of test development, the lexis of a corpus of compulsory readings is analyzed in order to provide a basis for designing a corpus-based test battery specifically adjusted to the lexical needs of English majors. The first sections give the reader an insight into the compilation of the corpus, followed by a description and comparison of the lexis found in the three sub-corpora. The findings are then applied in practical test design, as reported in the subsequent sections introducing the corpus-based test battery and the results of its administration to 134 first-year English majors.

7.2 Research questions

The following research questions are hence explored in this chapter:

1. What lexis do English majors at UP need to be familiar with to be likely to comprehend the required compulsory readings?

2. Based on evidence from corpus analysis, to what extent may the knowledge of the words included in the Academic Word List (Coxhead, 2000) be beneficial for students in text comprehension?

3. What low-frequency words are students likely to encounter in their compulsory readings?

4. How familiar are the students with the most frequent academic and low-frequency words occurring in the corpus of compulsory readings?

7.3 CORES: A Corpus of Readings in English Studies

Corpus linguistics has now become a widely researched and established field of applied linguistics and its achievements have largely contributed to the success of research on language description (Andor, 2003; Csomay, 2005a, 2005b, 2006, 2007; Hyland, 2007; Mackenzie, 2003; Martsa, 2007), language pedagogy for setting targets (Doró, 2008; Henry & Roseberry, 2007, Horváth, 2001; Tribble, 2002), language assessment for establishing basic and advanced levels (Laufer, 1992b; Moris & Cobb, 2003; Quian, 2002), or even translation studies to establish the characteristics of translated and original texts in comparable or parallel corpora (Baker, 1995, Laviosa, 1998). Academic English (EAP) has been widely studied; one specific application of a 3.5 million word corpus of academic texts is Averil Coxhead's (2000) Academic Word List compiled to aid the learning of core items in written academic vocabulary. Based on this list, Schmitt and Schmitt (2005, cited in O'Keefee, McCarthy & Carter, 2007, p. 216) designed each chapter of their book using words taken from the AWL for readers to self-test their vocabulary knowledge. Tim Johns' website, the *Kibbitzer* (www.eisu.bham.ac.uk. johnstf/timeap3.htm) and Horváth's JPU corpus (2001), on the other hand, are authentic examples of how concordancing can be applied in data-driven language learning in an EAP context.

7.3.1 Compiling the corpus

In order to compile a corpus specific to our English studies programme at UP that can later similarly serve as a basis for data-driven learning and assessment, therefore, each and every colleague at the three departments (English Linguistics, English Applied Linguistics, English and American Literatures and Cultures) was requested to provide a representative sample of authentic texts given as compulsory readings in their courses in the actual semester. According to the design rationale of the corpus the authentic texts had to be representative of the courses in the sense that, based on the intuitions of the tutors, all students attending the given course can rightfully be expected to be able to cope with the selected text language-, content- and vocabulary-wise in order to pass the course. In other words, those students who cannot understand the text are likely to be unsuccessful in completing the course.

Colleagues were asked to give a thorough consideration to their choice and provide one such text for each of their courses supplemented with detailed information on its source (name of author, title, date and place of publication) to be stored in a separate database. Ideally, this would have resulted in a much larger corpus than the one finally received. Some input texts came in electronic files of different format that had to be converted into 'Word' and 'plain text' formats suitable for computer analysis, some had to be scanned. All the texts were saved both in separate files and into one of the three sub-corpora: applied linguistics (AL), linguistics (L), and literature and culture (LC). The size of each sub-corpus is, consequently, proportionate to the share each field represents in the curriculum. The resulting joint CORES corpus is monolingual, written, L1, dynamic and developing, as it is open for continuous expansion, and un-annotated, according to Horváth's matrix of corpus typology (2001, p. 44).

As analysed by Cobb's Vocabprofile (2000), the specialized CORES corpus is made up of 118,808 word tokens (Table 41), stripped of proper names and numbers to exclude citation references to authors. It is small compared to large general combined corpora, such as the hundred million word British National Corpus (BNC), the one

million word Brown corpus, the one billion word Cambridge International Corpus (CIC), or the 220 million word Macmillan World English Corpus (for a comprehensive overview of existing corpora see O'Keefee *et al*., 2007).

It can even be considered small among specialized corpora such as the Corpus of English as a Lingua Franca in Academic Settings (ELFA), a 500,000 word compilation of spoken English; the 1.8 million word Michigan Corpus of Academic Spoken English (MICASE); or the 2.8 million words TOEFL 2000 Spoken and Written Academic Language Corpus (T2K SWAL) representing both spoken and written language use at four US universities. However, as O'Keeffee *et al*. (2007, p. 200) emphasize, small specialized corpora can give a unique insight into the linguistic features of specialized domains, in this case EAP vocabulary within the field of English studies at UP.

7.3.2 The vocabulary of the corpus

As has been discussed in the previous chapter introducing a student corpus, lexical variation is the most widely used descriptor of the lexis of texts. Therefore, it is wise to begin our explorations by examining the type-token ratio of CORES. Although it does not indicate high lexical variation (T/t = 0.10), the ratio of content words and the total number of words in the text (lexical density= 0.57) shows that almost 2 out of every 3 lexical words are supposedly different, imposing a considerable lexical challenge on non-native readers.

The lexical profile of the corpus reveals that the first 2,000 most frequent English words (K1+K2) provide its main lexical basis totaling at 77 percent of all tokens in the text (Table 41). It means that knowing the words of the General Service List (West, 1953) often referred to earlier in the previous chapters would enable the students to understand over three quarters of the words in the compulsory readings required at the English Department. In numerous previous studies this figure exceeds 80 percent of all the words in any general text (O'Keefee *et al*., 2007).

Table 41. Descriptive statistics of the frequency of lexis in CORES

CORES	Families	Types	Tokens	Percent
K1 Words (1-1000)	963	2,798	85,387	71.87%
Function words			(50,658)	(42.64%)
Content words			(34,727)	(29.23%)
K2 Words (1001-2000):	733	1,479	6,296	5.30%
1K+2K:	1,696	4,277	91,683	(77.17%)
AWL Words (academic):	545	1,630	8,958	7.54%
Off-List Words:	?	6,941	18,167	15.29%
Overall	2,241+?	12,848	118,808	100%

This is not a sufficient proportion though, as successful text comprehension may take place provided that 95-98 percent of all the words in the text are familiar to the reader (Hirsch & Nation, 1992; Laufer, 2005a; Nation & Waring, 1997), as discussed in Chapter 2. However, adding the 7.54% of academic words in the corpus that is advised to be learnt from the Academic Word List (Coxhead, 2000) by Nation (2006), the figure amounts close to 85 percent of all the words in the corpus. As has been mentioned earlier in the previous chapters, cognate effects may contribute to better word recognition and inferencing in reading comprehension. Although the L1 of the students is Hungarian, a Finno-Ugric language, which does not help much in understanding English (Jones, 1995), some students may gain a slight advantage in this lexical plight by speaking other languages of Germanic or Latin origin: 68.13 percent of the on-list words in the corpus have Anglo-Saxon origins, while the remaining 31.87 percent is a Greco-Latin or French cognate.

But even this leaves the learners with approximately 15 percent, still a high proportion of low-frequency words supposedly specific to this discipline to be known to some degree in order to comprehend the texts they are required to read at UP. These words may either be acquired through the extensive reading of academic texts in this discipline or learnt consciously by intentionally focusing on the vocabulary items frequent in this corpus: the major questions in the ongoing debate on how vocabulary is most efficiently acquired, as already outlined in Chapter 2. Krashen and his colleagues (Krashen, 1989, 2004), however, maintain that extensive free reading "can

and does provide acquirers with sufficient resources to reach a high level of literacy development" (McQuillian & Krashen, 2008, p. 104) both in L1 and L2 acquisition and support their views by reading rate research (Fraser, 2007).

Advocates of conscious learning argue that uncommon words (off-list words in the CORES corpus, for instance) do not appear frequently enough in free reading only to reach the exposure frequency required for vocabulary acquisition to take place (Cobb, 2007, 2008; Laufer, 1992b, 1997a, 2005b; Nation, 2001; Nation & Waring, 1997). They claim that "the majority of learners in the normal time frame of instructed L2 learning" (Cobb, 2008, p. 109) are unable to build an adequate functional lexicon from free reading only and believe that the explicit teaching of target vocabulary can bridge the gap. Without taking a stance for the time being, let us presume that some further analysis of the CORES corpus might help reveal what type of infrequent lexis our students need to be familiar with and this analysis of lexical needs may give the rationale for suggestions on the future applications of the corpus in the enhancement of student development and the design of the corpus-based vocabulary test battery.

Table 42. Types appearing 1,000 times or more in the CORES corpus

TYPE	RANGE	FREQ	AL	L	LC
THE	3	8,526	2,000	3,424	3,102
OF	3	5,261	1,045	2,201	2,015
TO	3	3,526	934	1,460	1,132
AND	3	3,405	794	1,215	1,396
IN	3	3,400	768	1,596	1,036
A	3	3,341	858	1,480	1,003
THAT	3	1,864	553	813	498
IS	3	1,681	506	764	411
AS	3	1,336	240	623	473
FOR	3	1,099	254	517	328
IT	3	1,007	295	361	351

The most frequent words in the CORES corpus are all function words, as shown in Table 42, although their rank orders slightly differ in the three sub-corpora. Similar rank orders were found in large written and spoken corpora such as the Cambridge International Corpus (CIC), the Cambridge and Nottingham Corpus of Discourse in

English (CANCODE) (cited in O'Keefee *et al.,* 2007, pp. 34-35), or Horváth's JPU student corpus (Horváth, 2001, p. 112), as indicated in Table 43 below.

Table 43. Rank orders and frequencies of the top 10 words in three corpora and the CORES corpus (based on O'Keefee *et al*., 2007, pp. 34-2-35; and Horváth, 2001, pp. 112-115)

CORES	CIC	CANCODE	JPU
the (8,526)	the (439,723)	the (169,335)	the (32,231)
of (5,261)	and (256,879)	I (150,989)	of (14,754)
to (3,526)	to (230,431)	and (141,206)	to (11602)
and (3,405)	a (210,178)	you (137,522)	and (10,835)
in (3,400)	of (194,659)	it (106,249)	in (9,102)
a (3,341)	I (192,961)	to (105,854)	a (8,526)
that (1,864)	you (164,021)	a (103,524)	is (6,409)
is (1,681)	it (150,707)	yeah (91,481)	it (4,149)
as (1,336)	in (142,812)	that (84,930)	that (4,123)
for (1,099)	that (124,250)	of (78,207)	I (3,695)

More informative about the lexis of a corpus is, however, a profile of lexical words in the texts. The first such word, *language*, comes only twentieth in the list. On closer examination, the distribution of the most common content words in the CORES corpus reveals that only K1 and K2 words appear over a hundred times in the corpus, as presented in Tables 44 and 45. Not surprisingly, the most frequent content word in linguistics and applied linguistics texts proved to be *language*, while it had considerably fewer occurrences in the LC sub-corpus, where interestingly the words *new* and *people* appeared the most often. Concerning the academic words (Table 46) *motivation* came first in the AL sub-corpus, with frequent occurrences of *attitude, individual, research* and *process*. In the LC sub-corpus *culture* and its derived forms were most frequent, together with the words *civil, immigrate, individual, complex* and *text*, while the linguistics (L) sub-corpus featured the words *tense, adult, input, principle* and *structure*. As the corpus contains 545 academic word families out of the 570 word families included in the AWL (Coxhead, 2000), it is legitimate to assume that the knowledge of these words may contribute to a better understanding of the compulsory EAP readings at UP.

Table 44. K1 content word types appearing over 100 times in CORES

TYPE	RANGE	FREQ	AL	L	LC
LANGUAGE	3	592	244	324	24
SO	3	344	62	199	83
CHILDREN	3	281	9	246	26
WORDS	3	259	25	219	15
NEW	3	245	47	62	136
CHILD	2	237	0	231	6
LEARNING	3	223	125	86	12
WORD	3	206	16	175	15
USE	3	199	65	112	22
ENGLISH	3	194	26	147	21
ONLY	3	191	38	90	63
EVEN	3	183	43	71	69
PEOPLE	3	180	68	26	86
FORMS	3	170	6	147	17
SEE	3	165	40	107	18
ALSO	3	138	41	58	39
WAY	3	138	40	62	36
AGE	3	133	6	114	13
MAN	3	131	4	93	34
FORM	3	128	17	89	22
PAST	3	128	15	61	52
LANGUAGES	3	125	13	111	1
TIME	3	125	48	43	34
VERY	3	119	56	37	26
EXAMPLE	3	115	19	85	11
JUST	3	113	46	50	17
SAME	3	111	31	52	28
SPEECH	3	109	17	85	7
CHANGE	3	108	33	63	12
RULE	3	107	22	80	5
SAY	3	107	29	53	25
OWN	3	106	25	26	55
THEN	3	102	25	44	33
WELL	3	101	29	42	30

Table 45. K2 content word types appearing over 100 times in CORES

TYPE	RANGE	FREQ	AL	L	LC
SENTENCE	3	139	39	97	3
VERBS	1	127	0	127	0
REGULAR	2	120	2	118	0
VERB	2	116	2	114	0
IRREGULAR	1	108	0	108	0
SENTENCES	3	107	24	81	2

Table 46. AWL content word types appearing over 40 times in CORES

TYPE	RANGE	FREQ	AL	L	LC
MOTIVATION	3	91	89	1	1
ACQUISITION	2	75	22	53	0
CULTURAL	3	73	8	2	63
TENSE	2	73	9	64	0
INDIVIDUAL	3	72	38	16	18
PROCESS	3	70	28	30	12
CULTURE	3	68	6	2	60
SIMILAR	3	62	13	40	9
PRINCIPLE	3	61	7	45	9
ADULT	2	60	1	59	0
INPUT	1	57	0	57	0
ATTITUDES	3	56	50	1	5
EVIDENCE	3	56	8	44	4
STRUCTURE	3	55	2	48	5
COMPLEX	3	54	12	27	15
TEXT	3	54	7	30	17
THEORY	3	52	10	38	4
CONTEXT	3	50	15	28	7
RESEARCH	3	47	36	9	2
ROLE	3	45	14	16	15
SPECIFIC	3	44	5	31	8
CHAPTER	2	42	5	37	0
CIVIL	2	41	1	0	40

It may be even more interesting to examine what low-frequency words constitute over twelve percent of the lexis in the corpus. As Cobb (2008, p. 109) notes, research has shown that a word needs to appear six times minimally for incidental learning to take place; therefore, the list of the low-frequency word types appearing six times or more in the CORES corpus is shown in Appendix K. Even the most frequent word (*semantic*) in this list appears only 93 times in the corpus, and the list involves a large number of technical terms, for instance *bilingual, pragmatic, phonological, suffix*, or *connectionist*, that need to be studied in specific courses in the curriculum. Others, such as *appeal, convey, prominent* or *entitled*, are less specialized in nature and may be useful items in the general mental lexicon of any advanced learner of English. Assuming that Cobb (2008) and other researchers are right in suggesting that six encounters with a word in reading may result in some degree of word learning, these low-frequency items in the list may have the greatest chance to be acquired incidentally in the course of English studies at UP, therefore, may be considered for vocabulary assessment purposes. However, the majority of the off-list words in the corpus occur only once in one of the three sub-corpora, raising the much debated question whether these items should be consciously dealt with or left for incidental vocabulary acquisition.

It has to be noted here, however, that although the CORES corpus is representative of the compulsory readings required in English studies at UP, it is also a cross-sectional collection of reading materials; consequently, first-year students may have encountered only a tiny fraction of these readings in their first two semesters, which greatly reduces their chances for a sufficient number of low-frequency word exposures at the time they are required to take the proficiency test. Another limitation of the present state of the CORES corpus that has to be mentioned here is that it involves only one article per course, while according to the syllabuses, students are required to read four articles on average to pass each course they take.

Furthermore, even though literature courses do not concern first-year students involved in the present study, the reading load greatly increases from the third

semester with the authentic British and American novels, short stories and poems, which are excluded from the CORES corpus but may provide a rich and vast additional source of vocabulary acquisition from reading. Although the lexical analysis of such an immense corpus is beyond the aims and available resources of this study, its contribution to the vocabulary development of the students can by no means be ignored. Yet, even with these limitations, the possible applications of the corpus and the implications of the above reported lexical analysis may be multifold and are going to be discussed in the following sections of this chapter.

7.4 Corpus-based vocabulary testing at UP

One such area where the results of the above corpus analysis may be applied is vocabulary assessment, more specifically vocabulary test design. The rationale for compiling the CORES corpus was to provide a basis for corpus-based vocabulary test development in order to diagnose and filter out students lacking in receptive vocabulary at the end of the second semester in English studies at UP. It was hypothesized at the outset of the investigation that insufficient vocabulary size hinders reading and listening comprehension and consequently results in poor academic achievement, as implied by past research overviewed in Chapter 2. In the next stage of the research we explored how familiar English majors at UP were with academic vocabulary supposedly essential for pursuing university studies and how knowledge of low-frequency vocabulary differentiated among proficient and less prepared students. It was concluded that although the 10K word level of the VLT adequately separated high and low achievers for our filtering purposes, a corpus of readings students were expected to cope with in our specific context would better indicate the special lexical needs of our learners, a factor that modern test design needs to consider. This method excludes intuitive judgements of word frequency often equaled with the usefulness of a word which was the main weakness of the study reported as stage 3 in the previous chapter.

7.4.1 Instruments

For reasons discussed earlier in detail in Chapter 6, the new corpus-based battery was again to be inserted in the grammar and usage component of the proficiency test; therefore, its structure and item number were kept identical to the ones administered and trialed in the previous semesters. Based on the lexical profile of the CORES corpus two 30-item versions were produced (Appendix H). Building on the lessons learnt from the previous stages, the main principle guiding the selection of lexical items to be included in the corpus-based battery was to test both academic vocabulary and low-frequency vocabulary occurring frequently enough in the corpus to allow for the supposedly sufficient number of encounters necessary for learning to take place.

Therefore, all the 30-30 items and the 30-30 distractors were selected from amongst the AWL and low-frequency words occurring in the corpus on both versions. Target words within the clusters were listed in alphabetical order as recommended by Schmitt *et al.* (2001). Target words to be defined were randomly selected, and as much as possible, the target words in the clusters begin with different letters and have different orthographic forms. As analysed by VocabProfile (Cobb, 2000), in Version 1 there were 17 AWL words (28.33%) and 43 low-frequency words (71.67%) among the items. Supposedly easier academic words were placed at the beginning of the battery: the first three clusters (18 words) comprised academic words only except for two low-frequency words: *debilitating* and *gloomy* in cluster three.

The subsequent six clusters involved low-frequency words except for the AWL word *initiate*. Concerning this ratio among test- and distractor items, nine AWL and 21 low-frequency words were tested, i.e. provided with a definition or synonym to match. Test- and distractor items within a cluster were intended to represent the same frequency level. In a similar fashion, in Version 2 there were 17 AWL (28.33%) and 43 low-frequency word items (71.67%), all AWL words were placed in the first three clusters supplemented with one low-frequency distractor item: *flawless*. The lexical profiles of the two test versions are shown in Table 47.

Table 47. The lexical profile of the CORES-based test

	Version 1	Version 2
AWL sublist 1	concept, estimate	approach, consistent, indicate
AWL sublist 2	appropriate, impact	obtain, primary
AWL sublist 3	imply	demonstrate, locate
AWL sublist 4	retain	subsequent
AWL sublist 5	fundamental, objective, pursuit	exposure, sustain, target
AWL sublist 6	edition, initiate, subsidy	incentive, utility
AWL sublist 7	comprise, successive	voluntary
AWL sublist 8	contemporary, terminate	arbitrary
AWL sublist 9	devote	controversy, violate
Off-list (low-frequency words)	advent, adversary, alleged, allot, amplify, apprentice, assault, breakthrough, chivalry, civic, conceal, congenial, contemplate, conviction, debilitating, denounce, descendant, designate, despised, doctrine, eloquence, elusive, endow, era, evoke, feasible, fragment, glimpse, gloomy, haunted, impending, lucrative, menace, merchandise, mushroom, obsession, pace, portable, postulate, preface, proliferation, redeem, subtle	adherence, allude, appeal, apprehension, articulated, assertion, candidate, clergy, concede, condemn, conspicuous, depict, disrupt, drawback, flawless, gratitude, heritage, indispensable, mandatory, preface, resemblance, retreat, revive, salient, salvation, scholar, scramble, specimen, stunning, surmount, tactful, testimony, tormented, trait, tranquility, uncanny, unfold, upheaval, vernacular, wane, whimsical, wit, yearn

Table 48. Frequency of words in the definitions in the corpus-based vocabulary test Version 1.

1K types: [families 47 : types 47 : tokens 58] a_[2] and_[3] arrival_[1] as_[1] at_[1] attack_[1] begin_[1] bring_[1] broken_[1] by_[1] carry_[1] catch_[1] cause_[1] coming_[1] danger_[1] dark_[1] difficult_[1] easy_[1] effect_[1] fast_[1] finish_[1] following_[1] for_[1] good_[1] grow_[1] idea_[1] in_[1] include_[1] look_[1] make_[1] mind_[1] necessary_[1] of_[1] off_[1] or_[4] order_[1] part_[1] producing_[1] profit_[1] sale_[1] spread_[1] stronger_[1] suggest_[1] time_[1] to_[5] true_[2] visited_[1]

2K types: [5:5:5] aim_[1] hide_[1] nice_[1] suitable_[1] violent_[1]

AWL types: [6:6:6] achieve_[1] assume_[1] brief_[1] notion_[1] period_[1] principle_[1]

OFF types: [?:6:6] ghosts_[1] goods_[1] mania_[1] pessimistic_[1] questionably_[1] sociable_[1]

Table 49. Frequency of words in the definitions in the corpus-based vocabulary test Version 2

1K types: [families 43 : types 43 : tokens 48] a_[1] after_[1] back_[1] beautiful_[1] become_[1] better_[1] break_[1] bring_[1] coming_[1] considering_[1] desire_[1] disadvantage_[1] expected_[1] feelings_[1] find_[1] for_[1] from_[1] full_[1] important_[1] indirectly_[1] introduction_[1] keep_[1] life_[1] mind_[1] most_[1] name_[1] of_[2] open_[1] or_[3] others_[1] place_[1] property_[1] reach_[1] rule_[1] something_[1] strongly_[1] than_[1] the_[1] to_[3] together_[1] type_[1] unchanging_[1] up_[1]

2K types: [11:11:11] aim_[1] calmness_[1] clever_[1] collective_[1] essential_[1] gradually_[1] mix_[1] mysterious_[1] priests_[1] quarrel_[1] refer_[1]

AWL types: [4:4:4] attachment_[1] feature_[1] image_[1] motivation_[1]

OFF types: [?:6:7] ancestors_[1] compulsory_[1] faithfulness_[1] humour_[2] playful_[1] supernatural_[1]

In line with the recommendations of Schmitt *et al.* (2001), the definitions were kept short and simple in order to minimize the amount of reading necessary and to allow for more items to be taken within the time available. The definitions were to be more frequent than the target words and wherever possible, the target words were defined with words from the GSL.

7.4.2 Participants and procedures

The test was administered to 134 first-year English majors at UP. All participants were native speakers of Hungarian. Version 1 and Version 2 were taken by 86 and 48 candidates, respectively. The two test versions were administered on 21 May (Version 1) and 29 May (Version 2) 2008, between 9 and 12.30. Both exams followed the same procedure as with the previous tests reported in Chapter 6. However, the sequence of the proficiency test components was this time slightly modified upon the students' request as follows: 1: listening comprehension test (approx. 30 mins); 2: writing test (60 mins) followed by a short break of 10 minutes, 3: reading comprehension test (40 mins); 4: grammar and usage test (40 mins). Answers on the listening and reading comprehension, the grammar and the receptive vocabulary tests were typed into a computer and analyzed by ITEMAN version 3.5 (1993).

7.4.3 Results and discussion

Descriptive statistics on the corpus-based receptive vocabulary tests reveal that the means on both test versions (18.31 and 16.83 on Version 1 and Version 2, respectively, standard deviations 4.69 and 3.92) exceed the fifty percent of the maximum score (30), as indicated in Table 50. The range of scores is wider on Version 1 (5-29) than on Version 2 (10-26), as shown by variance values. As in this stage of the test development process the focus of investigation is on the receptive vocabulary test battery, the scores on the other four subtests of the proficiency test are not reported here in detail. However, like in all the previous stages of assessing receptive vocabulary, correlations among the three subtests of the grammar and usage component were again examined and significant correlations were found between receptive vocabulary and the two grammar test components (Version 1: r_{13}= .581, p< 0.01; r_{23}= .580, p< 0.01; Version 2: r_{13}= .503, p< 0.001; r_{23}= .457, p< 0.01) providing further support to the assumption that grammar and vocabulary are related constructs.

Table 50. Descriptive statistics on the corpus-based receptive vocabulary tests

Corpus-based receptive vocabulary test	*Version 1*	*Version 2*
N. of Items	30	30
N. of Examinees	86	48
Mean	18.314	16.833
Variance	21.983	15.347
Std. Dev	4.689	3.918
Skew	-0.012	0.179
Kurtosis	0.056	-0.624
Minimum score	5	10
Maximum score	29	26
Median	18	17
Cronbach's	0.779	0.665

The graphical illustration of the above data provides further insights into the findings (Figure 17). The skewness of a dataset, a measure of asymmetry in the probability distribution, is in the negative on Version 1 of the corpus-based vocabulary test implying a longer left tail of the distribution curve with relatively few low values.

Whereas on Version 2, score distribution is a little shifted towards the left of the figure, indicating relatively less high values than on Version 1. Kurtosis, the measure of the peakedness of the distribution, also differs on the two test occasions. The negative kurtosis (-0.624) on Version 2, as opposed to the positive kurtosis (0.056) on Version 1, indicates that the variance in the distribution on Version 2 is rather due to frequent modestly sized deviations. On Version 1, on the other hand, some extreme deviations from the mean, such as the outlier student scoring five, also play a role.
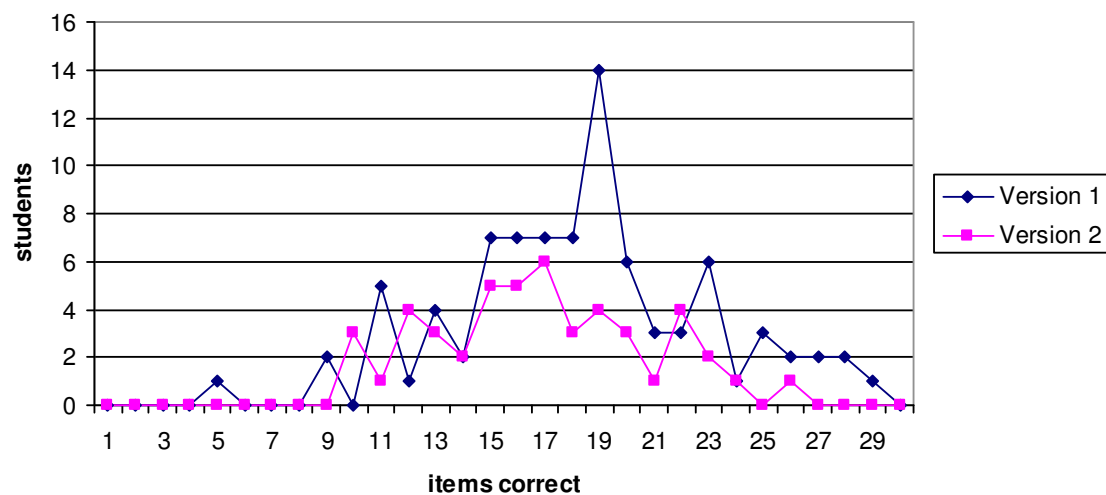
Figure 17. Score distribution on the corpus-based vocabulary test

Similarly to the procedure applied in the case of the receptive vocabulary test discussed in the previous chapter, in order to find out about how the items of the corpus-based vocabulary test worked, facility values and discrimination indices need to be examined more closely. The mean facility values on Version 1 and Version 2 are 61.07 (sd= 24.27) and 56.13 (sd= 25.62), respectively. Mean discrimination indices were 0.33 (sd= 0.14) on Version 1 and 0.31 (sd= 0.18) on Version 2 suggesting that the majority of the items require little or no revision (Szabó, 2008, p. 32). Two-tailed independent samples T-tests were run to reveal the difference between the two test versions and no significant difference was found either in the facility values or the discrimination indices of the two test versions (t_{fv}= .45, t_{DI}= .59, p> .05).

Table 51. Item statistics on the corpus-based vocabulary test

	AWL items					Off-list items			
	N	Mean %correct	sd	Mean DI	sd	N	Mean %correct	sd	Mean DI	sd
Version 1	9	59,89	28,26	0,31	0,17	21	61,57	23,10	0,34	0,12
Version 2	9	65,89	22,44	0,38	0,17	21	51,95	26,26	0,28	0,18

Besides comparing overall scores and figures on the two tests, a comparison of the behaviour of the academic and low-frequency items provides further insights into the nature of the corpus-based receptive vocabulary test. Table 51 indicates means on the academic (AWL) and low-frequency (Off-list) items separately. On the second test, the mean facility value on the academic word items (65.89) tends to be higher than on the low-frequency word items (51.95), whereas on Version 1, testees scored higher on the off-list items (61.57) and the academic words proved to be somewhat more difficult (59.89). Interestingly, however, the off-list items discriminated slightly better between high and low achievers on Version 1, whereas it seems to be the other way around on Version 2, where the academic words produced higher discrimination indices (0.38). Repeated two tailed independent samples T-tests, however, showed that neither the academic nor the off-list items were significantly different in the two versions (% correct: t_{AWL}= .62; t_{Off} = .24; discrimination index: t_{AWL} = .37; t_{Off} = .19; p>0.05).

Table 52. Corpus-based item characteristics

Version 1	*List*	*% correct*	*DI*	*Version 2*	*List*	*% correct*	*DI*
1. impact	AWL	97	0.10	1. target	AWL	69	0.41
2. objective	AWL	49	0.34	2. incentive	AWL	27	0.55
3. concept	AWL	95	-0.03	3. controversy	AWL	52	0.52
4. terminate	AWL	66	0.41	4. locate	AWL	100	0.00
5. imply	AWL	26	0.38	5. sustain	AWL	46	0.46
6. comprise	AWL	23	0.38	6. violate	AWL	65	0.50
7. successive	AWL	35	0.55	7. primary	AWL	90	0.31
8. appropriate	AWL	81	0.38	8. subsequent	AWL	79	0.31
9. gloomy	OFF	84	0.24	9. consistent	AWL	65	0.40
10. era	OFF	88	0.31	10. drawback	OFF	79	0.34
11. advent	OFF	55	0.38	11. tranquility	OFF	58	0.21
12. menace	OFF	55	0.55	12. adherence	OFF	31	0.35
13. assault	OFF	76	0.34	13. unfold	OFF	40	0.21
14. fragment	OFF	87	0.28	14. allude	OFF	33	0.49
15. glimpse	OFF	69	0.59	15. scramble	OFF	85	0.09
16. conceal	OFF	63	0.34	16. tactful	OFF	46	0.12
17. initiate	AWL	67	0.28	17. uncanny	OFF	8	0.06
18. evoke	OFF	55	0.52	18. whimsical	OFF	48	0.25
19. merchandise	OFF	80	0.21	19. heritage	OFF	94	0.08
20. obsession	OFF	84	0.28	20. wit	OFF	75	0.35
21. doctrine	OFF	81	0.41	21. clergy	OFF	77	0.62
22. lucrative	OFF	35	0.55	22. surmount	OFF	40	0.13
23. congenial	OFF	15	0.24	23. revive	OFF	94	-0.06
24. elusive	OFF	47	0.34	24. yearn	OFF	48	0.30
25. postulate	OFF	24	0.38	25. indispensable	OFF	35	0.49
26. mushroom	OFF	34	0.31	26. stunning	OFF	60	0.43
27. amplify	OFF	42	0.24	27. mandatory	OFF	35	0.53
28. portable	OFF	88	0.14	28. preface	OFF	77	0.46
29. haunted	OFF	86	0.28	29. trait	OFF	13	0.17
30. alleged	OFF	45	0.28	30. apprehension	OFF	15	0.25

Table 52 presents facility values and discrimination indices for each item on both test versions. There was one item familiar to all students on Version 2 (*locate*). Eleven items were known to over 80 % of the test-takers on Version 1 (AWL: *impact, concept, appropriate*; Off-list: *gloomy, era, fragment, merchandise, obsession, doctrine, portable, haunted*), surprisingly the three quarters of which came from among the low-frequency items. On Version 2 there were only five such items; two academic words (*locate, primary*) and three Off-list words (*scramble, heritage, revive*). The items that proved to be most difficult to the majority of the testees were *comprise, congenial,* and *postulate* on Version 1, and *incentive, trait* and *apprehension* on Version 2, known to fewer than 30 percent of the participants, as shown graphically in Figure 18.
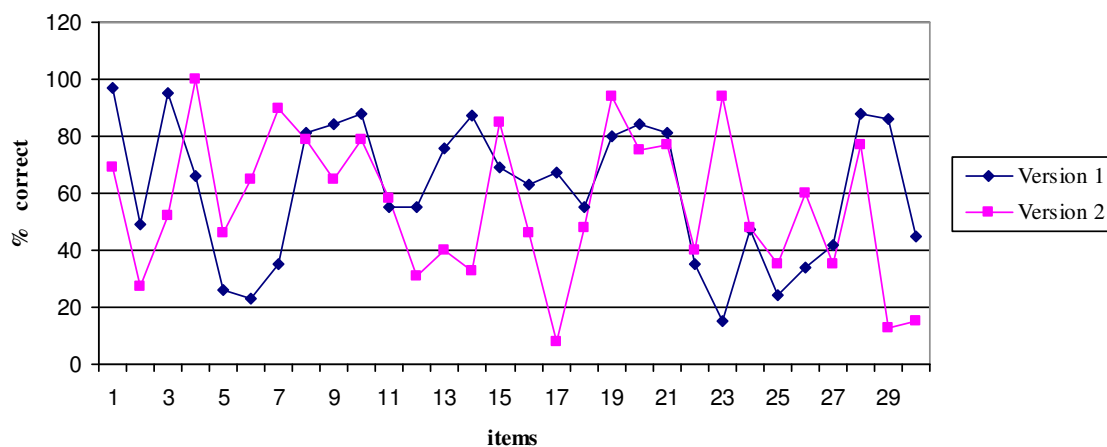
Figure 18. Facility values on the corpus-based vocabulary tests

Discrimination indices range from –0.03 *(concept)* to 0.59 *(glimpse)* on Version 1 and from -0.06 *(revive)* to 0.62 *(clergy)* on Version 2 (Table 52). Negative discrimination indices imply that more low-scorers than high-scorers get the item right, therefore, the item requires revision or need to be completely eliminated in possible future applications of the tests. According to the guidelines provided by Szabó (2008, p. 32), two items (*impact, portable*) on Version 1 and seven items (*locate, scramble, tactful, uncanny, heritage, surmount, trait)* on Version 2 should be completely revised (Figure 19). Further nine items on Version 1 and three items on Version 2 seem to be marginal

(.20 DI .29) and need revision. However, the remaining 18 items (DI .30) on Version 1 and 19 items on Version 2 require little or no revision, out of which six and twelve on Version 1 and Version 2, respectively, can be said to function quite satisfactorily (DI .40).
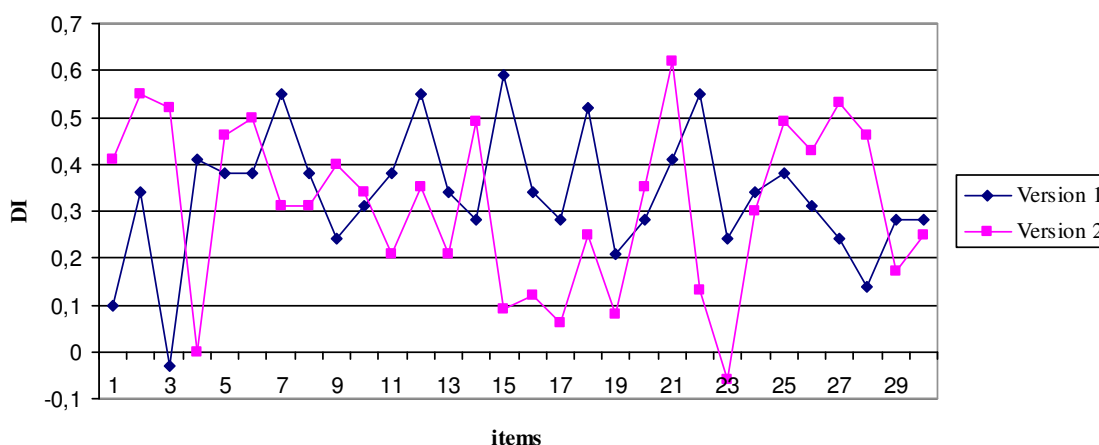
Figure 19. Discrimination indices on the corpus-based vocabulary test

However, as Szabó (2008, p. 33) emphasizes, a limitation of such traditional statistics on item difficulty and discriminability is that it "cannot be considered objective", as both the facility value and the discrimination index are sample dependent measures. As the facility value is expressed as the percentage of the correct responses on the item in the particular test-taker population, it cannot remain constant from administration to administration or from population to population; a need expressed by numerous researchers (e.g., Lord & Novick, 1968). Similarly, as demonstrated by Szabó (2008), the discrimination index is influenced by the facility value, and "because it is calculated on the basis of the performance of a particular population" (Szabó, 2008, p. 33), it is also relativistic in nature.

Therefore, it is important to treat the above discussed item statistics of the corpus-based receptive vocabulary tests with caution, as different results may be obtained with higher or lower ability groups of participants on any future administrations of the tests. This relativity of figures may also serve as a possible explanation to the

196

observation that lower facility values were obtained on Version 2 of each test administration, as probably the less able students were more reluctant to sign up for the first of the two optional test occasions every year.

The reliability of a test is also influenced by the number of testees, therefore, another reason for concern is the low number of candidates involved in each stage of the study. As has been discussed in the previous chapters, the Filter/Proficiency test is administered only twice a year and the number of test-takers vary from year to year and from test version to test version (Table 53). It can rightfully be assumed that a higher number of participants may have yielded better reliability figures on every test version. For instance, in the first two stages of vocabulary assessment two validated and reliable tests were adopted and administered (Schmitt *et al.,* 2001). Comparing the reliability figures to those of the relevant Schmitt *et al*. (2001) test levels it is clear that employing 801 participants showed higher reliability of the same instrument than working with 66-154 participants in our case. On the Academic section of the Levels Test (Schmitt *et al*., 2001) adopted as Stage 1 in our test development process (Table 53), Schmitt and his colleagues (2001) reported reliabilities of 0.958 and 0.960 for Version 1 and Version 2, respectively. Similarly high reliability of the instrument was shown on the 10K word level section adopted as Stage 2 in this study (0.915 and 0.924, respectively). As the tests applied were the same, it is legitimate to suppose that the difference is rather due to the difference in the number of candidates than the unchanged characteristics of the tests.

Table 53. Reliability data in the test development process

| | Stage 1 | | Stage 2 | | Stage 3 | | Stage 4 | |
	V1	V2	V1	V2	V1	V2	V1	V2
items	30	30	30	30	30	30	30	30
N	154	66	94	117	61	74	86	48
Cronbach's	0.78	0.55	0.59	0.57	0.83	0.75	0.78	0.67

In the light of the above line of argument, the reliability figures obtained in Stage 3 and 4 of the test development process are not unacceptably low (Table 53). In fact, the figures are higher than the reliability figures in the previous two stages, in spite of the decrease in the number of candidates (except for Stage 3 Version 2).

However, reliability is not only dependent on the number of candidates involved in measurement, but also on the number of items applied. In his monograph on applying Item Response Theory in item banking Szabó (2008, p. 93) reported that increasing the number of items from 50 to 70 with the inclusion of the vocabulary subtest in the Grammar and Usage component of the Proficiency test in 2002 increased the reliability estimates of the GAU component. Further increasing the number of items measuring receptive vocabulary size seems unnecessary, however, for two reasons. First, it would shift the focus of the GAU test component on vocabulary instead of the present balanced ratio of grammar and vocabulary items. Second, Schmitt *et al.* (2001) have shown in their Levels Test that 30 items per level produce good reliability and recommend that unless there is a particular need to apply more items, this is a sufficient number to measure vocabulary size in this test format.

7.5 Conclusion and future directions

This chapter has discussed the compilation, analysis and application of a corpus of compulsory readings in English studies at UP (CORES). The rationale for the compilation of the corpus was the inadequacy of the currently available tests of lexis for the purposes of assessing English majors's vocabulary knowledge in Hungary. As concluded from the studies reported in the previous chapters, besides the obvious need for testing the receptive knowledge of academic words claimed to be essential by part research in pursuing successful university studies, a necessity of focusing on low-frequency words that are key to text comprehension and are specific to this particular discipline also surfaced.

A remedy to the problem seemed to be the lexical analysis of a collection of texts students are required to read, in order to find out more about the lexical needs of this special group of language learners. The analysis revealed that the kowledge of the most common English words, together with a receptive knowledge of the words most common in academic texts, does contribute to successful text comprehension to a great extent. However, English majors are still left with a devastating twelve percent of all the words in the corpus that constitute rare words they are not likely to encounter frequently enough for incidental learning from reading to take place. It implies that these words require special attention on the part of the tutors and the students alike.

One possible application of the information gained from the corpus analyis undertaken has been reported in the rest of the chapter. Two versions of a test battery intended to measure the students's knowledge of the academic and low-frequency words most common in the CORES corpus were devised and tested. The results are promising, however, the initial data revealed that the battery needs further revision before entering a validation process that would make it adequate for wider application at other institutions of higher education in Hungary. Other encouraging future applications of the CORES corpus may have implications for in-class vocabulary teaching, may contribute to fostering autonomous learning and exam preparation and even serve as a basis of research for linguists and corpus linguists. These aspects are elaborated on in the general conclusions.

General conclusions and implications for future research

This dissertation has undertaken to explore the vocabulary knowledge of first-year English majors with the aim of identifying problem areas and offering novel paths for vocabulary assessment in a Hungarian EAP context. The scaffold of the dissertation comprising two parts and seven chapters was provided by the major questions distilled from the vast literature of the field. Whereas the first three chapters in Part I providing an overview of past research centred around three major theoretical issues of what vocabulary is, how it is learnt and used, and how it may be assessed; the empirical chapters in Part II investigated how many words first-year English majors at UP know, how well they know words at certain levels of lexical frequency and what words they need to know based on their curricular requirements.

After introducing the reader to the essential terminology of vocabulary studies, Chapter 1 discussed the place of vocabulary in language competence and overviewed how major models describe the organization as well as the fundamental processes of the mental lexicon. The scope of investigation in Chapter 2 was the acquisition and use of lexis. We examined how various cognitive factors, lexical features and the L1 may affect vocabulary acquisition and clarified the similarities and differences between the 'breadth' and 'depth' approaches to word learning. In discussing incidental acquisition versus intentional learning of vocabulary we touched upon the role of lexical coverage in texts, the learner's vocabulary size, learner strategy use, the frequency of exposure to lexical items and the modification of the input in acquiring lexis. Finally, Chapter 3 overviewed both qualitative and quantitative measures of word knowledge and introduced the vocabulary tests relevant to the empirical studies in Part II.

In Part II, Chapter 4 described the context of the studies. The three small-scale pilot studies reported in Chapter 5 tapped into different aspects of word knowledge and the findings set the rationale for and served as starting points of further investigations into the field. The results of the innovative first pilot study measuring the receptive

vocabulary size of first-year English majors in Hungary provided evidence for the hypothesized gap between the minimal receptive vocabulary size necessary for pursuing successful academic studies as defined by past research and the assessed receptive vocabulary size of our English majors. This inspired the researcher to embark on a deeper examination of what words English majors need to be familiar with in order to be successful in their studies, as reported in Chapters 6 and 7. Decisions on methods and procedures to be applied in the investigation, however, had to be preceded by trials of currently available tools and measures in vocabulary research to find out more about the relationship between receptive and productive vocabulary and the efficiency of intentional and incidental vocabulary learning. These orientating trials discussed in the other two pilot studies lead the researcher to the design of the series of empirical studies aimed at introducing and developing a separate vocabulary test component in the milestone proficiency examination prescribed for first-year English majors in their curriculum at UP.

The development of the vocabulary test battery had four major stages. In the first stage, critical initial decisions had to be made on the aims, format and place of the battery in the proficiency test. It was finally inserted into the grammar and usage component in the form of adapting the academic level items of a widely used and acknowledged vocabulary test. Statistical analysis, however, revealed that this battery was not fully convenient for the purposes of identifying and filtering students with insufficient lexical knowledge. In Stage 2, therefore, it was hypothesized that the knowledge of low-frequency words would better differentiate between candidates with larger and smaller vocabularies. Although the findings supported this hypothesis, the battery adapted nevertheless proved to be inadequate in terms of the irrelevance of the test items to our context.

Stage 3, therefore, was innovative in adhereing to the old test format but devising new items based on the intuitively perceived usefulness of particular words in the discipline of English studies. The investigation at this point reached a little further into the nature of vocabulary knowledge by exploring the relationship between vocabulary and

grammar, reading, listening and writing skills, general language proficiency and academic success. The findings imply that receptive vocabulary and grammar seem to be underlying constructs in the three skills and are closely related. Furthermore, concerning the productive vocabulary measures applied in the study it can be concluded that there is more to writing quality than sophisticated vocabulary use and that the numerical measures of productive vocabulary use may not be sufficient for determining the lexical qualities of a text. The findings also revealed the relatedness of receptive vocabulary and general language proficiency which may have implications for both language pedagogy and assessment.

However, as has been discussed in the first part of the dissertation, intuitive judgements of the frequency or the usefulness of a word may be misleading (McCrostie, 2007). Therefore, Stage 4 of the test development process relied on a corpus compiled by the researcher from readings required in English studies at UP, named CORES. The compilation aimed to serve as a basis for selecting relevant lexical items for our assessment purposes to meet the real lexical needs of English majors. The initial goals defined at the outset of the study seem to have been met with this approach as shown by the statistical analysis of the battery. Yet, we cannot claim to have come to a berth. The battery needs to go through a lengthy process of revision and validation before it can be applied in a wider context.

Besides, the studies reported in the dissertation are not without limitations. First, as has been discussed in the last chapter, the number of participants has an influence on the results of statistical analysis. Consequently, involving more testees may yield more reliable and generalizable results. Second, a caveat to bear in bind about lexical frequency profiling used in the analysis of both the student corpus and the CORES corpus is its incapability of handling lexical phrases or idioms generally considered to be evidences of sophisticated language use (Muncie, 2002). Involving this qualitative aspect of lexical knowledge in the present assessment process was abandoned for the sake of quantitative considerations of assessing the knowledge of a relatively large number of words in a relatively small time and space available in the proficiency

examination, as implied by Laufer *et al.* (2004). Yet, it would be beneficial to conduct qualitative investigations to explore other levels of lexical knowledge and the role of context in both corpora in the future.

Yet another potential problem in the test development procedure is the limited range of information gained from piloting the tests applied in Stages 3 and 4. Piloting tests is considered to be an "indispensable, yet highly problematic stage of test design" (Szabó, 2008, p. 89; see also Alderson, Clapham, & Wall, 1995; Hughes, 1991). Piloting in this study proved to be problematic because ideally the size and characteristics of the pilot group should be as close as possible to those of the target test population. Unfortunately, we cannot claim to have met this condition as the tests were piloted on a few colleagues in order to prevent that the test items become known to the target students before due time. However, based on the information gained from piloting ambiguous or poor definitions or cluster items have been replaced or completely revised before live administrations of the tests. At this point, we are, therefore, in the first stage of the corpus-based test development process, having gathered initial information on how the items in the battery work with this particular population. Consequently, the CORES-based vocabulary test at present cannot be claimed reliable and valid; it needs to undergo a thorough process of validation in the future before it can be recommended for use in other English studies programmes in Hungarian tertiary education.

Finally, another possible factor that needs to be considered in the revision of the corpus-based test and in the design of further similar corpus-based batteries in the future is the key word status of the lexical items to be selected for testing. Keywords, as defined by Scott (1999, cited in O'Keefee *et al.,* 2007, p. 12), are not necessarily the most frequent words, but the most "unusually frequent" words in a text, as compared to some norm. For instance, the software *Wordsmith Tools,* not available to the researcher at the design phase of the test, compares the words of any input text to a large reference or benchmark corpus and highlights any words that occur with an unusual frequency or "keyness" (O'Keefee *et al.,* 2007, p.12) in the input text. This

approach has been criticised by Andor (2009) who argues for the relevance of conceptual networks and frames in defining the keyword status of words in a text. However, producing keyword lists from the CORES corpus supplementing the information that less refined frequency-based item selection can provide may help better identify key specialized vocabulary worth attending to in vocabulary assessment.

Even with these limitations, the results of the studies reported have contributed to the description, the pedagogy and the assessment of English majors' EFL vocabulary in multiple ways and torched directions for future research.

Implications for corpus linguistics

- The student corpus, although relatively small compared to other corpora, may provide insights into the nature of Hungarian students' language use in an EAP context and advance the description of learner EFL writing. The peculiarities of student writings may be further explored in comparison to student writings in other disciplines.
- CORES as a specialized corpus of texts in English studies may contribute to the description of the specificities of language use in this discipline.
- The corpus of compulsory readings needs to be extended further and made available for other researchers working in institutes of Hungarian tertiary education.

Implications for language pedagogy

- The student corpus can be employed effectively in writing pedagogy, as previously shown by Horváth (2001).
- Frequency- and keyword lists produced from the CORES corpus inform language teachers about what words to focus on in preparing students for reading academic texts in their field of studies. It has been suggested that besides a good knowledge of frequent and academic words, being familiar with specific lexis rare in general English texts but frequent in the discipline of

English studies highly increases the potentials of students in academic text comprehension. Therefore, seminars devoted to language development in the first year may benefit from exploiting the CORES corpus in many ways not only at the UP but also in other similar English studies programmes nationwide.

- Autonomous learning and exam preparation can be facilitated by further future applications of CORES. Frequency- and keyword lists produced from the corpus and made available for students on the departmental website may serve as guidelines for focused intentional vocabulary learning outside the classroom.

- Tools of corpus analysis such as concordancers run on the corpus may help students identify and remedy individual problem areas at all possible levels of language use. The corpus may be exploited and consulted for reference in the process of essay writing for instance by checking out typical collocations of words in academic language use that single word lists cannot provide, thus contributing to more sophisticated vocabulary use and enhancing conscious vocabulary development.

- Based on the CORES corpus a software is being developed to further assist English majors in their vocabulary development and practice both in- and out of class. According to my plans the new tool would generate C-tests and/or cloze tests out of the whole corpus or any of the three sub-corpora at various word frequency levels requested by the user. The tool could be made available free to all English majors at UP and perhaps to a wider audience, however, that will require further technical and financial resources.

Implications for language assessment
- Converging quantitative and qualitative methods of vocabulary assessment in a single test format has been reported in two of the pilot studies attempting to bridge the gap in research literature in this field.
- The results of the empirical studies seem to suggest that in language testing vocabulary and grammar are closely related and may be treated as one

construct.

- The findings of this research has advised vocabulary assessment on what words to focus on in the selection of lexis to be tested for the purposes of filtering students with insufficient receptive vocabulary size for pursuing successful studies in English at university.

- The CORES-based battery may contribute to more effective vocabulary assessment based on the special lexical needs of English majors.

It is yet to be seen how these suggestions meet the interest of the target audiences. I hope, however, that the thoughts and questions raised in these studies contribute to a better understanding of the lexical knowledge and needs of English majors and fill the gap in the field of vocabulary assessment.

References

Abutalebi, J., Cappa, S.F., & Perani, D. (2005). What can functional neuroimaging tell us about the bilingual brain? In J.F. Kroll & A.M.B. de Groot (Eds.), *Handbook of bilingualism. Psycholinguistic Approaches* (pp. 497-515). Oxford: Oxford University Press.

Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics, 24* (4), 425-438.

Aitchison, J. (1994a). *Words in the mind. An introduction to the mental lexicon.* Oxford: Blackwell.

Aitchison, J. (1994b). Understanding words. In G. Brown, K. Malmkjær, A. Pollit, & J. Williams (Eds.), *Language and understanding* (pp. 81-96). Oxford: Oxford University Press.

Alderson, J. C. (2000). *Assessing reading.* Cambridge: Cambridge University Press.

Alderson, J.C. (2007). Judging the frequency of English words. *Applied Linguistics, 28*(3), 383-409.

Alderson, J. C., & Banerjee, J. (2001). Language testing and assessment (Part 1). *Language Teaching, 34*, 213-236.

Alderson, J.C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation.* Cambridge: Cambridge University Press.

Andor, J. (1985). On the relevance of frames. *Quderni di Semantica, 6* (2), 212-221.

Andor, J. (1998). A komplex lexikális-jegy analízis és a mormota esete [The case of complex lexical feature analyisis and the marmot]. In Cs. Pléh & M. Gy ri (Eds.), *A kognitív szemlélet és a nyelv kutatása* (pp. 83-100). Budapest: Pólya Kiadó.

Andor, J. (2003). Functional studies in the polarity and gradation of amplifier adjectives and adverbs in English. In J. Andor, M. Nikolov, & J. Horváth (Eds.), *Studies in English theoretical and applied linguistics* (pp.43-59). Pécs: Lingua

Franca Csoport.

Andor, J. (2004). The master and his performance: An interview with Noam Chomsky. *Intercultural Pragmatics, 1* (1), 93-111.

Andor, J. (2009). A kulcsszavakról [On keywords]. In J. Sebestyén, & Zs. Surányi (Eds.), *A nyelv, az irodalom és a kultúra varázsa* (pp. 23-28). Veszprém: Veszprémi Egyetem.

Andor, J., Nikolov, M. & Horváth, J. (Eds.). (2003). *Studies in English theoretical and applied linguistics*. Pécs: Lingua Franca Csoport.

Arnaud, P. & Béjoint, H. (Eds.). (1992). *Vocabulary and Applied Linguistics.* London: Macmillan.

Asher, N. & Pustejovsky, J. (2000). The metaphysiscs of words in context. Retrieved on 11/01/2008. http://www.cs.brandeis.edu/~jamesp/articles/ metaphysics-words.pdf

Austin, J. (1962). *How to do things with words.* Oxford: Oxford University Press.

Avrutin, S. (2000). Comprehension of discourse-linked and non-discourse-linked questions by children and Broca's aphasics. In Y. Grodzinsky, L. Saphiro, & D. Swinney (Eds.), *Language and the brain. Representation and processing* (pp. 295-315). San Diego, California: Academic Press.

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

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.P., & Palmer, A.S. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Baker, M. (1995). Corpora in translation studies: an overview and some suggestions for future research. *Target, 7* (2), 223-243.

Baker, R. (1997). *Classical Test Theory and Item Response Theory in test analysis.* Language Testing Update, Special Report No. 2. Lancaster: Centre for English Language Education.

Bánréti, Z. (1999). *Nyelvi struktúra és az agy. Neurolingvisztikai tanulmányok* [Linguistic structure and the brain. Studies in neurolinguistics]. Budapest:

Corvina.

Bárdos, J. (2002). *Az idegennyelvi mérés és értékelés elmélete és gyakorlata* [The theory and practice of assessing foreign language proficiency]. Budapest: Nemzeti Tankönyvkiadó.

Bardovi-Harlig, K. (2002). Pragmatics and second language acquisition. In R.B. Kaplan (Ed.), *The Oxford handbook of applied linguistics* (pp. 182-192). Oxford: Oxford University Press.

Beeckmans, R., Eyckmans, J., Janssens, V., Dufranne, M., & Van de Velde, H. (2001). Examining the Yes/No vocabulary test: some methodological issues in theory and practice. *Language Testing, 18* (3), 235-274.

Berthier, M.L, Green, C., Lara, J.P., Higueras, C., Barbancho, M.A., Dávila, G., & Pulvermüller, F. (2009). Memantine and constraint-induced aphasia therapy in chronic poststroke aphasia. *Ann Neurol, 65* (5), 577-85.

Bialystok, E. (Ed.). (1991). *Language processing in bilingual children*. Cambridge: Cambridge University Press.

Bialystok, E. (2005). Consequences of bilingualism for cognitive development. In J.F. Kroll & A.M.B. de Groot (Eds.), *Handbook of bilingualism. Psycholinguistic Approaches* (pp. 417-432). Oxford: Oxford University Press.

Biber, D., Conrad, D., & Reppen, R. (1998). *Corpus linguistics.* Cambridge: Cambridge University Press.

Biber, D, & Finnegan, E. (1994). Socilolinguistic perspectives on register. Oxford: Oxford University Press.

Bloom, P. (2000). *How children learn the meanings of words.* Cambridge, MA: MIT Press.

Bloomfield, L. (1933). *Language.* New York: Holt, Rinehart and Winston.

Blumstein, S. E., & Milberg, W. P. (2000). Language deficits in Broca's and Wericke's aphasia: a singualr impairment. In Y. Grodzinsky, L. Saphiro, & D. Swinney (Eds.), *Language and the brain. Representation and processing* (pp. 167-184). San Diego, California: Academic Press.

Bogaards, P. (2000). Testing L2 vocabulary knowledge at a high level: the case of the

E*uralex* French tests. *Applied Linguistics, 21* (4), 490-516.

de Bot, K. (2002). Cognitive processing in bilinguals: language choice and code-switching. In R.B. Kaplan (Ed.), *The Oxford handbook of applied linguistics* (pp. 287–300). Oxford: Oxford University Press.

Boyd-Zimmerman, C. (1997). Historical trends in second language vocabulary instruction. In J. Coady & T. Huckin (Ed.), *Second language vocabulary acquisition* (pp. 5-19). Cambridge: Cambridge University Press.

Brown, G., Malmkjær, K., Pollit, A., & Williams, J. (Eds.). (1994). *Language and understanding.* Oxford: Oxford University Press.

Brown, C., Sagers, S.L., & LaPorte, C. (1999). Incidental vocabulary acquisition from oral and written dialogue journals. *Studies in Second Language Acquisition, 21* (2), 259-283.

Cameron, L. (2002). Measuring vocabulary size in English as a additional language. *Language Teaching Research, 6* (2), 145-173.

Canale, M. (1983). On some dimensions of language proficiency. In J.W. Jr. Oller, *Issues in Language Testing Research* (pp. 333-342). Rowley, Mass.: Newbury House.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1* (1), 1-47.

CANCODE. http://www.cambridge.org/elt/corpus/cancode.htm.

Caplan, D. (2000). Positron emission tomographic studies of syntactic processing. In Y. Grodzinsky, L. Saphiro & D. Swinney (Eds.), *Language and the brain. Representation and processing* (pp. 316-329). San Diego, California: Academic Press.

Caramazza, A. (1999). Hány feldolgozási szint van a lexikai hozzáférésben? In Z. Bánréti (Ed.), *Nyelvi struktúra és az agy. Neurolingvisztikai tanulmányok* (pp. 249-286). Budapest: Corvina.

Caramazza, A. (2000). Aspects of lexical access: evidence from aphasia. In Y. Grodzinsky, L. Saphiro, & D. Swinney (Eds.), *Language and the brain. Representation and processing* (pp. 204-228). San Diego, California: Academic Press.

Caron, J. (1992). *An introduction to psycholinguistics.* New York: Harvester and Wheatsheaf.

Carston, R. (1990). Language and cognition. In F. J. Newmeyer (Ed.), *Linguistics: The Cambridge Survey,* Vol. III. (pp. 38-69). Cambridge: Cambridge University Press.

Carter, R. (1998). *Vocabulary. Applied linguistic perspectives.* London: Routledge.

Celce-Murcia, N., Dörnyei, Z., & Thurrell, S. (1995). Communicative competence: a pedagogically motivated model with content specifications. *Issues in Applied Linguistics, 6* (2), 347-372.

Chapelle, C. A. (1994). Are C-tests valid measures for L2 vocabulary research? *Second Language Research, 10,* 157-187.

Chapelle, C.A. (1998). Some notes on Systemic-Functional linguistics. Retrieved on 10/27/2007 from www.public.iastate.edu/~carolc/LING511/sfl.html

Chiang, S. Y. (1999). Assessing grammatical and textual features in L2 writing samples: the case of French as a foreign language. *The Modern Language Journal, 83* (2), 219-232.

Chomsky, N. (1957). *Syntactic structures.* The Hague: Mouton.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Chomsky, N. (1968). *Language and mind*. New York: Brace & World, Inc.

Chomsky, N. (1970). Remarks on nominalization. In R.A. Jacobs & P. Rosenbaum (Ed.), *Readings in English transformational grammar* (pp.184-221). Waltham, MA.: Ginn.

Christoffels, I. K., & de Groot, A. M. B. (2005). Simultaneous interpreting: A cognitive perspective. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism. Psycholinguistic approaches* (pp. 454-479). Oxford: Oxford University Press.

Clark, H. H. (1992). *Arenas for language use.* Chicago: The University of Chicago Press.

Coady, J. (1997a). L2 vocabulary acquisition: A synthesis of the research. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 273-290)*. Cambridge: Cambridge University Press.

Coady, J. (1997b). L2 vocabulary acquisition through extensive reading. In J. Coady, & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 225-237). Cambridge: Cambridge University Press.

Coady, J., & Huckin, T. (Eds.). (1997). *Second language vocabulary acquisition.* Cambridge: Cambridge University Press.

Cobb, T. (2000). Compleat Lexical Tutor. [site Web] http://www.lextutor.ca.

Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning and Technology, 11* (3), 38-63.

Cobb, T. (2008). Commentary: Response to McQuillan and Krashen (2008). *Language Learning and Technology, 12* (1), 109-114.

Collins, A.M., & Loftus, E.F. (1975). A spreading activation theory of semantic processing. *Psychological Review, 82,* 407-428.

Colombo, L., Stoianov, I., Pasini, M., & Zorzi, M. (2006). The role of phonology in the inflection of Italian verbs: A connectionist investigation. *The Mental Lexicon*, *1* (1), 147-181.

Common European Framework of Reference for Languages. 2001. Strasbourgh: Council of Europe.

Cook, V. (1988). *Chomsky's Universal Grammar: an introduction.* Oxford: Blackwell.

Cortes, V., & Csomay, E. (2007). Positioning lexical bundles in university lectures. Spoken Corpora in applied linguistics. Linguistic insights: Studies in language and communication. Bern: Peter Lang.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213-238.

Cruise, D. A. (1986). *Lexical semantics.* Cambridge: Cambridge University Press.

Csomay, E. (2005a). Linguistic variation in the lexical episodes of university classroom talk. In A. Tyler, M. Takada, Y. Kim, & D. Marinova (Eds.), *Language in use. Cognitive and discourse perspectives on language and language learning. Georgetown University round table on languages and linguistics* (pp. 150-162). Georgetown: Georgetown University Press.

Csomay, E. (2005b). Linguistic variation within university classroom talk: A corpus-based perspective. *Linguistics and Education, 15*, 243–274.

Csomay, E. (2006). Academic talk in American university classrooms: crossing the boundaries of oral-literate discourse? *Journal of English for Academic Purposes, 5,* 117–135**.**

Csomay, E. (2007). (under review) A corpus-based look at linguistic variation in classroom interaction: Teacher talk versus student talk in American university classess. Journal of English for Academic Purposes. Special Edition.

D'Anna, C. L., Zeichmeister, E. B., & Hall, J. W. (1991). Toward a meaningful definition of vocabulary size. *Journal of Reading Behaviour, 23,* 109-22.

Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (Eds.). (1999). *Dictionary of language testing.* Cambridge: UCLES.

DeKeyser, R., & Larson-Hall, J. (2005). What does the critical period really mean? In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of Bilingualism. Psycholinguistic Approaches* (pp. 88-108). Oxford: Oxford University Press.

Dijkstra, A., & Van Heuven, W. J. B. (1998). The BIA model and bilingual word recognition. In J. Grainger & A.M. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 189-225). Mahwah, NJ: Erlbaum.

Doró, K. (2007). Advanced vocabulary for language majors: an advantage or a must? In T. Váradi (Ed.), *Válogatás az I. Alkalmazott Nyelvészeti Konferencia el adásaiból* [*Selected Papers from the 1st Applied Linguistics PhD Conference]* (pp. 22-31). Budapest: MTA Nelvtudományi Intézet.

Doró, K. (2008). *The written assessment of the vocabulary knowledge and use of English majors in Hungary.* Unpublished PhD thesis. University of Szeged.

Dörnyei, Z. (1998). Motivation in second and foreign language learning. *Language Teaching, 31*, 117-135.

Dörnyei, Z. (1999). Motivation. In B. Spolsky (Ed.), *Concise encyclopedia of educational linguistics* (pp. 525-532). Amsterdam: Elsevier.

Dörnyei, Z. (2001). *Teaching and researching motivation*. London: Longman.

Dörnyei, Z., & Csizér, K. (1998). Ten commandments for motivating language learners: results of an empirical research. *Language Teaching Research, 2*, 203-229.

Dörnyei, Z., Csizér, K., & Németh, N. (2006). *Motivational dynamics, language attitudes and language globalisation: A Hungarian perspective.* Clevedon, UK: Multilingual Matters.

Duran, P., Malvern, D., Richards, B., & Chipere, N. (2004). Developmental trends in lexical diversity. *Applied Linguistics, 25* (2), 220-242.

Eco, U., Santambrogio, M., & Violi, P. (Eds.). (1988). *Meaning and mental representations.* Bloomington: Indiana University Press.

Ellis, N.C. (1997). Vocabulary acquisition: word structure, collocation, word class, and meaning. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy,* (pp. 122-139). Cambridge: Cambridge University Press.

Ellis, R. (1994). *The study of second language acquisition.* Oxford: Oxford University Press.

Ellis, R. (1999). Item versus system learning: explaining free variation. *Applied Linguistics, 20* (4), 460-480.

Ellis, R., & He, X. (1999). The roles of modified input and output in the incidental acquisition of word meanings. *Studies in Second Language Acquisition, 21* (2), 285-301.

Ellman, J. L., & McClelland, J. L. (1986). Interactive processes in speech perception: the TRACE model. In J. L. McClelland, D. Rummelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition,* Vol. 2. (pp. 58-121). Cambridge, Mass.: MIT Press.

Emmorey, K., & Fromkin, V. (1988). The mental lexicon. In Newmeyer (Ed.), *Linguistics: The Cambridge Survey. III. Language: Psychological and Biological Aspects* (pp. 124-149). Cambridge: Cambridge University Press.

Erben, T. (2004). Emerging research and practices in immersion language learning and teaching. *Annual Review of Applied Linguistics, 24*, 320-338.

Eubank, L., & Gregg, K. (1995). "Et in amygdala ego"? UG, (S)LA, and neurobiology. *Studies in Second Language Acquisition, 17*, 35-57.

Feldman, L. B., Basnight-Brown, D., & Pastizzo, M. J. (2006). Semantic influences on

morphological facilitation: Concreteness and family size. *The mental lexicon*, *1* (1), 35-58.

Fillmore, C.J. (1985). Frames and the semantics of understanding. *Quaderni di Semantica, 6* (2), 222-254.

Flowerdew, J. (2001). Concordancing as a tool in course design. In M. Ghadessy, A. Henry, & R. Roseberry (Eds.), *Small corpus studies in ELT: Theory and practice* (pp. 71-93). Philadelphia: Benjamins.

Flowerdew, L. (2003). A combined corpus and systemic-functional analysis of the problem-solution pattern in a student and professional corpus of technical writing. *TESOL Quarterly, 37* (3), 489-510.

Flowerdew, L. (2005). An integration of corpus-based and genre-based approaches to text analysis in EAP/ESP: countering criticisms against corpus-based methodologies. *English for Specific Purposes, 24*, 321–332.

Fodor, J. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

Forster, K. I. (1976). Accessing the mental lexicon. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms* (pp. 257-287). Amsterdam: North Holland.

Forster, K.I. (2006). Early activation of category information in visual word recognition. *The mental lexicon*, *1*(1), 35-58.

Fraser, C. A. (1999). Lexical processing strategy use and vocabulary learning through reading. *Studies in Second Language Acquisition, 21* (2), 225-241.

Fraser, C. A. (2007). Reading rate in L1 Mandarin Chinese and L2 English across five reading tasks. *The Modern Language Journal, 91* (3), 372-394.

Galloway, J. (1989). *The trick is to keep breathing.* London: Minerva.

Gardner, D. (2004). Vocabulary input through extensive reading: A comparison of words found in children's narrative and expository reading materials. *Applied Linguistics, 25* (1), 1-37.

Gardner, R. C., & MacIntyre, P. D. (1992). A student's contributions to second language learning. Part I: Cognitive variables. *Language Teaching, 25*, 211-220.

Gardner, R.C., & MacIntyre, P. D. (1993a). A student's contribution to second language acquisition. Part II: Affective variables. *Language Teaching*, 26, 1-11.

Gardner, R.C., & MacIntyre, P.D. (1993b). On the measurement of affective variables in second language learning. *Language Learning, 43*, 157-194.

Garman, M. (1990). *Psycholinguistics.* Cambridge: Cambridge University Press.

Garrett, M. (2000). Remarks on the architecture of language processing systems. In Y. Grodzinsky, L. Saphiro, & D. Swinney, *Language and the brain. Representation and processing* (pp. 31-74). San Diego, California: Academic Press.

Garrod, S., & Pickering, M.J. (Eds.). (1999). *Language processing.* Hove: Psychology Press.

Gass, S. (1999). Incidental vocabulary learning. *Studies in second language acquisition, 21* (2), 319-333.

Gleitman, L., & Landau, B. (Eds). (1994). *The acquisition of the lexicon.* Cambridge, Massachusettes: The MIT Press.

Gósy, M. (1999). *Pszicholingvisztika* [Psycholinguistics]. Budapest: Corvina.

Goulden, R., Nation, P., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics, 11* (4), 341-363.

Grabe, W. (2002). Reading in a second language. In R.B. Kaplan (Ed.), *The Oxford handbook of applied linguistics* (pp. 49-59). Oxford: Oxford University Press.

Grabe, W., & Stoller, F. L. (1997). Reading and vocabulary development in a second language: A case study. In J. Coady, & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 98-122). Cambridge: Cambridge University Press.

Grainger, J., & Dijkstra, A. (1992). On the representation and use of language information in bilinguals. In R. Harris (Ed.), *Cognitive processing in bilinguals* (pp. 207-220). Amsterdam: Elsevier.

Granger S., Hung, J., & Petch-Tyson, S. (2002). *Computer learner corpora, second language acquisition and foreign language teaching.* Amsterdam, Philadephia: John Benjamins Publishing Company.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1,* 67-81.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics. Vol. 3: Speech acts* (pp. 41-58). New York: Academic Press.

Grodzinsky, Y., Saphiro, L., & Swinney, D. (Eds.). (2000). *Language and the brain.*

Representation an processing. San Diego, California: Academic Press.

de Groot, A. M. B. (1992). Bilingual lexical representation: A closer look at conceptual representations. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology and meaning* (pp. 389-412). Amsterdam: Elsevier Science.

de Groot, A. M. B., & Keijzer, R. (2000). What is hard to learn is easy to forget: The roles of word concreteness, cognate status, and word frequency in foreign language vocabulary learning and forgetting. *Language Learning, 50,* 1-56.

de Groot, A. M. B., & Poot, R. (1997). Word translation at three levels of proficiency in a second language: The ubiquitous involvement of conceptual memory. *Language Learning, 47*, 215-264.

de Groot, A. M. B., & van Hell, J. G. (2005). The learning of foreign language vocabulary. In J.F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism. Psycholinguistic approaches* (pp. 9-29). Oxford: Oxford University Press.

Gu, Y. P. (2003). Vocabulary learning in a second language: Person, task, context and strategies. *Teaching English as a Second or Foreign Language - Electronic Journal, 7* (2), A4. Retrieved on 03/11/2008 from http://www-writing.berkeley.edu/TESl-EJ/ej26/a4.html

Gu, Y. & Johnson, R.K. (1996). Vocabulary learning strategies and language learning outcomes. *Language Learning, 46* (4), 643-679.

Hall, C. J. (2002). The automatic cognate form assumption: Evidence for the parasitic model of vocabulary development. *IRAL, 40,* 69-87.

Halle, M. (1973). Prolegomena to a theory of word-formation. *Linguistic Inquiry, 4*, 3-16.

Halliday. M. A. K. (1985). *An introduction to functional grammar.* London: Edward Arnold.

Halliday, M. A. K., & Hasan, R. (1989). *Language, context and text: Aspects of language in a social-semiotic perspective.* Oxford: Oxford University Press.

Hatch, E., & Lazaraton, A. (1991). *The research manual. Design and statistics for applied linguistics.* Boston, Mass.: Heinle and Heinle Publishers.

Hazenberg, S., & Hulstijn, J. H. (1996). Defining a minimal receptive second-language vocabulary for non-native university students: An empirical investigation. *Applied Linguistics, 17* (2), 145-163.

Hedgecock, J., & Lefkowitz, N. (1999). Some input on input: two analyses of student response to expert feedback in L2 witing. *The Modern Language Journal, 83* (3), 287-308.

Heij, W. L. (2005). Selection processes in monolingual and bilingual lexical access. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of Bilingualism. Psycholinguistic Approaches* (pp. 289-307). Oxford: Oxford University Press.

van Hell, J. G., & de Groot, A. M. B. (1998). Conceptual representation in bilingual memory: effects of concreteness and cognate status. *Bilingualism: Language and Cognition, 1*, 193-211.

Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition, 21* (2), 303-317.

Henry, K. (1999). Early L2 writing development: a study of autobiographical essays by university-level students of Russian. *The Modern Language Journal, 83* (3), 309-326.

Henry, A., & Roseberry, R. L. (2007). Language errors in the genre-based writing of advanced academic ESL students. *RELC Journal, 38* (2), 171-198.

Hickok, G. (2000). Speech perception, conduction aphasia, and the functional neuroanatomy of language. In Y. Grodzinsky, L. Saphiro, & D. Swinney (Eds.), *Language and the brain. Representation and processing* (pp. 87-104). San Diego, California: Academic Press.

Hirsch, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language, 8* (2), 689-96.

Horváth, J. (1996). The assessment of essay writing skills in the first-year proficiency test. In M. Nikolov & J. Horváth (Eds.), *Learning lessons: innovation in teacher education and assessment* (pp. 88-106). Pécs: Lingua Franca Csoport.

Horváth, J. (2001). *Advanced writing in English as a foreign language. A corpus-based study of processes and products.* Pécs: Lingua Franca Csoport.

Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language. *Studies in Second Language Acquisition, 21* (2), 181-193.

Hughes, A. (1991). *Testing for language teachers.* Cambridge: Cambridge University Press.

Huibregtse, I., Admiraal, W., & Meara, P. (2002). Scores on a yes-no vocabulary test: correction for guessing and response style. *Language Testing, 19* (3), 227-245.

Hulstijn, J. (1997). Mnemonic methods in foreign language vocabulary learning: Theoretical considerations and pedagogical implications. In J. Coady, & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 203-224). Cambridge: Cambridge University Press.

Hulstijn, J.H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrance of unknown words. *The Modern Language Journal, 80*, 327-339.

Hunston, S. (2002). *Corpora in applied linguistics.* Cambridge: Cambridge University Press.

Hunt, A. & Beglar, D. (1998). Current research and practice in teaching vocabulary. *The Language Teacher.* Retrieved on 07/02/ 2008 from http://www.jalt-publications.org/tlt/articles/1998/01/hunt

Hyland, K. (2007). Applying a gloss: Exemplifying and reformulating in academic discourse. *Applied Linguistics, 28* (2), 266-285.

Hyland, K., & Tse, P. (2007). Is there an "academic vocabulary"? *TESOL Quarterly, 41* (2), 235-253.

Hymes, D. (1971). *On communicative competence.* Philadelphia, PA: University of Pennsylvania Press.

Jackendoff, R. (1990). *Semantic structures.* Cambridge, MA: MIT Press.

Jackendoff, R. (1993). *Languages of the mind. Essays on mental representation.* Cambridge, MA: MIT Press.

Jackendoff, R. (2000). Fodorian modularity and representational modularity. In Y. Grodzinsky, L. Saphiro, & D. Swinney (Eds.), *Language and the brain.*

Representation and processing (pp. 4-30). San Diego, California: Academic Press.

Jacobs, B. (1995). Dis-integrating perspectives of language acquisition. A response to Eubank and Gregg. *Studies in Second Language Acquisition, 17*, 65-71.

Jarvis, S. (2002). Short texts, best fitting curves and new measures of lexical diversity. *Language Testing, 19* (1), 57-84.

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics, 21* (1), 47-77.

Jones, F. R. (1995). Learning an alien lexicon: A teach-yourself case study. *Second Language Research, 11*, 149-158.

Kaplan, R.B. (Ed.). (2002). *The Oxford Handbook of Applied Linguistics.* Oxford: Oxford University Press.

Kempen, G., & Huijbers, P. (1983). The lexicalization process in sentence production and naming: Indirect election of words. *Cognition, 14*, 185-209.

Kétszint érettségi, angol nyelv, vizsgaleírás (2002). Oktatási és Közm vel dési Minisztérium. Retrieved on 22/09/2009 from http://www.okm.gov.hu/ letolt/okev/doc/ erettsegi_40_2002/angol_nyelv_vizsgaleiras.pdf

Kiefer, F. (2007). *Jelentéselmélet* [Theory of meaning]. Budapest: Corvina.

Kim, H., Relkin, N.R., Lee, K., & Hirsch, J. (1997). Distinct cortical areas associated with native and second languages. *Nature, 388,* 171-174.

Koda, K. (1997). Orthographic knowledge in L2 lexical processing: A cross-linguistic approach. In J. Coady, & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 35-54). Cambridge: Cambridge University Press.

Koda, K. (2005). *Insights into second language reading. A cross-linguistic approach.* Cambridge: Cambridge University Press.

Kojic-Sabo, I., & Lightbown, P.M. (1999). Students' approaches to vocabulary learning and their relationship to success. *Modern Language Journal, 83* (2), 176-192.

Kormos, J., Hegybiró-Kontra, E., & Csölle, A. (2002). Language wants of English majors in a non-native context. *System, 30,* 517-542.

Krashen, S. (1985). *The input hypothesis: Issues and implications.* Torrance, CA:

Laredo Publishing Company, Inc.

Krashen, S. (1989). We acquire vocabulary and spelling by reading: additional evidence for the input hypothesis. *The Modern Language Journal, 73,* 440-464.

Krashen, S. (2004). *The power of reading. 2nd edition.* Portsmouth, NH: Heinemann.

Kroll, J.F., & De Groot, A.M.B. (Eds.). (2005). *Handbook of bilingualism. Psycholinguistic approaches.* Oxford: Oxford University Press.

Kroll, J. F., & Dijkstra, T. (2002). The bilingual lexicon. In R. B. Kaplan (Ed.), *The Oxford handbook of applied linguistics* (pp. 301-321). Oxford: Oxford University Press.

Kroll, J. F., & Stewart, E. (1994). Category inference in translation and picture naming: Evidence for assymmetric connections between bilingual memory representations. J*ournal of Memory and Language, 33,* 149-174.

Kroll, J. F., & Sunderman, G. (2003). Cognitive processes in second language learners and bilinguals: the development of lexical and conceptual representations. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 104-129). Oxford: Blackwell Publishing.

Kroll, J. F., & Tokowicz, N. (2005). Models of bilingual representation and processing. Looking back and to the future. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism. Psycholinguistic Approaches* (pp. 531-553). Oxford: Oxford University Press.

Kuiken, F., Mos, M., & Vedder, I. (2005). Cognitive task complexity and second language writing performance. In S. H. Foster-Cohen, M. P. García-Mayo, & J. Cenoz (Eds.), *The EUROSLA Yearbook. Volume 5.* (pp. 195-222). Amsterdam: John Benjamins.

Langacker, R.W. (1987). *Foundations of cognitive grammar.* Vol. 1. Stanford, CA: Stanford University Press.

Langacker, R.W. (1991). *Concept, image, and symbol. A cognitive basis of grammar.* Berlin: Mouton de Gruyter.

Laufer, B. (1987). Teaching vocabulary: The lexical perspective of reading comprehension. *English Language Teaching Journal* (Israel), *35*, 58-67.

Laufer, B. (1992a). Reading in a foreign language: How does L2 lexical knowledge interact with the reader's general academic ability? *Journal of Research in Reading, 15* (2), 95-103.

Laufer, B. (1992b). How much lexis is necessary for reading comprehension? In P. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 126-132). London: Macmillan.

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 (Eds.), *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. (2001). Quantitative evaluation of vocabulary: how can it be done and what it is good for. In C. Elder, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lumley, T. McNamara, & K. O'Loughlin (Eds.), *Experimenting with uncertainty. Essays in honour of Alan Davies* (pp. 241-250). Cambridge: Cambridge University Press.

Laufer, B. (2005a). Focus on form in second language vocabulary learning. *EUROSLA Yearbook, 5*, 223-250.

Laufer, B. (2005b). Instructed second language vocabulary learning: The fault in the 'default hypothesis'. In A. Housen & M. Pierrard (Eds.), *Investigations in instructed second language learning* (pp. 311-329). Berlin: Walter De Gruyter**.**

Laufer, B., Elder, C., Hill, C., & Congdon, P. (2004). Size and strength: Do we need both to measure vocabulary knowledge? *Language Testing, 21* (2), 202-226.

Laufer, B., & Hill, M. (2000). What lexical information do L2 learners select in a CALL dictionary and how does it affect word retention? *Language Learning and*

Technology, 3 (2), 58-76.

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: the construct of task-induced involvement. *Applied Linguistics, 22* (1), 1-26.

Laufer, B., & Nation, P. (1995). Vocabulary size and use: lexical richness in L2 written production. *Applied Linguistics*, *16* (3), 307-322.

Laufer, B., & Nation, P. (1999). A vocabulary size test of controlled productive ability. *Language Testing, 16*(1), 33-51.

Laufer, B., & Paribakht, T. (1998). The relationship between passive and active vocabularies: Effects of language learning context. *Language Learning, 48*, 365–391.

Laviosa, S. (1998). Core patterns of lexical use in a comparable corpus of English narrative prose. *Meta, 43* (4), 557-570.

Lawson, M. J., & Hogben, D. (1996). The vocabulary-learning strategies of foreign-language students. *Language Learning, 46* (1), 101-135.

Lee, D. Y. W. (2001). Genres, registers, text types, domains and styles: Clarifying the concepts and navigating a path through the BNC jungle. *Language Learning and Technology, 5* (3), 37-72.

Lehmann, M. (2007a). Is intentional or incidental vocabulary learning more effective? *International Journal of Foreign Language Teaching, 3* (1), 23-28. http://www.tprstories.com/ijflt

Lehmann, M. (2007b). The lexical diversity of short texts: Exploring the productive-receptive continuum of lexical knowledge. In J. Horváth, & M. Nikolov (Eds.), *UPRT 2007: Empirical studies in English applied linguistics* (pp. 293-305). Pécs: Lingua Franca Csoport.

Lehmann, M. (2007c). 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 Theoretical and Applied Linguistics* (pp. 25-38).

Lehmann, M. (2006a). Vocabulary as a filter with first-year English majors. In M.

Nikolov, & J. Horváth, *UPRT 2006. Empirical studies in English Applied Linguistics*, (pp. 121-137). Pécs: Lingua Franca Csoport.

Lehmann, M. (2006b). Magoljunk vagy olvassunk?: El tanulmány a tudatos és járulékos szótanulás összehasonlítására. *Modern nyelvoktatás, 12* (1), 44-58.

Lehmann, M. (2004). Az idegennyelvi szókincs elsajátítás és mérés innovatív irányzatai [Innovative trends in the aquisition and assessment of foreign language vocabulary]. *Alkalmazott nyelvtudomány, 4* (1), 41-57.

Lehmann, M. (2003a). Range. Software review. *Modern nyelvoktatás, 9* (4), 88-89.

Lehmann, M. (2003b). 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,* (172-182). Pécs: Lingua Franca Csoport.

Levelt, W. (1989). *Speaking: from intention to articulation.* Cambridge, MA: MIT Press.

Lieber, R. (1980). *The organization of the lexicon.* PhD dissertation, MIT. Distributed by the Indiana University Linguistics Club 1981.

Lindgren, E., & Sullivan, K.P.H. (2002). The LS graph: A methodology for visualizing writing revision. *Language Learning*, *52* (3), 565-595.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, Mass.: Addison-Wesley.

Mackenzie, J. L. (2003). Such and textual connectivity. In J. Andor, M. Nikolov, & J. Horváth (Eds.), *Studies in English theoretical and applied linguistics* (pp. 60-73). Pécs: Lingua Franca Csoport.

MacWhinney, B. (2005). A unified model of language acquisition. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism. Psycholinguistic approaches* (pp. 49-67). Oxford: Oxford University Press.

MacWhinney, B., & Leinbach, J. (1991). Implementations are not conceptualizations: Revising the verb learning model. *Cognition, 40,* 121-157.

Marslen-Wilson, W. (1987). Functional parallelism in spoken word recognition. *Cognition, 25*, 71-102.

Marslen-Wilson, W. (1999). Abstractness and combination: The morphemic lexicon.

In S. Garrod & M. J. Pickering (Eds.), *Language processing* (101-119) Hove: Psychology Press.

Martsa, S. (2007). *English morphology. An introduction.* Budapest: Nemzeti Tankönyvkiadó.

McCarthy, M. (1990). *Vocabulary.* Oxford: Oxford University Press.

McCrostie, J. (2007). Investigating the accuracy of teachers' word frequency intuitions. *RELC Journal, 38*(1), 53-66.

McLaughlin, B. & Robbins, S. (1999). Second language learning. In B. Spolsky (Ed.). *Concise encyclopedia of educational linguistics* (pp. 540-552). Amsterdam: Elsevier.

McQuillan, J., & Krashen, S. (2008). Can free reading take you all the way? A response to Cobb (2007). *Language Learning and Technology, 12*(1), 104-108.

Meara, P. (1992). *EFL vocabulary test*. Swansea, UK.: Center for Applied Language Studies.

Meara, P. (1999). Lexis: Acquisition. In B. Spolsky (Ed.), *Concise encyclopedia of educational linguistics* (pp. 565-567). Amsterdam: Elsevier.

Meara, P. (2004). Modelling vocabulary loss. *Applied Linguistics, 25* (2), 137-155.

Meara, P. (2005). Reactivating a dormant vocabulary. *Eurosla Yearbook, 5*, 269-280.

Meara, P., & Jones, G. (1990). *Eurocentres vocabulary size test. Version E1.1/K10.* Zurich: Eurocentres Learning Service.

Medgyes, P. and Nikolov, M. (2002) Curriculum development: The interface between political and professional decisions. In R. Kaplan (Ed.), *The Oxford Handbook of Applied Linguistics* (pp. 195-206). London: Oxford University Press.

Melka, F. (1997). Receptive versus productive aspects of vocabulary knowledge. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: description, acquisition and pedagogy* (pp.84-102) . Cambridge: Cambridge University Press.

Milroy, L. (1994). Sociolinguistics and second language learning. In G. Brown, K. Malmkjær, A. Pollit, & J. Williams (Eds.), L*anguage and understanding* (pp. 153-168). Oxford: Oxford University Press.

Mochida, A. & Harrington, M. (2006). The Yes/No test as a measure of receptive vocabulary knowledge. *Language Testing, 23* (1), 73-98.

Morris, L., & Cobb, T. (2003). Vocabulary profiles as predictors of the academic performance of Teaching English as a Foreign Language trainees. *System, 32* (1),75-87.

Morrison, C. M., & Ellis, A. W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21*, 116-153.

Morton, J. (1982). Disintegrating the lexicon: An information processing approach. In J. Mehler, E. C. T. Walker & M. F. Garret (Eds.), *Perspectives on mental representation.* (pp. 89-109). Hillsdale, NJ: Erlbaum.

Munby, J. (1978). *Communicative syllabus design.* Cambridge: Cambridge University Press.

Muncie, J. (2002). Process writing and vocabulary development: comparing lexical frequency profiles across drafts. *System, 30,* 225-235.

Murphy, J., & Kandil, M. (2004). Word-level stress patterns in the academic word list. *System, 32* (1), 61-74.

Murre, J. M. J. (2005). Models of monolingual and bilingual language acquisition. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of Bilingualism. Psycholinguistic Approaches* (pp. 154-169). Oxford: Oxford University Press.

Myers-Scotton, C. (2005). Supporting a differential access hypothesis: Code switching and other contact data. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of Bilingualism. Psycholinguistic Approaches* (pp. 326-348). Oxford: Oxford University Press.

Nagy, B.C. (2008). *"To will or not to will" Exploring advanced EFL learners' willingness to communicate in English*. Unpublished PhD thesis. University of Pécs.

Nagy, W. (1997). On the role of context in vocabulary learning. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, Acquisition and Pedagogy* (pp.64-83). Cambridge: Cambridge University Press.

Nassaji, H. (2003). L2 vocabulary learnin from context: strategies, knowledge sources, snd their relationship with success in L2 lexical inferencing. *TESOL Quarterly,*

37 (4), 645-670.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston, Mass.: Heinle and Heinle Publishers.

Nation, I. S. P. (1993a). Using dictionaries to estimate vocabulary size: essential, but rarely followed procedures. *Language Testing, 10,* 27-40.

Nation, I. S. P. (1993b). Vocabulary size, growth and use. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon* (pp. 115-134). Amsterdam: Benjamins.

Nation, I. S. P. (2001). L*earning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review, 63*(1), 1-12 .

Nation, I. S. P., & Newton, J. (1997). Teaching vocabulary. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 238-254). Cambridge: Cambridge University Press.

Nation, I. S. P., & Wang, Ming-Tzu, K. (1999). Graded readers and vocabulary. *Reading in a Second Language, 12*, 355-379.

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.

Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching.* Oxford: Oxford University Press.

Nassaji, H. (2003). L2 vocabualry learning from context: strategies, knowledge sources, and their relationship with success in L2 lexical inferencing. *TESOL Quarterly, 37* (4), 645-670.

Nemzeti alaptanterv [National core curriculum]. (2003). Budapest: Oktatási Minisztérium.

New, E. (1999). Computer-aided writing in French as a foreign language: a qualitative and quantitative look at the process of revision. *The Modern Language Journal*, 83, 80-97.

Nicol, J. L., & Love, T. (2000). Overarching agrammatism: when comprehension

involves production. In Y. Grodzinsky, L. Saphiro, & D. Swinney (Eds.), *Language and the brain. Representation and processing* (pp. 105-122). San Diego, California: Academic Press.

Nikolov, M. (1999). Classroom observation project. In M. Nikolov (Ed.), *English Language Education in Hungary. A Baseline Study* (pp. 221-246). Budapest: The British Council Hungary.

Nikolov, M. (2002). *Issues in English language education.* Bern: Peter Lang.

Nusbaum, H. C., Pisoni, D. B., & Davis, C.K. (1984). Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words. *Research on Speech Perception, Progress Report No. 10*. Bloomington, IN: Indiana University.

Oller, J.W. (1979). *Language tests at school.* London: Longman.

O'Keefee, A., McCarthy, M., & Carter, R. (2007*). From corpus to classroom: language use and language teaching.* Cambridge: Cambridge University Press.

Ötvösné Vadnai, M. (2002). A szótanulás nehézségei: szókincsmérések egyetemi hallgatókkal [The difficulties of vocabulary learning: assessing student vocabulary]. *Iskolakultúra, 12*, 112-122.

Oxford, R. L. (1999). Second language learning: Individual differences. In B. Spolsky (Ed.),*Concise encyclopedia of educational linguistics* (pp. 552-560). Amsterdam: Elsevier.

Palmer, F. R. (1981). *Semantics.* Second edition. Cambridge: Cambridge University Press.

Paradis, M. (1994). When the interpretation does not fit the facts, alter the facts, not the interpretation! A comment on Pulvermüller and Schumann (1994). *Language Learning, 45* (4), 725-727.

Paribakht, T. S., & Wesche, M. (1993). The relationship between reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal, 11* (1), 9-29.

Paribakht, T. S., & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 174-200).

Cambridge: Cambridge University Press.

Paribakht, T. S., & Wesche, M. (1999). Reading and "incidental" L2 vocabulary acquisition. An introspective study of lexical inferencing. *Studies in Second Language Acquisition, 21* (2), 195-224.

Partington, A. (1998). *Patterns and meanings. Using corpora for English language research and teaching*. Amsterdam: John Benjamins Publishing Company.

Pawley, A., & Syder, F.H. (1983). Two puzzles for linguistic theory: native-like selection and native-like fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191-226). London: Longman.

Piaget, J. (1952). *The origins of the intelligence in children.* New York: Norton.

Pickering, M. J. (1999). Sentence comprehension. In S. Garrod & M. J. Pickering (Eds.), *Language processing* (pp. 123-154). Hove: Psychology Press.

Piñango, M. M. (2000). Canonicity in Broca's sentence comprehension: the case of psychological verbs. In Y. Grodzinsky, L. Saphiro, & D. Swinney (Eds.), *Language and the brain. Representation and processing* (pp. 330-353). San Diego, California: Academic Press.

Pinker, S. (1989). *Learnability and cognition*. Cambridge, Massachusetts: MIT Press.

Pinker, S. (1999). *A nyelvi ösztön*. Budapest: Typotex.

Pinker, S. (2007). *The stuff of thought. Language as a window into human nature.* London: Penguin Books.

PISA 2000. Knowledge and skills for life. (2001). Paris: OECD.

Plunkett, K., & Marchman, V. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition, 48*, 21-69.

Poulisse, N., & Bongaerts, T. (1994). First language use in second language production. *Applied Linguistics, 15*, 36-57.

Pulido, D. (2004). The relationship between text comprehension and second language incidental vocabulary acquisition: A matter of topic familiarity? *Language Learning, 54* (3), 469-523.

Pulido, D. (2007). The effects of topic famailiarity and passage sight vocabualry on L2 lexical inferencing and retention through reading. *Applied Linguistics, 28* (1), 66-

86.

Pulvermüller, F. (1995). What neurobiology can buy language theory: A response to Eubank and Gregg. *Studies in Second Language Acquisition, 17*, 73-77.

Pulvermüller, F., & Schumann, J. H. (1994). Neurobiological Mechanisms of Language Acquisition. *Language Learning, 44* (4), 681-734.

Pustejovsky, J. (1995). *The generative lexicon.* Cambridge, Mass: MIT Press.

Pustejovsky, J. (2001). Type construction and the logic of concepts. In P. Bouillon & F. Busa (Eds.), *The syntax of word meaning*. Cambridge: CUP. Retrieved on 11/01/2008 from http://www.cs.brandeis.edu/~jamesp/articles/type-construction.pdf

Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: an assessment perspective. *Language Learning, 52* (3), 513-536.

Qian, D. D., & Schedl, M. (2004). Evaluation of an in-depth vocabulary knowledge measure for assessing reading performance. *Language Testing, 21* (1), 28-52.

Quillian, M. R. (1967). Word concepts: A theory of and simulation of some basic semantic capabilities. *Behavioural Science, 12*, 410-430.

Rasekh, Z. E., & Ranjbary, R. (2003). Metacognitive strategy training for vocabulary learning. *Teaching English as a Second or Foreign Language, 7*(2), A-5. Retrieved on 06/11/2008 from http://www-writing.berkeley.edu/TESl-EJ/ej26/a5.html

RANGE. Retreived on 16/04/2007 from http://www.vuw.ac.nz/lals/staff/paul-nation/nation.aspx .

Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing, 10,* 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.

Reeves, L. M., Hirsh-Pasek, K., & Golinkoff, R. (1998). Words and meaning: from primitives to complex organization. In J. B. Gleason & N. B. Ratner (Eds.),

Psycholinguistics (pp. 157-226). London: Thomson.

Richards, J. C. (1976). The role of vocabulary in language teaching. *TESOL Quarterly, 10,* 77-89.

Richards, J. C., & Schmidt, R. W. (Eds.). (1983). *Language and communication*. London: Longman.

Robinson, P. (2001a). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics, 22* (1), 27-57.

Robinson, P. (Ed.). (2001b). *Cognition and second language acquisition.* Cambridge: Cambridge University Press.

Rodd, J., Gaskell, G., & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: semantic competition in lexical access. *Journal of Memory and Language, 46,* 245-266.

Rodriguez, M. & Sadoski, M. (2000). Effects of rote, context, keyword, and context/keyword methods on retention of vocabualry in EFL classrooms. *Language Learning, 50* (2), 385-412.

Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition, 21*, 589-619.

Rott, S., Williams, J., & Cameron, R. (2002). The effect of multiple-choice L1 glosses and input-output cycles on lexical acquisition and retention. *Language Teaching Research, 6* (3), 183-222.

Rummelhart, D., & McClelland, J. (1986). On learning the past tenses of English verbs: Implicit rules or parallel distributed processing? In J. L. McClelland, D. E. Rummelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the micro-structure of cognition* (Vol. 2, pp. 216-271). Cambridge, MA: MIT Press.

Rummelhart, D., Lindsay, P. H., & Norman, D. A. (1972). A process model for long-term memory. In E. Tulvig & W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press.

Sanford, A. J. (1999). Word meaning and discourse processing. In S. Garrod & M. J.

Pickering (Eds.), *Language processing* (pp. 301-334). Hove: Psychology Press.

Sankó, Gy. (2008). *The effects of form- and meaning-focused hypertextual input modification on L2 vocabulary acquisition and retention*. Debrecen: Kossuth Egyetemi Kiadó.

Saphiro, A.M., & Waters, D. L. (2005). An investigation of the cognitive processes underlying the keyword method of foreign vocabulary learning. *Language Teaching Research, 9* (2), 129-146.

Schmitt, D., & Schmitt, N. (2005). *Focus on vocabualry: Mastering the Academic Word List.* Longman: White Plains, NY.

Schmitt, N. (1997). Vocabulary learning strategies. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 199-228). Cambridge: Cambridge University Press.

Schmitt, N. (2000). *Vocabulary in language teaching.* Cambridge: CUP.

Schmitt, N. (2008). Instructed second language vocabulary learning. Review article. *Language Teaching Research, 12* (3), 329-364.

Schmitt, N., & McCarthy, M. (Eds.). (1997). *Vocabulary: Description, acquisition and pedagogy.* Cambridge: Cambridge University Press.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing, 18* (1), 55-88.

Schneider, V. I., Healy, A. F., & Bourne, L. E. Jr. (2002). What is learned under difficult conditions is hard to forget: contextual interference effects in foreign vocabulary acquisition, retention, and transfer. *Journal of Memory and Language, 46,* 419-440.

Schoonen, R., & Verhallen, M. (2008). The assessment of deep word knowledge in young first and second language learners. *Language Testing, 25* (2), 211-233.

Schumann, J. (1994). Where is cognition? Emotion and cognition in second language acquisition. *Studies in Second Language Acquisition, 16,* 231-242.

Schumann, J. (1995). Ad minorem theoriae gloriam: A response to Eubank and Gregg. *Studies in Second Language Acquisition*, *17*, 59-63.

Scott, M. (1999). *Wordsmith Tools*. Software. Oxford: Oxford University Press.

Searle, J. R. (1969). *Speech acts.* Cambridge: Cambridge University Press.

Selinker, L. (1972). Interlanguage. *The International Review of Applied Linguistics, 10,* 209-231.

Sifakis, N., Oder, T., Lehmann, M. & Bluma, D. (2007). Aspects of learner autonomy in the national curricula of four European countries. In A-B. Fenner, & D. Newby (Eds.), *Coherence of principles, cohesion of competences: Exploring theories and designing materials for teacher education* (pp. 132-151). Council of Europe Publishing: Strasbourg.

Sinclair, J. (1991). *Corpus, concordance, collocation.* Oxford: Oxford University Press.

Singleton, D. (1989). *Language acquisition: The age factor*. Clevendon, Avon: Multilingual Matters.

Singleton, D. (1999). *Exploring the second language mental lexicon.* Cambridge: Cambridge University Press.

Singleton, D. (2000). *Language and the lexicon. An introduction.* London: Arnold.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and processes in semantic memory: A featural model for semantic decision. *Psychological Review, 81*, 214-241.

Slabakova, R. (2006). Is there a critical period for semantics? *Second Language Research, 22* (3), 302-338.

Spolsky, B. (Ed.). (1999). *Concise encyclopedia of educational linguistics*. Amsterdam: Elsevier.

Stockall, L., & Marantz, A. (2006). A single route, full decomposition model of morphological complexity: MEG evidence. *The mental lexicon*, *1* (1), 85-124.

Stubbs, M. (2001). *Words and Phrases: Corpus Studies of Lexical Semantics.* Cambridge, MA: Blackwell Publishers.

Swan, M. (2008). Talking sense about learning strategies. *RELC Journal, 39*, 262-273.

Swain, M. (1985). Communicative competence: some roles of comprehensible input and comprehensible output in its development. In S. Gass & M. Madden (Eds.), *Input in second language acquisition* (pp. 235-255). Rowley, MA: Newbury House.

Swinney, D., Prather, P., & Love, T. (2000). The time-course of lexical access and the role of context: converging evidence from normal and aphasic processing. In Y. Grodzinsky, L. Saphiro, & D. Swinney (Eds.), *Language and the brain. Representation and processing* (pp. 273-294). San Diego, California: Academic Press.

Szabó, G. (1996). Effective filter? Proficiency testing of first-year English majors. In M. Nikolov & J. Horváth (Eds.), *Learning lessons: innovation in teacher education and assessment* (pp. 76-87). Pécs: Lingua Franca Csoport.

Szabó, G. (2006). Anchors Aweigh! An analysis of the impact of anchor item's number and difficulty range on item difficulty calibrations. In M. Nikolov, & J. Horváth (Eds.), *UPRT 2006* (pp.249-262). Pécs: Lingua Franca Csoport.

Szabó, G. (2008). *Applying item response theory in language test item bank building*. Frankfurt am Main: Peter Lang.

Taguchi, N. (2007). Task difficulty in oral speech act production. *Applied Linguistics, 28* (1), 113-135.

Tannen, D. (1989). *Talking voices: repetition, dialogue and imagery in conversational discourse.* Cambridge: Cambridge Univeristy Press.

Tannenhaus, M. K. (1988). Psycholinguistics: an overview. In S. Newmeyer (Ed.), *Linguistics: The Cambridge Survey. III. Language: Psychological and Biological Aspects* (pp. 1-37). Cambridge: Cambridge University Press.

Tarone, E. (1999) Interlanguage. In B. Spolsky (Ed.), *Concise encyclopedia of educational linguistics* (pp. 507-512). Amsterdam: Elsevier.

Taylor, I. (1990). *Psycholinguistics. Learning and using language.* Englewood Cliffs, New Jersey: Prentice Hall.

The common European framework of reference for languages. Retrieved on 22/06/2009 from http://www.coe.int/T/DG4/Linguistic/Source/Framework_EN.pdf

Thomas, J., & Short, M. (Eds.). (1996). *Using corpora for language research*. London: Longman.

Tinkham, T. (1997). The effects of semantic and thematic clustering on the learning of second language vocabulary. *Second Language Research, 13* (2), 138-163.

Tribble, C. (2002). Corpora and corpus analysis: New windows on academic writing. In J. Flowerdew (Ed.), *Academic discourse* (pp. 131-149) London: Longman.

Turvey, M. T., & Moreno, M. A. (2006). Physical metaphors for the mental lexicon. *The mental lexicon*, *1* (1), 7-35.

Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: the declarative/procedural model. *Bilingualism: Language and Cognition, 4* (1), 105-122.

Ure, J. N. (1971). Lexical density and register differentiation. In G.E. Perren & J. L. M. Trim (Eds.), *Application of linguistics: Selected papers of the Second International Congress of Applied Linguistics* (pp. 443-452). Cambridge: Cambridge University Press.

Vermeer, A. (2000) .Coming to grips with lexical richness in spontaneous speech data. *Language Testing, 17* (1), 65-83.

Walters, J. (2004). Teaching the use of context to infer meaning: a longitudinal survey of L1 and L2 vocabulary research. *Language Teaching, 37*, 243-252.

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics, 28* (1), 46-65.

Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition, 30* (1), 79-95.

Weir, C. J. (1993). *Understanding and developing language tests*. London: Prentice Hall.

West, M. (1953). *A general service list of English words.* London: Longman.

Widdowson, H. G. (1978). *Teaching language as communication.* Oxford: Oxford University Press.

Wierzbicka, A. (1996). *Semantics. Primes and universals.* Oxford: Oxford University Press.

Wilks, C., Meara, P., & Wolter, B. (2008). A further note on simulating word association behaviour in a second language. *Second Language Research, 21* (4), 359–372.

Williams, E. (1994). Remarks on lexical knowledge. In L. Gleitman & B. Landau (Eds.), *The acquisition of the lexicon* (pp. 7-35). Cambridge, MA: The MIT Press.

Wong-Fillmore, L. (1991). Second-language learning in children: A model of language learning in social context. In E. Bialystok (Ed.), *Language processing in bilingual children* (pp. 49-69). Cambridge: Cambridge University Press.

Wode, H. (1999). Incidental vocabulary acquisition in the foreign language classroom. *Studies in Second Language Acquisition, 21* (2), 243-258.

Wray, A. (2002). *Formulaic language and the lexicon.* Cambridge: Cambridge University Press.

Wurm, L. H., Ernestus, M., Schreuder, R., & Baayen, R. H. (2006). Dynamics of the auditory comprehension of prefixed words: Cohort entropies and conditional root uniqueness points. *The mental lexicon*, *1 (*1), 125-146.

Xue, G., & Nation, I. S. P. (1984). A university word list. *Language Learning and Communication, 3*, 215-229.

Yu, G. (2009). Lexical diversity in writing and speaking task performances. *Applied Linguistics Advance Access* published on June 4, 2009. doi: 10.1093/applin/amp024

Zareva, A. (2005). Models of lexical knowledge assessment of second language learners of English at higher levels of language proficiency. *System, 33*, 547-562.

Zechmeister, E. B., D'Anna, C., Hall, J. W., Paus, C. H., & Smith, J. A. (1993). Metacognitive and other knowledge about the mental lexicon: Do we know how many words we know? *Applied Linguistics, 14 (*2), 188-206.

Zobl, H. (1995). Converging evidence for the 'acquisition-learning' distinction. *Applied Linguistics, 16* (1), 35-56.

Appendices

Appendix A. Vocabulary learning strategies (adapted from Schmitt, 1997, p.208)

Strategy group	Strategy
Strategies for the discovery of a new word`s meaning	
DET	Analyze part of speech
DET	Analyze affixes and roots
DET	Check for L1 cognate
DET	Analyze any available pictures or gestures
DET	Guess from textual context
DET	Bilingual dictionary
DET	Monolingual dictionary
DET	Word lists
DET	Flash cards
SOC	Ask teacher for an L1 translation of new word
SOC	Ask teacher for paraphrase or synonym of new word
SOC	Ask classmates for meaning
SOC	Discover new meanings through group work activity
Strategies for consolidating a word once it has been encountered	
SOC	Study and practice meaning in a group
SOC	Teacher checks students` flash cards or word lists for accuracy
SOC	Interact with native speakers
MEM	Study word with a pictorial representation of its meaning
MEM	Image word`s meaning
MEM	Connect word to a previous personal experience
MEM	Associate the word with its coordinates
MEM	Connect word to its synonyms and antonyms
MEM	Use semantic maps
MEM	Use 'scales' for gradable adjectives
MEM	Peg method

MEM	Loci method
MEM	Group words together to study them
MEM	Group words together spatially on a page
MEM	Use new word in sentences
MEM	Group words together within a storyline
MEM	Study the spelling of the word
MEM	Say new word aloud when studying
MEM	Image word form
MEM	Underline initial letter of the word
MEM	Use Keyword Method
MEM	Affixes and roots (remembering)
MEM	Part of speech (remembering)
MEM	Paraphrase the word`s meaning
MEM	Use cognates in study
MEM	Learn the word of an idiom together
MEM	Use physical action when learning a word
MEM	Use semantic feature grids
COG	Verbal repetition
COG	Written repetition
COG	Word lists
COG	Flash cards
COG	Take notes in class
COG	Use the vocabulary section in your textbook
COG	Listen to tape of word lists
COG	Put English labels on physical objects
COG	Keep a vocabulary notebook
MET	Use English-language media (songs, movies, newscasts, etc.)
MET	Test oneself with word tests
MET	Use spaced word practice
MET	Skip or pass new word
MET	Continue to study word over time

Appendix B. Four strands of vocabulary teaching (Nation, 2001, p. 390)

Strand	*General conditions*	*Vocabulary requirements*	*Activities and techniques*
Meaning-focused input	Focus on the message Some unfamiliar items Understanding Noticing	95%+ coverage (preferably 98%) Skill at guessing from context opportunity to negotiate Incidental defining and attention drawing	• Reading graded readers • Listening to stories • Communication activities
Meaning-focused output	• Focus on the message • Some unfamiliar items • Understanding • Noticing	• 95%+ coverage (preferably 98%) • Encouragement to use unfamiliar items • Supportive input	• Communication activities with written input • Prepared writing • Linked skills
Language-focused learning	• Focus on language items	• Skill in vocabulary learning strategies • Appropriate teacher focus n high-frequency words and strategies for low-frequency words	• Direct teaching of vocabulary • Direct learning • Intensive reading • Training in vocabulary strategies
Fluency development	• Focus on the message • Little or no unfamiliar language • Pressure to perform faster	• 99%+ coverage • Repetition	• Reading easy graded readers • Repeated reading • Speed reading • Listening to easy input • 4/3/2 speaking exercise • rehearsal tasks • 10-minute writing • Linked skills •

Appendix C. The vocabulary size test

List of words	I. I don't remember having seen this word before ✓	II. I have seen this word before, but I don't know what it means ✓	III. I have seen this word before, and I **think** it means: (synonym or translation)	IV. I **know** this word. It means: (synonym or translation)	V. I can use this word in a sentence. (Write a sentence.) If you do this section, please also do Section IV.
1. cool					
2. kitchen					
3. lead					
4. cow					
5. frog					
6. scent					
7. harsh					
8. ascertain					
9. sprig					
10. matron					
11. coronet					
12. jut					
13. amorpous					
14. bagpipe					
15. choleric					
16. crock					
17. incumbent					
18. offal					
19. untoward					
20. amphitrite					
21. carpel					
22. doss					
23. havelock					
24. nominative					

List of words	I. I don't remember having seen this word before ✓	II. I have seen this word before, but I don't know what it means ✓	III. I have seen this word before, and I **think** it means: (synonym or translation)	IV. I **know** this word. It means: (synonym or translation)	V. I can use this word in a sentence. (Write a sentence.) If you do this section, please also do Section IV.
25. pilotage					
26. serried					
27. blurb					
28. scriber					
29. appositive					
30. capybara					
31. directrix					
32. footage					
33. horseshit					
34. nighthawk					
35. ravioli					
36. aeroplankton					
37. tandoor					
38. cogito					
39. corvette					
40. chanterelle					
41. hyperthyroid					
42. pica					
43. immunoassay					
44. apertometer					
45. scandium					
46. gusli					
47. chuckie					
48. mendeleyevite					
49. matelasse					
50. slipper					

Appendix D. The post-test in Pilot study 3

List of words	I. I don't remember having seen this word before ✓	II. I have seen this word before, but I don't know what it means ✓	III. I have seen this word before, and I **think** it means: (synonym or translation)	IV. If the word is familiar to you, can you identify where you learnt it? (context)	V. I **know** this word. It means: (synonym or translation)	V. I can use this word in a sentence. (Write a sentence.) If you do this section, please also do Section V.
1. affidavit						
2. alleged						
3. bid						
4. caitiff						
5. comeuppance						
6. delicacy						
7. desperation						
8. dime-store						
9. dimly						
10. dungeon						
11. echelons						
12. efficacious						
13. elusive						
14. go bankrupt						
15. hard-liner						
16. hookah						

List of words	I. I don't remember having seen this word before ✓	II. I have seen this word before, but I don't know what it means ✓	III. I have seen this word before, and I **think** it means: (synonym or translation)	IV. If the word is familiar to you, can you identify where you learnt it? (context)	V. I **know** this word. It means: (synonym or translation)	V. I can use this word in a sentence. (Write a sentence.) If you do this section, please also do Section V.
17. immense						
18. inclination						
19. kaboosh						
20. limelight						
21. meritocracy						
22. miffed						
23. naturalized						
24. nightingale						
25. numb						
26. persecution						
27. pixel						
28. preview						
29. prospective						
30. rafting						
31. rave						
32. recurring						
33. repulsive						

List of words	I. I don't remember having seen this word before ✓	II. I have seen this word before, but I don't know what it means ✓	III. I have seen this word before, and I **think** it means: (synonym or translation)	IV. If the word is familiar to you, can you identify where you learnt it? (context)	V. I **know** this word. It means: (synonym or translation)	V. I can use this word in a sentence. (Write a sentence.) If you do this section, please also do Section V.
34. smack						
35. spooks						
36. surrender						
37. to confide						
38. to drown						
39. to flee						
40. to muse on						
41. to ponder						
42. to reveal						
43. to strive						
44. to thwart						
45. to trigger						
46. torrent						
48. ultimate						
49. wee						
50. yearn						

Appendix E. Course evaluation form for Pilot study 3

Course evaluation form

Please, finish the following sentences:

The course …..

The teacher.....

The novel.....

Vocab tests.....

I could have been motivated by.....

Thank you!

Version 1

In this section you must choose the right word to go with each meaning.   Write the letter of that word next to its meaning.   Here is an example.

A business
B clock   ___F___ part of a house
C horse   ___C___ animal with four legs
D pencil   ___D___ something used for writing
E shoe
F wall

1.
A benefit
B labor   _____ work
C percent   _____ part of 100
D principle   _____ general idea used to
E source   guide one's actions
F survey

2.
A element   _____ money for a special
B fund   purpose
C layer   _____ skilled way of doing
D philosophy   something
E proportion   _____ study of the meaning
F technique   of life

3.
A consent
B enforcement   _____ total
C investigation   _____ agreement or permission
D parameter   _____ trying to find information
E sum   about something
F trend

4.
A decade
B fee   _____ 10 years
C file   _____ subject of a discussion
D incidence   _____ money paid for services
E perspective
F topic

5.
A colleague
B erosion   _____ action against the law
C format   _____ wearing away gradually
D inclination   _____ shape or size of
E panel   something
F violation

6.
A achieve
B conceive   _____ change
C grant   _____ connect together
D link   _____ finish successfully
E modify
F offset

7.
A convert
B design   _____ keep out
C exclude   _____ stay alive
D facilitate   _____ change from one thing
E indicate   into another
F survive

8.
A anticipate
B compile   _____ control something
C convince   skillfully
D denote   _____ expect something will
E manipulate   happen
F publish   _____ produce books and
                     newspapers

9.
A equivalent
B financial   _____ most important
C forthcoming   _____ concerning sight
D primary   _____ concerning money
E random
F visual

10.
A alternative
B ambiguous   _____ last or most important
C empirical   _____ something different that
D ethnic   can be chosen
E mutual   _____ concerning people from
F ultimate   a certain nation

Version 2

In this section you must choose the right word to go with each meaning.   Write the letter of that word next to its meaning.   Here is an example.

A business
B clock   ___F___ part of a house
C horse   ___C___ animal with four legs
D pencil   ___D___ something used for writing
E shoe
F wall

1.
A area
B contract   _____ written agreement
C definition   _____ way of doing something
D evidence   _____ reason for believing
E method   something is or is not true
F role

2.
A debate
B exposure   _____ plan
C integration   _____ choice
D option   _____ joining something into a
E scheme   whole
F stability

3.
A access
B gender   _____ male or female
C implementation _____ study of the mind
D license   _____ entrance or way in
E orientation
F psychology

4.
A accumulation
B edition   _____ collecting things over time
C guarantee   _____ promise to repair a broken
D media   product
E motivation   _____ feeling a strong reason or
F phenomenon   need to do something

5.
A adult
B exploitation _____ end
C infrastructure _____ machine used to move
D schedule   people or goods
E termination   _____ list of things to do at
F vehicle   certain times

6.
A alter
B coincide   _____ change
C deny   _____ say something is
D devote   not true
E release   _____ describe clearly and
F specify   exactly

7.
A correspond
B diminish   _____ keep
C emerge   _____ match or be in
D highlight   agreement with
E invoke   _____ give special
F retain   attention to something

8.
A bond
B channel   _____ make smaller
C estimate   _____ guess the number or
D identify   size of something
E mediate   _____ recognizing and
F minimize   naming a person or
                      thing

9.
A explicit
B final   _____ last
C negative   _____ stiff
D professional _____ meaning 'no' or
E rigid   'not'
F sole

10.
A abstract
B adjacent   _____ next to
C controversial _____ added to
D global   _____ concerning the whole
E neutral   world
F supplementary

Version 1

In this section you must choose the right word to go with each meaning.   Write the letter of that word next to its meaning.   Here is an example.

A business

B clock   ____F___ part of a house
C horse   ____C___ animal with four legs
D pencil   ____D___ something used for writing
E shoe
F wall

1.
A colleague
B erosion   _____ action against the law
C format   _____ wearing away gradually
D inclination   _____ shape or size of
E panel   something
F violation

2.
A casualty
B flurry   _____ someone killed or
C froth   injured
D revelry   _____ being away from other
E rut   people
F seclusion   _____ noisy and happy
       celebration

3.
A anticipate
B compile   _____ control something
C convince   skillfully
D denote   _____ expect something will
E manipulate   happen
F publish   _____ produce books and
       newspapers

4.
A apparition
B botany   _____ ghost
C expulsion   _____ study of plants
D insolence   _____ small pool of water
E leash
F puddle

5.
A arsenal
B barracks   _____ happiness
C deacon   _____ difficult situation
D felicity   _____ minister in a church
E predicament
F spore

6.
A acquiesce
B bask   _____ to accept without protest
C crease   _____ sit or lie enjoying
D demolish   warmth
E overhaul   _____ make a fold on cloth or paper
F rape

7.
A blaspheme
B endorse   _____ slip or slide
C nurture   _____ give care and food to
D skid   _____ speak badly about God
E squint
F straggle

8.
A clinch
B jot   _____ move very fast
C mutilate   _____ injure or damage
D smolder   _____ burn slowly without flame
E topple
F whiz

9.
A auxiliary
B candid   _____ bad-tempered
C luscious   _____ full of self-importance
D morose   _____ helping, adding support
E pallid
F pompous

10.
A dubious
B impudent   _____ rude
C languid   _____ very ancient
D motley   _____ of many different kinds
E opaque
F primeval

Version 2

In this section you must choose the right word to go with each meaning.   Write the letter of that word next to its meaning.   Here is an example.

A business
B clock   ___F___ part of a house
C horse   ___C___ animal with four legs
D pencil   ___D___ something used for writing
E shoe
F wall

1.
A debate
B exposure   _____ plan
C integration   _____ choice
D option   _____ joining something into a
E scheme   whole
F stability

2.
A benevolence
B convoy   _____ kindness
C lien   _____ set of musical notes
D octave   _____ speed control for an
E stint   engine
F throttle

3.
A access
B gender   _____ male or female
C implementation   _____ study of the mind
D license   _____ entrance or way in
E orientation
F psychology

4.
A bourgeois
B brocade   _____ middle class people
C consonant   _____ row or level of something
D prelude   _____ cloth with a pattern or
E stupor   gold or silver threads
F tier

5.
A alcove
B impetus   _____ priest
C maggot   _____ release from prison early
D parole   _____ medicine to put on
E salve   wounds
F vicar

6.
A alkali
B banter   _____ light joking talk
C coop   _____ a rank of British nobility
D mosaic   _____ picture made of small
E stealth   pieces of glass or stone
F viscount

7.
A dissipate
B flaunt   _____ steal
C impede   _____ scatter or vanish
D loot   _____ twist the body about
E squirm
uncomfortably
F vie

8.
A contaminate
B cringe   _____ write carelessly
C immerse   _____ move back because of fear
D peek   _____ put something under water
E relay
F scrawl

9.
A illicit
B lewd   _____ immense
C mammoth   _____ against the law
D slick   _____ wanting revenge
E temporal
F vindictive

10.
A indolent
B nocturnal   _____ lazy
C obsolete   _____ no longer used
D torrid   _____ clever and tricky
E translucent
F wily

Appendix H. The test of lexis intuitively perceived useful in English studies (with key)

Version 1

In this section you must choose the right word to go with each meaning.   Write the letter of that word next to its meaning.   Here is an example.

A business
B clock   ___F___ part of a house
C horse   ___C___ animal with four legs
D pencil   ___D___ something used for writing
E shoe
F wall

1.
A. abandon
B. conceive   ___A__ leave
C. enhance   ___B__ understand
D. facilitate   ___D__ make easier
E. supplement
F. thrive

2.
A. animosity
B. consistency   ___C__ talk
C. discourse   ___A__ ill will
D. perception   ___E__ support
E. scaffold
F. verve

3.
A. capacious
B. diminishing   ___C__ same
C. identical   ___E__ continuing without
D. marginal   change
E. persistent   ___A_ containing a lot
F. valid

4.
A. arbitrary
B. diverse   ___F__ enough
C. intrinsic   ___E__ relating to the future
D. obtainable   ___A__ determined by chance
E. prospective
F. sufficient

5.
A. pursue
B. initiate   ___F__ get from
C. depict   ___E__ fall quickly
D. dub   ___C__ represent in drawing
E. plummet
F. derive

6.
A. eliminate
B. harness   __E___ support publicly
C. scan   __B___ bring under control
D. consume   __A___ remove completely
E. advocate
F. varnish

7.
A. semicolon
B. impact   ___C__ scheme to harm
C. vendetta   ___D__ using instead of
D. substitution   something else
E. narrative   ___F__ joining word
F. conjunction

8.
A. digest
B. connotation   ___E__ strong feeling of fear
C. collage   ___A__ concise publication
D. intrigue   ___C__ picture
E. dismay
F. nugget

9.
A. vast
B. ferocious   __E___ known for something bad
C. proficient   __F___ not enough
D. adjacent   __C___ showing skill
E. infamous
F. scarce

10.
A. illicit
B. indolent   __A___ not legal
C. mammoth   __B___ lazy
D. salient   __D___ most important
E. vindictive
F. wily

Version 2

In this section you must choose the right word to go with each meaning.   Write the letter of that word next to its meaning.   Here is an example.

A business
B clock   ___F___ part of a house
C horse   ___C___ animal with four legs
D pencil   ___D___ something used for writing
E shoe
F wall

1.
A. access
B. confer   __D___ give
C. improve   __F___ find by following signs
D. retain   __A___ reach
E. target
F. trace

2.
A. annals
B. clause
C. grant   __A___ chronicles
D. junction   __C___ money for studies
E. mature   __E___ fully developed
F. script

3.
A. appropriate
B. eternal   __C___ native
C. indigenous   __D___ responsible
D. liable   __E___ shared, common
E. mutual
F. persistent

4.
A. contemporary
B. crucial   __D___ based on observation
C. declining   __F___ tolerable
D. empirical   __B___ very important
E. ignorant
F. sustainable

5.
A. accumulate
B. construct   __D___ find direction
C. dispose   __F___ put in order
D. orient   __A___ bring together
E. reject
F. sequence

6.
A. obtain
B. adapt
C. reduce   __G___ get something
D. cultivate   __C___ make smaller
E. unveil   __B___ turn into a film
F. overhaul

7.
A. chase
B. commence
C. expose   __E___ decrease quickly
D. skim   __A___ run after
E. plunge   __D___ read through
F. allot

8.
A. abolish
B. yoke
C. inaugurate   __E___ support, encourage
D. devour   __B___ eat eagerly
E. promote   __C___ declare open
F. glaze

9.
A. clue
B. tinge
C. hodgepodge   __F___ a small amount of
D. stratagem   something
E. apprehension __C___ confused mixture
F. lump   __B___ a solid piece of
something

10.
A. colossal
B. unruly
C. adroit   __C___ next to each other
D. contiguous   __D___ quick in behavior
E. distinct   __B___ difficult to control
F. persuasive

Appendix I. Regression tables for vocabulary and general language proficiency on Version 1

Descriptive Statistics

	Mean	Std. Deviation	N
Filter ovarall score	,6293	,13195	61
Beyond 2000	11,2672	3,71742	61
Receptive vocab	,5809	,18403	61

Correlations

		Filter ovarall score	Beyond 2000	Receptive vocab
Pearson Correlation	Filter overall score	1,000	,181	,718
	Beyond 2000	,181	1,000	,277
	Receptive vocab	,718	,277	1,000
Sig. (1-tailed)	Filter ovarall score	.	,081	,000
	Beyond 2000	,081	.	,015
	Receptive vocab	,000	,015	.
N	Filter ovarall score	61	61	61
	Beyond 2000	61	61	61
	Receptive vocab	61	61	61

Variables Entered/Removed

Model	Variables Entered	Variables Removed	Method
1	Receptive vocab, Beyond 2000[a]	.	Enter

a. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,718[a]	,516	,499	,09338

a. Predictors: (Constant), Receptive vocab, Beyond 2000

Model Summary

Model	Change Statistics				
	R Square Change	F Change	df1	df2	Sig. F Change
1	,516	30,908	2	58	,000

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	,539	2	,269	30,908	,000[a]
	Residual	,506	58	,009		
	Total	1,045	60			

a. Predictors: (Constant), Receptive vocab, Beyond 2000

b. Dependent Variable: Filter overall score

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	,336	,048		6,969	,000
	Beyond 2000	,000	,003	-,020	-,207	,837
	Receptive vocab	,519	,068	,723	7,609	,000

a. Dependent Variable: Filter overall score

Appendix J. Regression tables for vocabulary and general language proficiency on Version 2

Descriptive Statistics

	Mean	Std. Deviation	N
Filter overall	,5110	,09066	74
Beyond 2000	10,6392	3,81345	74
Receptive vocab	,5149	,16303	74

Correlations

		Filter overall	Beyond 2000	Receptive vocab
Pearson Correlation	Filter overall	1,000	,204	,400
	Beyond 2000	,204	1,000	,239
	Receptive vocab	,400	,239	1,000
Sig. (1-tailed)	Filter overall	.	,041	,000
	Beyond 2000	,041	.	,020
	Receptive vocab	,000	,020	.
N	Filter overall	74	74	74
	Beyond 2000	74	74	74
	Receptive vocab	74	74	74

Variables Entered/Removed

Model	Variables Entered	Variables Removed	Method
1	Receptive vocab, Beyond 2000[a]	.	Enter

a. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,415[a]	,172	,149	,08364

a. Predictors: (Constant), Receptive vocab, Beyond 2000

Model Summary

	Change Statistics				
Model	R Square Change	F Change	df1	df2	Sig. F Change
1	,172	7,391	2	71	,001

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	,103	2	,052	7,391	,001[a]
	Residual	,497	71	,007		
	Total	,600	73			

a. Predictors: (Constant), Receptive vocab, Beyond 2000

b. Dependent Variable: Filter overall

Appendix K. Low-frequency word types appearing six times or more in the CORES corpus

TYPE	RANGE	FREQ	AL	L	LC
SEMANTIC	1	93	0	93	0
GRAMMATICAL	2	78	2	76	0
PHONOLOGICAL	1	50	0	50	0
PHRASE	3	50	4	44	2
NOVEL	3	47	2	20	25
REANALYSIS	1	47	0	47	0
LINGUISTIC	3	39	10	28	1
BILINGUAL	2	38	2	36	0
NEGRO	1	37	0	0	37
CATHOLIC	1	35	0	0	35
COGNITIVE	2	33	3	30	0
NATIONALISM	1	32	0	0	32
VICTORIAN	1	30	0	0	30
LEXICAL	1	29	0	29	0
SYNTAX	3	29	1	27	1
SUFFIX	1	27	0	27	0
INTEGRATIVENESS	1	26	26	0	0
PHONOLOGY	2	26	3	23	0
PRAGMATIC	3	26	1	24	1
SEMANTICS	1	24	0	24	0
AESTHETIC	1	23	0	0	23
FICTION	2	23	1	0	22
INFLECTION	1	23	0	23	0
METAPHOR	3	23	10	5	8
NARRATIVE	3	22	6	1	15
PROTESTANT	1	21	0	0	21
REALISM	2	21	1	0	20
SYNTACTIC	1	21	0	21	0
CLASSROOM	2	20	19	0	1
CONSONANT	1	20	0	20	0
REPERTOIRE	3	19	11	7	1
TALMY	1	19	0	19	0
DEFAULT	1	18	0	18	0
LEXICON	2	18	1	17	0
REVIVAL	1	18	0	0	18
CRITICISM	2	17	2	0	15
INTEGRATIVE	2	16	15	0	1
NEGROES	1	16	0	0	16
SCORES	1	16	16	0	0
VOCABULARY	3	16	6	9	1
DISCOURSE	3	15	1	2	12
GRAMMATICALIZATION	1	15	0	15	0
IMPRESSION	3	15	2	4	9
INFLECTIONAL	1	15	0	15	0
NATIVISM	1	14	0	0	14
POSTMODERN	1	14	0	0	14
ASSOCIATOR	1	13	0	13	0
CIVIC	1	13	0	0	13
ESSAY	2	13	1	0	12

INNATE	1	13	0	13	0
TELEVISION	3	13	4	3	6
TEMPLATES	1	13	0	13	0
TORY	1	13	0	13	0
UTTERANCE	2	13	0	11	2
VAST	3	13	2	7	4
ASSIMILATION	2	12	0	6	6
BILINGUALISM	1	12	0	12	0
CELTIC	3	12	1	1	10
DIALECT	2	12	10	2	0
ERA	2	12	2	0	10
FICTIVE	2	12	0	11	1
FLUENT	2	12	4	8	0
MOTIVATIONAL	1	12	12	0	0
NASAL	1	12	0	12	0
REFERENCE	3	12	4	5	3
ADVERTISERS	2	11	3	0	8
ANIMATION	1	11	11	0	0
APPEAL	3	11	1	3	7
BREAKED	1	11	0	11	0
BUDGET	2	11	1	0	10
CAMPAIGN	2	11	2	0	9
CHAOS	3	11	7	1	3
CLIMATE	2	11	10	0	1
CONCISE	2	11	7	4	0
DEMOCRATIC	2	11	3	0	8
DIALECTS	2	11	7	4	0
EMISSIONS	1	11	11	0	0
EMOTIONAL	3	11	1	1	9
FANTASTIC	1	11	0	0	11
FICTIONAL	1	11	0	0	11
LITERAL	2	11	0	8	3
MONSTER	3	11	1	8	2
NEGATION	1	11	0	11	0
SUFFIXES	1	11	0	11	0
UTTERANCES	1	11	0	11	0
VICTORIANISM	1	11	0	0	11
ANTECEDENT	2	10	4	6	0
CONGRESS	2	10	6	0	4
CONSERVATIVE	3	10	3	4	3
CONVEY	2	10	1	9	0
DEMOCRATS	2	10	3	0	7
DERIVATIONAL	1	10	0	10	0
DESPERATION	2	10	6	0	4
INFANTS	2	10	0	8	2
INTERESTINGLY	2	10	1	9	0
ONTOLOGICAL	1	10	0	0	10
PROMINENT	3	10	4	2	4
PROSODY	1	10	0	10	0
STANCE	2	10	0	9	1
STATIC	2	10	2	8	0
SUBSET	1	10	0	10	0
COASTAL	1	9	9	0	0
CORRELATION	2	9	1	8	0
DEMOCRACY	1	9	0	0	9
ENTITLED	2	9	1	0	8

GENERALIZATION	2	9	1	8	0
GENRE	2	9	2	0	7
GLANCE	2	9	0	6	3
GREENHOUSE	1	9	9	0	0
IRREGULARS	1	9	0	9	0
LINEAR	3	9	2	4	3
LINGUISTS	2	9	1	8	0
LITERALLY	3	9	1	5	3
OVERLOAD	1	9	0	9	0
PROPOSITION	3	9	1	4	4
PROTESTANTISM	1	9	0	0	9
RECALL	3	9	2	6	1
SENATE	2	9	7	0	2
SETTINGS	2	9	3	6	0
TEMPLATE	1	9	0	9	0
VICE-VERSA	1	9	0	9	0
AGENDA	3	8	2	1	5
BOURGEOIS	1	8	0	0	8
CAPTURE	3	8	1	6	1
CLUSTERS	2	8	1	7	0
CONNECTIONIST	1	8	0	8	0
DUAL	1	8	0	8	0
EASTER	1	8	0	0	8
ELITE	2	8	0	1	7
ENTHUSIASM	2	8	4	0	4
FEEDBACK	1	8	0	8	0
FLU	1	8	0	8	0
HUMANITY	2	8	2	0	6
IMPAIRED	2	8	1	7	0
INFERNAL	1	8	0	0	8
MORPHEME	1	8	0	8	0
ONSET	1	8	0	8	0
ORGANIC	2	8	4	0	4
PLAYWRIGHTS	1	8	0	0	8
PROSE	3	8	1	1	6
PROSODIC	1	8	0	8	0
PROTAGONIST	2	8	0	1	7
RETRIEVAL	1	8	0	8	0
SCHOLARS	3	8	1	2	5
SEGMENTS	2	8	0	7	1
SOPHISTICATED	3	8	3	4	1
VELAR	1	8	0	8	0
APTITUDE	1	7	7	0	0
ASSERTION	2	7	0	3	4
BAKERY	1	7	0	7	0
CARBON	1	7	7	0	0
CAREER	2	7	2	0	5
COMPREHENSION	3	7	3	3	1
CORRELATIONS	2	7	1	6	0
CRISIS	1	7	0	0	7
EMBEDDED	1	7	0	7	0
GENERALIZE	1	7	0	7	0
GENUINE	3	7	2	1	4
HEADLESS	2	7	0	6	1
INHERITED	2	7	0	5	2
INSISTENCE	2	7	1	0	6

INTACT	2	7	2	5	0
MASSIVE	3	7	2	3	2
MELODY	2	7	3	4	0
MOVIE	3	7	2	1	4
PHONETICS	1	7	0	7	0
PREDICATE	2	7	3	4	0
PROFESSOR	3	7	4	1	2
PROMINENCE	2	7	3	0	4
REALIST	1	7	0	0	7
REFORM	2	7	1	0	6
REFUGE	1	7	0	0	7
RURAL	2	7	0	1	6
SALIENT	2	7	1	6	0
SECULAR	1	7	0	0	7
SELF-REFLEXIVE	1	7	0	0	7
SINGULAR	2	7	0	6	1
STEM-CELL	1	7	7	0	0
TONE	3	7	2	1	4
USAGE	2	7	1	6	0
VERBAL	2	7	1	6	0
VICTORIANS	1	7	0	0	7
VOCAL	1	7	0	7	0
ANIMATED	2	6	5	1	0
BOTHERS	1	6	0	6	0
COMBINATORIAL	1	6	0	6	0
DELIBERATELY	3	6	2	1	3
DIARY	2	6	0	5	1
ENTHUSIASTIC	2	6	2	0	4
FANTASY	2	6	3	0	3
FASCINATING	3	6	2	1	3
FEEDING	1	6	6	0	0
FREQUENCIES	1	6	0	6	0
GENERALIZED	2	6	0	3	3
GENERATIVE	2	6	1	5	0
GEOGRAPHICAL	1	6	6	0	0
GERM	1	6	0	6	0
HEIGHTENED	2	6	1	0	5
HINT	2	6	0	1	5
INFERIORITY	1	6	0	0	6
JUDICIAL	2	6	5	0	1
LEGITIMATE	3	6	3	1	2
MASCULINE	2	6	0	5	1
MONOLINGUAL	1	6	0	6	0
MONSTROUS	1	6	0	0	6
MORPHOLOGY	1	6	0	6	0
NODES	1	6	0	6	0
NOVELS	1	6	0	0	6
OPPOSITION	2	6	0	1	5
PLURALISM	1	6	0	0	6
PROFILE	3	6	4	1	1
QUIRY	1	6	0	6	0
RHETORIC	1	6	0	0	6
SATELLITE	2	6	2	4	0
SENTIMENT	2	6	1	0	5
SIBILANT	1	6	0	6	0
STRIVE	2	6	1	0	5

Appendix L. The CORES-based vocabulary tests (with key)

Version 1

In this section you must choose the right word to go with each meaning. Write the letter of that word next to its meaning. Here is an example.

A business
B clock   ___F___ part of a house
C horse   ___C___ animal with four legs
D pencil   ___D___ something used for writing
E shoe
F wall

1.
A concept
B edition   __C___ effect
C impact   __D___ aim to achieve
D objective   __A___ idea or notion
E pursuit
F subsidy

2.
A comprise
B devote   __F___ finish
C estimate   __D___ suggest
D imply   __A___ include
E retain
F terminate

3.
A appropriate
B contemporary   __F___ following in order
C debilitating   __A___ good or suitable
D fundamental   __E___ dark and pessimistic
E gloomy
F successive

4.
A advent
B chivalry   __D___ period of time
C descendant   __A___ coming or arrival
D era   __E___ danger
E menace
F pace

5.
A assault
B conviction   __A___ violent attack
C eloquence   __D___ a part broken off
D fragment   __E___ a brief look at
E glimpse
F preface

6.
A allot
B conceal   __B___ hide
C denounce   __E___ cause to begin
D evoke   __D___ bring to mind
E initiate
F redeem

7.
A apprentice
B breakthrough   __D___ goods for sale
C doctrine   __E___ a mania
D merchandise   __C___ a principle
E obsession
F proliferation

8.
A adversary
B congenial   __F___ producing profit
C despised   __B___ nice and sociable
D elusive   __D___ difficult to catch
E feasible
F lucrative

9.
A amplify
B contemplate   __F___ assume as true or
C designate   necessary
D endow   __E___ spread and grow fast
E mushroom   __A___ make stronger
F postulate

10.
A alleged
B civic   __E___ easy to carry
C haunted   __C___ visited by ghosts
D impending   __A___ questionably true
E portable
F subtle

Version 2

In this section you must choose the right word to go with each meaning.   Write the letter of that word next to its meaning.   Here is an example.

A business
B clock   ___F___ part of a house
C horse   ___C___ animal with four legs
D pencil   ___D___ something used for writing
E shoe
F wall

1.
A   approach
B   controversy   e_____ aim to reach
C   exposure   d_____ motivation
D   incentive   b_____ quarrel
E   target
F   utility

2.
A   demonstrate
B   indicate   c_____ find a place
C   locate   e_____ keep up
D   obtain   f_____ break the rule
E   sustain
F   violate

3.
A   arbitrary
B   consistent   d_____ most important
C   flawless   e_____ coming after
D   primary   something
E   subsequent   b_____ unchanging
F   voluntary

4.
A   adherence
B   candidate   c_____ disadvantage
C   drawback   e_____ calmness
D   salvation   a_____ attachment or
E   tranquillity   faithfulness
F   upheaval

5.
A   allude
B   concede   f_____ gradually open
C   disrupt   a_____ indirectly refer to
D   retreat   e_____ mix together
E   scramble
F   unfold

6.
A   articulated
B   conspicuous   d_____ considering others'
C   salient   feelings
D   tactful   e_____ mysterious or
E   uncanny   supernatural
F   whimsical   f_____ full of playful humour

7.
A   assertion
B   clergy   c_____ property from
C   heritage   ancestors
D   specimen   f_____ clever type of humour
E   testimony   b_____ collective name for
F   wit   priests

8.
A   appeal
B   condemn   e_____ become better than
C   depict   expected
D   revive   d_____ bring back to life or
E   surmount   mind
F   yearn   f_____ desire strongly

9.
A   indispensable
B   mandatory   a_____ essential
C   stunning   c_____ beautiful
D   tormented   b_____ compulsory
E   vernacular
F   wane

10.
A   apprehension
B   gratitude   c_____ introduction
C   preface   f_____ feature
D   resemblance   d_____image
E   scholar
F   trait

Doktori (PhD) értekezés tézisei

ASSESSING ENGLISH MAJORS' VOCABULARY
AT THE UNIVERSITY OF PÉCS

WORKING TOWARDS A CORPUS-BASED APPROACH

ANGOL SZAKOSOK SZÓKINCSÉNEK KORPUSZ-ALAPÚ ESZKÖZÖKKEL
TÖRTÉN MÉRÉSE A PÉCSI TUDOMÁNYEGYETEMEN

Lehmann Magdolna

Témavezet : Dr. Andor József

PÉCSI TUDOMÁNYEGYETEM
Nyelvtudományi Doktori Iskola
Alkalmazott Nyelvészeti Doktori Program

2009

Az értekezés témája és céljai

A fehér papíron rejtelmesen tekergő, kacskaringós fekete betűk látványa már az első találkozások alkalmával lenyűgözött. Emlékszem, szinte könyörögtem a szüleimnek, árulják el, melyik jel mit jelent pontosan és hogyan lesznek belőlük szavak, ők pedig vonakodtak, mert akkoriban az elsős tanító nénik nemigen kedvelték az iskolában azokat a gyerekeket, akik már tudtak olvasni. De én nem tágítottam, mert éreztem, ha sikerül megtudnom a betűk értelmét, egy új világ tárul majd fel előttem, és én alig vártam, hogy felfedezhessem. Azóta kiderült számomra, hogy a szavak nem csupán összefűzött betűk és ma már úgy tűnik, minél többet tudok meg róluk, annál lehetetlenebbnek látszik, hogy tökéletesen megértsem az általuk megnyitott kapuk mögött rejlő világok minden lehetséges dimenzióját.

Nyelvészként és egyetemi oktatóként a szavak világával foglalkozom és az egyik ilyen mikro-világ, mely megragadta érdeklődésemet, az angol szakos hallgatók szókincse. Saját és kollégáim tapasztalatai alapján az elsőéves hallgatók igen nagy nehézségekkel küzdenek a kurzusaikon számukra kötelező olvasmányok megértésével és írásfeladataik elkészítésével. Korábbi kutatások rámutattak, hogy e jelenség hátterében szókincsbeli hiányosságok állhatnak. Kutatásaim tehát arra irányultak, hogy felmérjem, hány angol szót ismernek az első és másodéves, nyelvi alapvizsga előtt álló angol szakos hallgatók a Pécsi Tudományegyetemen, vagyis mekkora szókinccsel rendelkeznek; ez a szókincs elegendő-e számukra ahhoz, hogy tanulmányaikat sikeresen végezhessék; illetve hogy ismerik-e az egyetemi angol nyelvű szövegekben leggyakrabban előforduló, s ezáltal számukra leghasznosabbnak tekinthető szavakat, és milyen jól ismerik azokat. E kérdések köré csoportosulnak az értekezésben bemutatott empirikus vizsgálatok, meghatározva a dolgozat vázát. A kutatási kérdések megválaszolása előtt azonban számos egyéb kérdést kellett tisztázni, így például azt, hogy mi is pontosan a szó és a szókincs, mit értünk azon, hogy ismerünk egy szót, ez hogyan mérhető, és milyen szerepet játszik a szavak ismerete az olvasott szöveg megértésében, illetve az önálló szövegalkotásban.

A szó fogalmának meghatározása nem egységes a nyelvészeti szakirodalomban, attól függ, mennyire absztrakt módon, morfológiai, szintaktikai, szemantikai vagy pragmatikai szemszögb l közelítjük meg a kérdést (lásd Bloomfield, 1933; Caramazza, 1999; Caron, 1992; Carter, 1998; Kempen & Huijbers, 1983; Martsa, 2007; Nattinger & DeCarrico, 1992; Pawley & Syder, 1983; Singleton, 1999; Wray, 2002). Szókincsen azonban általában az adott nyelvben, szövegben, az anyanyelvi vagy nem-anyanyelvi beszél k fejében (mentális lexikonjában) létez összes szót értjük (Carter, 1998; Ellis, 1994; Gleitman és Landau, 1994; Singleton, 2000). A lexikai kompetencia felépítése Henriksen (1999) szerint háromréteg és több, egymással kapcsolatban álló egyenesen ábrázolható. Ezek egyike a részleges tudástól a precíz jelentés ismeretig terjed, egy másikon a szóismeret mélysége jeleníthet meg a teljesen ismeretlent l a pontosan ismert jelentésig, míg egy harmadik ábrázolásmód a receptív (szófelismerés) és a produktív (szóhasználat) végpontok között helyezi el a lexikális tudás mértékét.

Az adott nyelvtanuló vagy anyanyelvi beszél szókincse adott id ben mindig az egyenesek más-más szakaszán található, tehát a lexikai tudás egyáltalán nem tekinthet statikus állapotnak, hanem egy dinamikusan változó jelenségként értelmezhet . Paribakht és Wesche (1999) úgy vélekedik, hogy a szókincs elsajátítás folyamata az adott szóval történ els jelentéssel bíró találkozással kezd dik és apró lépésekkel el rehaladva jut el addig, hogy a szó, jellegzetes elemeinek, vonásainak megismerésével, beépülhessen a mentális lexikonba. Tanulmányában Laufer (1998) egy adott szó ismeretének három komponensét vizsgálja: (1) alap receptív (passzív) tudás, mely a szó leggyakoribb és központi jelentésének ismerete (Laufer, 1998, p. 257); (2) kontrollált produktív ismeret, mely szerint a nyelvhasználó képes egy adott feladat által sugalmazott szavak el hívására és alkalmazására, például egy adott mondatba beírni egy olyan szót, melynek els néhány bet je van csak megadva; a harmadik pedig a (3) szabad produktív ismeret, vagyis amikor a nyelvhasználó képes saját akarata és választása szerint alkalmazni az adott szót, például egy írásbeli fogalmazási feladatban. Eredményei azt mutatják, hogy a három tudástípus nem azonos mértékben fejl dött az egy tanévet felölel kísérlet során, tehát valóban meg kell különböztetni egymástól a receptív és produktív lexikális ismeretet. A receptív

szókincs a tapasztalatok szerint mindig nagyobb a produktív szókincsnél (Webb, 2008).

A nyelvtanárok és nyelvtanulók körében leggyakrabban felmerül kérdések többnyire arra irányulnak, hány szóra és pontosan milyen szavakra van szükségük a nyelvtanulóknak. Nation és Waring (1997) rámutat arra, hogy egy nyelvtanuló számára nem minden szó lehet egyformán hasznos, és hogy mi számít hasznos vagy kevésbé hasznos szónak egy adott nyelvben, azt használatának, el fordulásának gyakorisága határozhatja meg. A tanítás és tanulás folyamatának el segítésére a nyelvészek számos gyakorisági szólistát hoztak létre nagy szöveggy jtemények, korpuszok alapján. Közülük az egyik legrégebbi, de ma is legtöbbet hivatkozott lista a *General Service List* (GSL), melyet Michael West dolgozott ki 1953-ban. A GSL az angol nyelvben el forduló 2.000 leggyakoribb szót tartalmazza egy 5 millió szavas írott korpusz alapján, és Nation és Waring (1997) szerint ez tekinthet a nyelvtanárok és kurrikulum tervez k számára a leggyakoribb angol szavak klasszikus gy jteményének.

A GSL mellett léteznek általános és speciális tanulmányi célokra készült listák is, mint a *University Word List* (UWL, egyetemi szólista) (Read, 2000, p. 159), valamint az ennél újabb kelet , Coxhead által 2000-ben kidolgozott, 570 szócsaládot tartalmazó *Academic Word List* (AWL), melyet egyetemeken széles körben el forduló szövegekben leggyakoribb szavak alapján állítottak össze. A szakirodalom szerint azok a szavak, melyek e listák egyikében sem találhatók meg, ritka (low-frequency) szavaknak tekinthet k az angol nyelvben és Nation és Waring (1997) szerint a gyors lexikai fejl dés érdekében a nyelvtanulóknak el ször a kétezer leggyakoribb szót kellene elsajátítani.

Az anyanyelv nyelvhasználó és a nyelvtanuló szókincse közötti legszembet n bb különbség az általuk ismert szavak száma, vagyis szókincsük nagysága. Goulden, Nation és Read (1990) vizsgálata alapján egy egyetemet végzett, angol anyanyelv nyelvhasználó átlagosan 17.000 alapszót ismer. Az olvasás során fejl dik a szókincs (Ellis 1994, Krashen 1989), ezt a folyamatot nevezzük saját fordításomban *járulékos* szótanulásnak [*incidental vocabualry learning*], hiszen az új szavak illetve jelentések megismerése ez esetben az olvasás kísér jelensége. Számos

vizsgálat próbálkozott a folyamat mechanizmusának megvilágításával (Rott, 1999; Paribakht és Wesche, 1999; Wode, 1999; Brown és munkatársai, 1999; Ellis és He, 1999), ám a terminus használata még nem egységes (Huckin és Coady, 1999). Nation (2006) szerint 5-6, míg Webb (2007) szerint tíznél is többször kell találkozni a szövegben egy adott szóval ahhoz, hogy jelentését az olvasás során elsajátítsuk.

Az olvasott szövegértés és a szókincs közötti kapcsolatot sokan vizsgálták (Alderson, 2000; Gass, 1999; Grabe & Stoller, 1997; Hulstijn, Hollander & Greidanus, 1996; Laufer, 1997; Nation, 2006, Paribakht & Wesche, 1997; Quian, 2002; Quian & Schedl, 2004; Zareva, 2005). A szövegek megértéséhez szükséges szókincs nagysága sokat vitatott kérdés. Carter (1998) szerint a GSL els kétezer szava 80 százalékát fedi le bármely vizsgált szövegnek, az els 5.000 leggyakoribb szó ismerete Hirsch és Nation (1992) szerint 90-95%-os lefedettséget biztosít, míg Nation (2001) szerint a GSL és az AWL szavainak ismerete a beszélt nyelvi szókincs 92%-át, a magazinok szókincsének pedig 84%-át fedi le. Egy olvasott szöveg megértéséhez a benne el forduló szavak 95-98%-át kell ismerni (Hirsch & Nation, 1992; Laufer, 2005a, 2005b; Nation & Waring, 1997), tehát minden ötvenedik szó lehet ismeretlen csupán (Schmitt, 2008, p. 330). Hazenberg és Hulstijn (1996) az angolt idegen nyelvként tanuló holland egyetemi hallgatók számára határozta meg azt a minimális szókincset, mely lehet vé teszi az angol nyelv egyetemi tanulmányokhoz szükséges anyagok ért olvasását. Eredményeik alapján ez 10.000 angol alapszó ismeretét jelenti. Korábbi hasonló holland tanulmányok szerint 3.000 és 5.000 alapszó passzív ismerete már elegend lenne. Nation (2006) számításai szerint a British National Corpus alapján a nyelvtanulóknak 8.000-9.000 szócsaládot kell ismerniük ahhoz, hogy autentikus szövegeket tudjanak olvasni, korábban Laufer (1992) 5.000 szócsalád ismeretét javasolta.

A szóismeret mérésének legfontosabb kérdései, hogy miért mérjük, mely szavak ismeretét mérjük, a szóismeret melyik aspektusát akarjuk mérni, és hogyan mérjük azt. A szókincs mérésére kvantitatív és kvalitatív eszközök is rendelkezésre állnak. A kvalitatív megközelítés lehet séget nyújt a szóismeret mélységének pontos meghatározására, hátránya azonban, hogy id igényes és ezért korlátozott mintán alkalmazható. A kvantitatív megközelítés lehet vé teszi, hogy rövid id alatt

nagyszámú szó ismeretér l alkossunk képet, de nem tudjuk a szóismeret teljes spektrumát vizsgálni. Veszteségek tehát mindkét módszer alkalmazása során fellépnek, ezért mindig az adott kontextus függvénye az, hogy melyik a célravezet bb. Kvalitatív mér eszközök a személyes interjúk (Verhallen & Schoonen, 1993, in Read, 2000, p. 94), a szóasszociáción alapuló tesztek (Read, 2000), illetve a széles körben alkalmazott *Vocabulary Knowledge Scale* (saját fordításomban *szóismereti skála*, Paribakht és Wesche, 1997). A kvantitatív mér eszközök között a receptív szókincs nagyságát méri Goulden, Nation és Read (1990) tesztje, míg az elterjedt *Vocabulary Levels Test* (VLT, Schmitt, Schmitt & Clapham, 2001) öt szógyakorisági szinten határozza meg a receptív szókincs nagyságát.

A produktív szókincs meghatározásakor kvantitatív módon közelítve egy szöveghez megállapítható, hány szóból áll összesen, ez a 'szövegszó' ('token') érték, míg a szövegben el forduló különböz szavak számát 'szóalak' értéknek (type) nevezzük. A kett hányadosát (T/t) gyakran alkalmazzák a szöveg lexikai változatosságának leírására. Mivel azonban e hányados nagymértékben függ a szövegek hosszától, nem tekinthet elég megbízhatónak, ezért helyette számos matematikai formulát dolgoztak ki e hatás kiküszöbölésére (Guiraud R-, Herdan C-, Zipf Z-, vagy Uber U-index). Megoldásként Laufer és Nation (1995) az írás lexikai min ségének vizsgálatára új, objektívebb mér eszközt alkotott *Lexical Frequency Profile* (LFP) (Szógyakorisági Profil- saját fordításomban) néven. Elméletük szerint a nyelvtanuló szókincsének nagysága jól tükröz dik írásképességében, a szógyakorisági profil pedig megmutatja, milyen arányban fordulnak el az írott szövegben a különböz szógyakorisági szinteknek megfelel szavak (Laufer és Nation, 1995, p. 311).

Az LFP legújabb (Paul Nation honlapjáról szabadon letölthet) változata a RANGE nev szoftver, melynek segítségével könnyen elvégezhet az íráselemzés. A program továbbfejlesztett változata, és sok egyéb hasznos alkalmazás található *Compleat Lexical Tutor* néven Tom Cobb (2000) honlapján is. Mivel azonban jelenleg nem létezik olyan eszköz, amely a szó ismeretének receptív és produktív aspektusát egyetlen tesztformátumban lenne képes mérni, ezért Laufer és munkatársai (2004) több kvantitatív teszttípus egyidej alkalmazását javasolják.

A kutatás ismertetése és az értekezés felépítése

Vizsgálataim megtervezése során a fent említett kutatási eredményekre támaszkodtam. Kutatási kérdéseim megválaszolása érdekében az értelezés két f részre és hét fejezetre tagolódik, amint az 1. számú táblázatban látható. Az els rész három fejezete átfogó képet nyújt a második részben bemutatott vizsgálatok elméleti hátterér l és a szókincskutatás három alapvet kérdését járja körül: a szókincs leírását, elsajátítását és mérését. Az els fejezet betekintést nyújt az olvasó számára a szókincskutatás terminológiájába, meghatározza a szókincs nyelvi kompetenciában betöltött helyét, tárgyalja a szó és a szóismeret értelmezési lehet ségeit, és bemutatja a mentális lexikon felépítésének, a szavak tárolásának és el hívásának f bb modelljeit. A második fejezet a szókincs elsajátítás szakirodalmát tekinti át, érinti az els - és második nyelvi szókincs elsajátítási mechanizmusának hasonlóságait és különbségeit, és szót ejt a sokat vitatott tudatos szótanulás és az olvasás során fellép járulékos szóelsajátítás lehetséges el nyeir l és hátrányairól. A harmadik fejezet fókuszában a szókincs mérése áll. A fejezet bemutatja a szókincs értékelésének elméleti hátterét, módszereit és a jelenleg elérhet legelterjedtebb, illetve az empirikus vizsgálatok során alkalmazott szókincsmér eszközöket.

Az értekezés els részében áttekintett kérdések kiindulópontként szolgáltak a második részben tárgyalt empirikus vizsgálatokhoz, melyeket a Pécsi Tudományegyetem Angol Alkalmazott Nyelvészeti Tanszékén folytattam. A kutatás célja az els éves angol szakos hallgatók szókincsének vizsgálata annak fényében, milyen lexikai követelményeket támaszt számukra a tanterv, kötelez olvasmányaik és írásfeladataik. A negyedik fejezet ismerteti az empirikus vizsgálatok kontextusát, bemutatja a résztvev ket, az angol szakos hallgatók tantervét és a tanszéken folytatott szókincsmérések el zményeit és hátterét, valamint rövid módszertani áttekintést nyújt a vizsgálatokról. Az ezt követ empirikus fejezetek felépítése az els éves angol szakos hallgatók szókincsének feltérképezésére és értékelésére kidolgozott szókincsmér eszköz kifejlesztésének fázisait tükrözi.

1. sz. táblázat: Az értekezés felépítése

I. A szókincskutatás a második nyelv elsajátítás területén

1. fejezet. Mi a szókincs?
- A szókincskutatás terminológiája
- A szókincs és a nyelvi kompetencia
- A szóismeret összetev i
- A mentális lexikon
2. fejezet: Hogyan sajátítható el a szókincs?
- A szókincs elsajátítás menetét befolyásoló tényez k
- A szókincs elsajátítás mélységi és mennyiségi aspektusai
- A szókincs elsajátítás és az olvasás kapcsolata
- A tudatos szótanulás és pedagógiai vonatkozásai
3. fejezet: Hogyan mérhet a szókincs?
- A szóismeret globális mérésének lehet ségei
- Kvalitatív mér eszközök
- Korpusz-alapú szólisták
- Kvantitatív mér eszközök
- Az értekezésben alkalmazott mér eszközök bemutatása

II. Empirikus tanulmányok az angol szakos hallgatók szókincsének és lexikális szükségleteinek területén

4. fejezet: Háttér a kutatásokhoz
- Kutatás kontextusa
- Résztvev k ismertetése
- Kutatás módszertani ismertetése
5. fejezet: Hány szót? El tanulmányok az angol szakosok receptív szókincsének feltárására
  A. Az eljárások
  B. Eredmények
  C. Az eredmények tárgyalása
6. fejezet: Milyen jól? Szókincsmérések els éves angol szakos hallgatókkal
- Az eljárások
- Eredmények
- Az eredmények tárgyalása
7. fejezet: Mely szavakat? Korpusz-alapú szókincsmér eszköz kifejlesztése
- A korpusz bemutatása
- A mér eszköz kifejlesztése
- Mérési eredmények
- Az eredmények tárgyalása

Konklúzió és jöv beni kutatási irányok
- A tanulmányok összefoglalása
- A vizsgálatok korlátai
- Korpusznyelvészeti, pedagógiai és nyelvi értékelési vonatkozások

Az ötödik fejezetben három olyan el tanulmányról esik szó, melyek fontos információval szolgáltak a kutatás további menetét illet en. Az els tanulmány 93 els - és másodéves angol szakos hallgató receptív szókincsének nagyságát mérte fel annak felderítésére, hogy a nemzetközi szakirodalom alapján a hallgatók szókincsének nagysága alkalmassá teszi-e ket egyetemi tanulmányok folytatására angol nyelven. A kapott eredményeket azután összevetettem az egyetemi végzettség angol anyanyelv ek receptív szókincsének nagyságára, valamint az angolt, mint idegen nyelvet tanuló egyetemi hallgatók receptív szókincsének nagyságára vonatkozó nemzetközi adatokkal. Magyarországon ilyen jelleg mérést korábban nem végeztek. Emellett, a további két, kis mintán végzett vizsgálat el zetes betekintést nyújt a receptív szókincs és az írott szóhasználat, valamint a tudatos szótanulás és az olvasás során fellép járulékos szóelsajátítás kapcsolatába, melyek tanulságait a következ fejezetekben tárgyalt vizsgálatok hasznosították.

A legfrissebb kutatási eredmények gyakorlatban történ alkalmazása mellett a tesztkészítés egyik fontos jellemz je kell, hogy legyen a változó szükségletekhez és feltételekhez való folyamatos alkalmazkodás és a javító célzatú fejlesztés. Ezt az irányelvet szem el tt tartva, a hatodik fejezet f célkit zése annak bemutatása, milyen közbüls fázisokon áthaladva alakítottam ki a hallgatók receptív szókincsének mérési eljárását a hallgatók valós lexikális igényeihez igazítva. Az els lépcs ben a bevezet ben említett, mások által gyakran használt és elismert szókincsmér teszt, az úgynevezett *Vocabulary Levels Test* (VLT, Schmitt, Schmitt, & Clapham, 2001) akadémikus szintjét használtam, azt feltételezve, hogy e szavak ismerete meghatározó lehet az angol szakosok tanulmányai során.

Mivel a nyelvi alapvizsgára csak évente két alkalommal kerülhet sor, és a januári vizsgán statisztikai elemzés szempontjából kevés jelentkez áll rendelkezésre, így a vizsgálat folytatására évente egy-egy alkalommal volt lehet ség. A második fázisban ugyanennek a tesztnek az angol nyelvben ritkán el forduló szavak ismeretét mér szekcióját alkalmaztam, feltételezve, hogy ez a szint jobban képes elkülöníteni a kisebb és nagyobb szókinccsel rendelkez hallgatókat, tehát jobban megfelel a nyelvi alapvizsga sz r funkciójának.

A tanulságok levonása után a harmadik fázisban egy saját kidolgozású mér eszközt használtam, melynek szavai részben a VLT akadémikus szintjét alkotják, részben pedig olyan itemek, melyek saját tapasztalataim és intuitív megítélésem szerint tanulmányaik során fontosak és hasznosak az angol szakos hallgatók számára. Emellett, a szókincsmér eszközzel nyert receptív szókincsre vonatkozó adatokat ebben a fázisban összevetettem a nyelvi alapvizsga egyéb komponenseib l nyert adatokkal, azzal a céllal, hogy behatóbban vizsgáljam a szókincs és az olvasás-, írás-, és hallott szövegértési képesség, az általános nyelvi szint, valamint az angol szakos tanulmányi eredmények összefüggéseit, illetve hogy pontosabb képet kapjak az egyik el tanulmányban már vizsgált receptív és produktív szókincs kapcsolatáról. A vizsgálat során tehát a hallgatók írásainak lexikális megítéléséhez a korpuszelemzés módszereit is alkalmaztam.

A fent bemutatott eljárás a szavak gyakoriságának illetve hasznosságának intuitív megítélésén alapul, ám ez nem feltétlenül esik egybe a szavak valódi gyakoriságával (McCrostie, 2007). Ezért a hetedik fejezet azt vizsgálja, milyen szókincsre van valójában szüksége a Pécsi Tudományegyetem angol szakos hallgatóinak ahhoz, hogy sikeresen végezhessék tanulmányaikat angol nyelven. Ehhez a hallgatók számára kurzusaikon kötelez olvasmányokból összeállított reprezentatív szövegmintán végeztem korpuszelemzést. A korpuszt CORES-nak neveztem el (Corpus of Readings in English Studies). A korpusz elemzésével nyert adatok azután forrásként szolgáltak egy olyan szókincsmér eszköz kidolgozásához, mely jobban tükrözi az angol szakos hallgatók valós lexikai igényeit, mint a nemzetközileg kidolgozott és korábban alkalmazott szókincsmér eszközök.

Az értekezésben bemutatott vizsgálatok kutatási kérdéseit, adatgy jt eszközeit és elemzési módszereit a 2. számú táblázat foglalja össze. A továbbiakban a kutatás négy f fázisának eredményeit mutatom be.

2 sz. táblázat: A kutatás fázisai, kutatási kérdései, adatgyűjtő eszközei, és módszertana

	Kutatási kérdések	Adatgyűjtő eszközök	Elemzés módszere
Első előtanulmány 93 résztvevő	Hány angol szót ismernek az első éves angol szakos hallgatók? Vagyis, mekkora az angol szókincsük és elég nagy-e ahhoz, hogy egyetemi tanulmányokat folytassanak angol nyelven?	A szókincs nagyságát mérő teszt (Goulden, Nation & Read, 1990)	Leíró statisztika
Második előtanulmány 33 résztvevő	Milyen összefüggés áll fenn a passzív szókincs nagysága és a hallgatók írásaiban megjelenő aktív szókincs között? A nagyobb passzív szókinccsel rendelkező hallgatók írásaiban változatosabb és kifinomultabb aktív szóhasználat jelenik-e meg, és fordítva?	A szókincs nagyságát mérő teszt (Goulden, Nation & Read, 1990) A hallgatók írásai	A hallgatók írásainak korpuszelemzése Korreláció analízis
Harmadik előtanulmány 33 résztvevő	Hatékonyabb-e hosszú távon a tudatos szótanulás és a szótesztekre való rendszeres felkészülés az olvasás során fellépő járulékos szóelsajátításnál?	A szókincs nagyságát mérő teszt (Goulden, Nation & Read, 1990) Szókincsfejlődést mérő teszt Kvalitatív kérdőív	Korreláció analízis
Első fázis 220 résztvevő	Milyen jól ismerik az első éves hallgatók az egyetemi szintű szövegek értő olvasásához elengedhetetlen (akadémikus) szókincset? Hogyan diszkriminál a hallgatók között a Schmitt, Schmitt & Clapham (2001) által kidolgozott szóteszt akadémikus szókincset mérő része?	Szókincsmérő teszt (VLT, Schmitt, Schmitt & Clapham, 2001) Szókincsmérő teszt (VLT, Schmitt, Schmitt & Clapham, 2001)	Leíró statisztika Korreláció analízis Item analízis
Második fázis 211 résztvevő	Az akadémikus vagy a ritka szavak receptív ismeretének mérése alkalmasabb-e a nem megfelelő szókinccsel rendelkező hallgatók szűrésére? Melyik diszkriminál jobban?	Szókincsmérő teszt (VLT, Schmitt, Schmitt & Clapham, 2001)	Korreláció analízis Item analízis
Harmadik fázis 135 résztvevő	Milyen kapcsolat áll fenn a passzív illetve az aktív szókincs és az írott- és hallott szövegértést, az íráskészséget, és a nyelvhelyességet mérő teszteken elért eredmények között? Milyen kapcsolat mutatható ki a passzív és az aktív szókincs illetve az általános nyelvi szint között? Milyen összefüggés áll fenn a passzív és az aktív szókincs között?	Nyelvtudásmérő teszt Az általam angol szakosok számára hasznosnak vélt szavakból kidolgozott szókincsmérő eszköz Nyelvtudásmérő teszt Az általam angol szakosok számára hasznosnak vélt szavakból kidolgozott szókincsmérő eszköz Nyelvtudásmérő teszt Az általam angol szakosok számára hasznosnak vélt szavakból kidolgozott szókincsmérő eszköz	A hallgatók írásainak korpuszelemzése Korreláció analízis Korreláció analízis Lineáris regresszió analízis

	Kutatási kérdések	Adatgy jt eszközök	Elemzés módszere
	A tanulmányban vizsgált 9 változó milyen összefüggést mutat az angol szakosok tanulmányi eredményeivel?	Nyelvtudásmér teszt Az általam angol szakosok számára hasznosnak vélt szavakból kidolgozott szókincsmér eszköz	Lineáris regresszió analízis
Negyedik fázis 134 résztvev	Mely szavak ismeretére lehet szüksége egy angol szakos hallgatónak a PTE-n, hogy képes legyen megérteni a kötelez olvasmányként feladott tanulmányokat?	Az általam összeállított CORES korpusz	Korpuszelemzés Szókincsprofil
	Korpuszelemezés alapján milyen mértékben segítheti a Coxhead (2000) által összeállított akadémikus szólista ismerete az angol szakosok olvasott szövegértését?	Az általam összeállított CORES korpusz	Korpuszelemzés Szókincsprofil
	Milyen ritka szavakkal találkozhatnak az angol szakos hallgatók kötelez olvasmányaikban?	Az általam összeállított CORES korpusz	Korpuszelemzés Szókincsprofil
	Mennyire ismerik a hallgatók a kötelez olvasmányaik korpuszában leggyakrabban el forduló akadémikus és ritka szavakat?	CORES korpusz alapú szókincsmér eszköz	Leíró statisztika Item analízis

A kutatás eredményei

Az el tanulmányok fontos adatokkal szolgáltak a további kutatás megtervezéséhez, ezért fontos ezekr l is szót ejteni. Az els el tanulmány azt vizsgálta, mekkora az els éves, nyelvi alapvizsga el tt álló hallgatók szókincsének nagysága és ez elegend - e ahhoz, hogy angol nyelven folytathassanak egyetemi tanulmányokat. A mérés eredménye szerint nagyok az egyéni eltérések, de szókincsük nagysága 4.000-11.000 szó között mozog, az átlag 7.123,66 (SD= 1777,97). Amennyiben azt vesszük alapul, hogy minimálisan 5.000 szó receptív ismerete szükséges az írott autentikus szövegek megértéséhez, akkor hallgatóink 15 százalékának kell nehézségekkel szembenéznie az olvasás során. Ha viszont a holland kutatások által javasolt 10.000 szó ismeretét tekintjük szükségesnek, akkor a hallgatók 94 százaléka nem éri el a megfelel szintet. Még a Nation (2006) által javasolt, mértékletesebb, 8.000 szóra vonatkozó adatok szerint is hallgatóink közel 60 százalékának kell lexikális nehézségekkel számolnia tanulmányai során.

Mivel a hallgatóknak a kurzusok teljesítése során írásfeladatokat is teljesítenie kell, a második tanulmányban igyekeztem el zetes képet kapni arról is, milyen kapcsolat állhat fenn a receptív szókincs nagysága és a produktív szóhasználat között. Bár a vizsgálat nem talált szignifikáns összefüggést a két tényez között, a szókincsprofil kimutatta, hogy a nagyobb szókinccsel rendelkez hallgatók írásaikban több nem gyakori szót használtak, mint kisebb szókincs társaik. Ez összhangban áll Laufer és munkatársainak (2004) eredményeivel, ám egyben arra is utal, hogy az íráskészségben nem csupán a szókincs játszik szerepet, ezért további vizsgálatokra van szükség ezen a téren.

A harmadik el tanulmány a szódolgozatokra való rendszeres és tudatos felkészülés és az olvasás során fellép, járulékos szóelsajátítás hatékonyságát vizsgálta. Az el tesztelés során a teszt- és a kontrollcsoport szókincsének nagysága hasonló volt. Mindkét csoport ugyanazokat az olvasmányokat olvasta a félév során azzal a különbséggel, hogy a tesztcsoport az órákon szókincsre fókuszált feladatokat végzett és rendszeres szódolgozatokat írt, míg a kontrollcsoport csupán az olvasmányok tartalmi vonatkozásaira koncentrált. A 14 héten át tartó kezelést követ utóteszten nem volt szignifikáns különbség a két csoport eredménye között az olvasmányok szavainak kvantitatív és kvalitatív módon is mért ismerete tekintetében. Ez arra utalhat, hogy az olvasás során is sajátíthatunk el szavakat és a hallgatók kérd ívben adott válaszaiból az is kiderült, hogy a szódolgozatokat hasznosnak és fontosnak tartják ugyan, de a megtanult szavak többsége nem rögzült bennük hosszú távon. Fontos következtetés tehát, hogy a szavak tartós rögzüléséhez többször és többféleképpen kell találkozni az adott szóval. Ezt a nyelvfejlesztés során lehet vé kell tenni, különösen a nem gyakori szavak esetében.

Láttuk tehát, hogy az els éves angol szakos hallgatók többségének szókincse a nemzetközi adatok tükrében nem elég nagy ahhoz, hogy a tanuló probléma nélkül képes legyen autentikus írott szövegek megértésére, továbbá azt is megfigyelhettük, hogy a receptív szókincs nagysága összefüggésben van az írott szóhasználattal is. A hallgatóknak kurzusaik során pedig szükséges egyetemi szint szakirodalmat olvasni és írásm veket készíteni angol nyelven, ebb l következ en tehát a szókincsbeli hiányosságok a tanulmányi sikerekre is hatással lehetnek. További kutatásaim

fókuszában ezért az állt, mely szavak ismerete lehet hasznára a hallgatóknak és ezeket milyen jól ismerik.

Korábbi kutatások igazolták, hogy magasabb nyelvi szinteken a nyelvtanuló szókincsér l pontosabb képet alkothatunk receptív szókincsének nagysága és produktív szóhasználata alapján, mint kvalitatív vizsgálatokkal (Zareva, 2005). Ezt figyelembe véve, valamint a praktikalitás, objektivitás, az eredmények számítógépes értékelhet sége, és a mérés megvalósítása lehetséges szempontjainak mérlegelésével, a hallgatók szókincsét a nyelvi alapvizsga nyelvhasználatot mér komponensének receptív szókincset mér teszttel való kiegészítése mellett döntöttem. Erre a célra el ször a VLT (Schmitt, Schmitt & Clapham, 2001) akadémikus szókincset mér itemeit választottam ki. A vizsgálat els fázisában megállapítottam, hogy a hallgatók a tanulmányaik szempontjából hasznosnak tekinthet szavakat többségében ismerik, és a teszt nem diszkriminál megfelel en. Ugyanakkor arra is fény derült, hogy a receptív szókincs és a nyelvtani teszt eredményei között szignifikáns összefüggés áll fenn, ami azt jelentheti, hogy a két konstruktum nem független egymástól, ahogyan arra Wray (2002) is utal. Az enyhe korreláció miatt azonban az is nyilvánvaló, hogy a két teszt eltér dolgot mér, szándékaink szerint.

A következ fázisban, egy évvel kés bb, arra kerestem a választ, hogy a nem gyakori szavak ismerete jobban diszkriminál-e a kisebb illetve nagyobb receptív szókincs hallgatók között az akadémikus szókincset mér tesztnél. Az eredmények igazolták e feltevést, ugyanakkor felvet dött a kérdés, mennyire tekinthet k hasznosnak a VLT tízezres gyakorisági szintjének szavai a mi hallgatóink számára. Következésképpen, felmerül az a kérdés, hogy valóban ezeket a szavakat kell-e szókincsmérés céljára alkalmaznunk, vagy talán létezik-e ennél jobb megoldás is az adott kontextusban.

A megoldást az jelentette, hogy az els fázisban itemelemzés alapján megfelel en diszkrimináló néhány akadémikus item megtartása mellett olyan új itemeket dolgoztam ki, melyek intuitív megítélésem szerint jobban tükrözik az angol szakos hallgatók lexikális igényeit. A statisztikai adatok azt mutatták, hogy az új itemek többsége megfelel en diszkriminál (átlag DI= 0,41 és 0,3 a két tesztváltozaton), néhány azonban revízióra szorul majd. Vizsgálataim ebben a fázisban

kiterjedtek a hallgatók írásfeladataiban megfigyelhet produktív szókincsre is. Az adatok arra utalnak, hogy a produktív szókincs mérésére alkalmazott három mér szám közül a T/t arány és az U index nem tökéletesen alkalmas, míg a szókincsprofil megfelel bbnek t nik. A receptív és produktív szókincs összevetéséb l a Pearson korrelációs táblázatok alapján az látszik, hogy a várt er s korreláció helyett csupán a szókincsprofil alapján meghatározott, a szövegben ritkának tekinthet szavak aránya (a GSL kétezer szaván kívüli szavak használatának mértéke = 'Beyond 2.000') mutat gyenge, de szignifikáns kapcsolatot a receptív szókinccsel. Ennek egyik magyarázata az lehet, hogy a receptív és a produktív szókincs fejl dése nem párhuzamosan zajlik, ahogy arra Laufer (1998) is rámutatott, illetve, hogy a hallgatók receptív szókincsének nagysága jóval nagyobb, mint amit aktívan használni képesek az írás folyamán. Másrészt, Jarvis (2002) szerint az írás lexikai változatosságát csupán a nyelvelsajátítás kezdeti szakaszaiban befolyásolja a passzív szókincs nagysága, a mi hallgatóink pedig magasabb szint nyelvtudással rendelkeznek. Emellett valószín síthet az is, hogy az írásokban megfigyelhet lexikai változatosság sokkal összetettebb jelenség, mint a benne el forduló szavak száma és gyakorisága, és nem írható le csupán matematikai összefüggésekkel.

Kutatási kérdésként megfogalmaztam azt is, milyen összefüggés lehet a receptív illetve a produktív szókincs, és a nyelvi alapvizsga tesztkomponensein (olvasott és hallott szövegértés, írás, nyelvtan és nyelvhasználat) elért eredmények között. A receptív szókincs szignifikáns korrelációt mutat mind a négy komponenssel, míg a produktív szókincs csupán a nyelvtani teszt eredményeivel mutat összefüggést. Ez az eredmény a receptív szókincs tekintetében nem meglep , hiszen az olvasott és hallott szövegértés szintén receptív készségeket feltételez. Az, hogy a produktív szókincs mér számai esetünkben nem mutatnak összefüggést az írásfeladaton elért pontszámmal, abból adódhat, hogy az írás megítélése során az értékel k nem csupán a szókincset vették figyelembe, hanem fontos szerepet játszott a kohézió, koherencia, a nyelvhelyesség, valamint a feladatban meghatározott követelmények teljesítésének mértéke is.

Vizsgáltam továbbá a teljes alapvizsgán elért eredmény (mely a nyelvtudás szintjét jellemzi) és a szókincs kapcsolatát is. Korreláció és regresszió analízis alapján

megállapítható, hogy a receptív szókincs és a nyelvtudás szintje nem független egymástól, míg a fent említett mér számokkal kifejezett produktív szókincs nem mutat összefüggést. Végül, kiszámítottam a résztvev k els évben elért tanulmányi eredményeinek átlagát. Ez azt mutatja, mennyire sikeres az adott hallgató tanulmányai során. Pearson korreláció analízis alkalmazásával ezeket az értékeket összevetettem a vizsgálatban szerepl kilenc változóval (receptív szókincs, a produktív szókincs három mér száma, az olvasott- és hallott szövegértés, írás, nyelvtan és nyelvhasználat tesztkomponenseken, valamint a nyelvi alapvizsga egészén elért pontszámok).

A két tesztváltozat közül az els ben gyenge, de szignifikáns összefüggés mutatható ki a tanulmányi eredmények és a hallott szövegértés, valamint az íráskomponensen elért eredmények között, míg a másik változat nem mutat összefüggést. A jelenség egyik magyarázata lehet, hogy a hallgatók a kreditrendszernek megfelel en különféle kurzusokat teljesítenek, melyeken az oktatók nem egységes értékelési szempontokat alkalmaznak. Ugyanakkor a kurzusok teljesítésének kritériuma többnyire egy esszé feladat elkészítése, várható tehát, hogy aki jól ír, sikeresebb is lehet a tanulmányai során. Érdekes aspektus emellett a hallott szövegértés összefüggése is, hiszen a tanulmányok során fontos a tanárok, el adók beszédének megértése. Nagy (2008) kimutatta, hogy a PTE angol szakos hallgatói a tanórákon nem szívesen kommunikálnak, véleményem és adataim szerint találó tehát a magyar nyelvben az egyetemistákra használatos „hallgató" elnevezés.

Összefoglalva, a harmadik fázisban kidolgozott és alkalmazott szókincsmér eszköz alkalmazása hatékonyabbnak bizonyult a korábbi, nemzetközileg használatos eszköznél a mi kontextusunkban, és az idegen nyelvi szókincs számos fontos aspektusára világított rá. Ugyanakkor, a szavak gyakoriságának, vagy hasznosságának intuitív megítélése megkérd jelezhet (McCrostie, 2007), melyet a teszt szókincsprofilja is bizonyított. A következ lépcs ben szükségessé vált tehát annak pontos meghatározása, milyen valós lexikai igényei vannak a PTE angol szakos hallgatóinak.

Ennek megválaszolására egy, a hallgatók számára kötelez olvasmányokból álló reprezentatív korpusz (CORES) létrehozására volt szükség. A CORES lexikai elemzését szógyakorisági profiljának elkészítésével végeztem. A korpuszt három

alkorpusz alkotja (nyelvészeti, alkalmazott nyelvészeti, és irodalom/kultúra szövegek) és összesen 118.808 szóból áll, leszámítva a tulajdonneveket és számokat. Ez az általános korpuszokhoz képest kicsinek tekinthet , ám specializált korpuszként alkalmas az adott diszciplína szóhasználatának feltérképezésére. Bár a lexikai változatosságot jellemz T/t érték nem magas (0,10), a korpusz lexikai s r sége azt mutatja, hogy minden három szóból kett különböz , s ezzel meglehet s lexikai kihívás elé állítja az olvasókat. A szókincsprofil alapján látszik továbbá, hogy a teljes szöveg szókészletének 77 %-át, vagyis több mint kétharmadát, a GSL els 2.000 szava alkotja, igazolható tehát a feltevés, miszerint e szavak ismerete jelent s szerepet játszik a szöveg megértésében. Korábbi vizsgálatok szerint ez az érték általában 80% körül mozog (O'Keefee és munkatársai, 2007). További 7,54%-ot tesz ki az AWL szólistában megtalálható szavak aránya, melyek ismerete tovább növelheti a hallgatók szövegértési képességét. Összeadva mindez a teljes korpusz szókészletének közel 85%-át jelenti, amely azonban még nem elegend arány, figyelembe véve, hogy a szavak 95-98%-ának ismerete szükséges a szöveg megértéséhez (Hirsch & Nation, 1992; Laufer, 2005a, 2005b; Nation & Waring, 1997). Ebb l kifolyólag, a fennmaradó 15 százalék arra utal, hogy a hallgatóknak ismerniük kell emellett olyan, egyébként a korábban említett kritériumok alapján ritkának nevezhet szavakat is, melyek viszonylag gyakran fordulnak el ezen a tudományterületen. Az elemzés ezen eredményeit az értekezés melléklete tartalmazza.

A CORES korpusz elemzése alapján a tesztfejlesztés negyedik fázisában tehát olyan szókincsmér eszközt dolgoztam ki, mely formátumában megegyezik a korábbiakkal, ám itemei a korpuszban el forduló szavak gyakorisága alapján kerültek kiválasztásra. A 134 résztvev s vizsgálat eredményei szerint a harminc itemes teszten az átlag változatonként 18,314 és 16,833 (SD_1=4,686 és SD_2= 3,918), az itemek többsége pedig kielégít en diszkriminál (DI_1= 0,33, SD_1= 0,14; és DI_2= 0,31, SD_2 = 0,18). Az eredmények biztatóak, ám a tesztfejlesztési folyamat ezzel még nem zárult le. Fontos lesz a továbbiakban a kevésbé jól m köd itemek revíziója, újabb itemek, esetleg újabb tesztváltozatok készítése, valamint egy nagyobb mintán végzett, gondos validálási eljárás lefolytatása, miel tt az eszköz széles körben is alkalmazható lesz a jöv ben.

A kutatás korlátai

A kutatás eredményeinek bemutatása mellett szólni kell a korlátairól is. Az értekezésben bemutatott vizsgálatok fókuszában az írott szókincs állt. Mivel a kutatás els dleges célja az volt, hogy az els éves hallgatók számára mérföldk ként el írt nyelvi alapvizsga részeként olyan szókincsmér eszközt fejlesszen ki, mely alkalmas lehet a diszciplína valós lexikai igényeinek figyelembe vétele alapján a nem megfelel szókinccsel rendelkez hallgatók sz résére, vizsgálataim a hallgatók írott szóhasználatára irányultak. Ennek oka az, hogy az alapvizsgának akkor még nem volt szóbeli része, ám a jöv ben hasznos lehet a beszédprodukció hasonló vizsgálata is a bolognai alapképzés (BA) és a mesterképzés (MA) kapcsán.

A vizsgálatokba bevont hallgatók száma a kutatáshoz adott volt, hiszen minden els éves számára kötelez a nyelvi alapvizsga, így a minta a kérdéses kontextusra reprezentatívnak tekinthet . Mivel azonban a statisztikai elemzésekben a résztvev k számának növekedésével bizonyítottan jobb megbízhatóság érhet el (Szabó, 2008), így a jöv ben a dolgozatban bemutatott eredmények alapján átdolgozott korpusz-alapú szókincsmér eszközt nagyobb populáción szükséges kipróbálni, miel tt a szélesebb közönség számára is elérhet vé tehet .

A hallgatók írásainak elemzése, illetve a CORES korpusz szókincsének vizsgálata során alkalmazott szógyakorisági profil képes megállapítani, melyek az adott szövegekben leggyakrabban el forduló szavak. Ugyanakkor a profil nem alkalmas a több szóból álló kifejezések vagy idiómák vizsgálatára, melyek el fordulása a szövegben Muncie (2002) szerint éppen a választékos nyelvhasználat egyik bizonyítéka. A szóismeret e kvalitatív aspektusának vizsgálatát az értekezésben az a megfontolás írta felül, hogy bár a kvalitatív megközelítés során a szóismeret mélységének pontosabb meghatározására van lehet ség, jelent sen kisebb mintán alkalmazható csupán, mint amire esetünkben szükség volt. Mivel jelenleg nem létezik olyan szókincsmér eszköz, mely a szóismeret min ségi és mennyiségi jellemz it egyetlen tesztformátumban lenne képes mérni (ennek egy lehetséges megoldására tettem kísérletet a harmadik el tanulmányban), Laufer és munkatársai (2004) a

szókincs mérését több kvantitatív mér eszköz együttes alkalmazásával javasolják. Az általuk említett feltételeknek az értekezésben bemutatott kutatás eleget tesz, hiszen a nyelvi alapvizsga tartalmaz egy passzív szókincset mér , korpusz alapú, valós lexikai igényelemzéssel kidolgozott eszközt; a hallgatók írásainak lexikális elemzése szógyakorisági profil készítésével történt; a hallott és írott szövegértést mér tesztkomponenseken pedig jelent s szerepet kap a szókincs és a szövegkörnyezet is. E kombináció együttesen átfogóbb képet nyújt a hallgatók lexikális ismereteir l, mint egyetlen mérési módszer alkalmazása.

A mér eszköz kifejlesztése során lehetséges problémaforrásként merül fel az el tesztelés során nyert információ korlátozott volta. A mér eszközök el tesztelése igen fontos, ugyanakkor ellentmondásos fázisa a tesztfejlesztésnek (Szabó, 2008; Alderson, Clapham, & Wall, 1995; Hughes, 1991). Az értekezésben tárgyalt kutatás során az el tesztelés azért okozott nehézséget, mert ideális esetben az el tesztelt csoport nagysága és tulajdonságai a lehet legnagyobb egyezést mutatnak a célcsoport megfelel paramétereivel. Sajnos, mivel a mér eszközöket csupán saját kollégáimon volt alkalmam el tesztelni - mert biztosítani kell az alapvizsga titkosságát -, nem állíthatom, hogy a tesztfejlesztés során ennek a kritériumnak tökéletesen eleget tettem. Azonban az el tesztelés során gyengének vagy félreértelmezhet nek bizonyuló itemeket cseréltem és újabb el tesztelésnek vetettem alá. S t, ilyen esetben az egész klaszter cseréjére sor került, mert azon belül az itemek bizonyítottan nem függetlenek egymástól (Schmitt, Schmitt & Clapham, 2001). A CORES-korpusz alapú mér eszköz jelenleg tehát nem tekinthet még tökéletesen megbízhatónak, és gondos validálási eljárásnak szükséges a jöv ben alávetni nagyobb mintán, miel tt más magyarországi egyetemek angol programjai számára is ajánlható lenne.

Végül, a jöv ben a korpusz-alapú szókincsmér eszköz revíziója és további itemek írása során egy lehetséges újabb szempontként merül fel a szavak korpuszbeli gyakorisága mellett a tesztelésre kiválasztandó szavak kulcsszó státusza. A kulcsszavak Scott (1999) szerint nem feltétlenül a leggyakoribb szavak a szövegben, hanem a „szokatlanul" gyakoriak egy adott normához képest. Egyes szoftverek, például a Wordsmith Tools, mely vizsgálatom során még nem állt rendelkezésemre, nagy referencia korpuszokhoz hasonlítva képesek meghatározni és listázni azokat a

szavakat, amelyek a beadott szövegben szokatlanul gyakran fordulnak el , vagyis kulcsszavai annak. A CORES korpuszra vonatkoztatva, egy ilyen kulcsszó lista segítségével még pontosabb képet kaphatunk arról, mely szavak ismeretét érdemes mérni az angol szakos hallgatóink esetében.

Összegzés és a kutatás jöv beli kiterjesztése

Mindemellett, kutatásom eredményei számos tekintetben hozzájárultak az angol szakos hallgatók angol, mint idegen nyelvi szókincsének leírásához, pedagógiájához és értékeléséhez egyaránt, s egyben új irányvonalakat jelöltek meg a további kutatások számára.

Korpusznyelvészeti vonatkozások

- A hallgatók írásaiból összeállított korpusz, mely bár más korpuszokhoz képest kicsinek tekinthet , betekintést nyújthat a magyar anyanyelv , angol szakos egyetemi hallgatók írott nyelvhasználatába és hozzájárulhat az angolt, mint idegen nyelvet tanulók íráskészségének tanulmányozásához. A hallgatói korpuszt a jöv ben tovább lehet majd vizsgálni más tanulói korpuszok viszonylatában is, illetve a szókincsen kívül egyéb nyelvi elemekre is kiterjesztve.

- A CORES, mint az angol szakos hallgatók számára kötelez en olvasandó tanulmányokból összeállított szöveggy jtemény, hozzájárulhat a diszciplínára jellemz nyelvhasználati sajátosságok feltárásához.

- A kötelez olvasmányok korpuszát a jöv ben folyamatosan b víteni kell és id vel elérhet vé tehet más magyarországi fels oktatási intézmények hallgatói, oktatói és kutatói számára is.

Nyelvpedagógiai vonatkozások

- A tanulói korpuszok hatékonyan alkalmazhatók az írásoktatásban, ahogy azt korábban Horváth (2001) már bizonyította, tehát hasznos lehet a jöv ben az

írást oktató kollégák és nyelvtanárok számára hozzáférhet vé tenni a hallgatók írásainak korpuszát.

- A CORES korpusz alapján összeállított gyakorisági- és kulcsszó listák fontos információval szolgálhatnak a nyelvfejlesztéssel foglalkozó kollégáknak arról, mely szavakra fókuszáljanak tudományterületünkön a hallgatók egyetemi szint , angol nyelv tanulmányok olvasására való felkészítése során. Eredményeim arra utalnak, hogy az angol nyelvben leggyakoribb és az egyetemi szövegekben leggyakrabban el forduló szavak ismerete mellett jelent sen el segítheti a hallgatók olvasott szövegértési képességét azon szavak ismerete, melyek általában az angol nyelvben ritkán, tudományterületünkön azonban viszonylag gyakran fordulnak el . Ebb l adódóan, az els évesek nyelvfejlesztésére irányuló kurzusok oktatói sokféle módon hasznosíthatják a CORES korpuszt a tanórán és azon kívül is a Pécsi Tudományegyetemen, de tágabb vonatkozásban más, hasonló angol szakos képzésben országszerte.

- A CORES korpusz jöv beli alkalmazási lehet ségei nagyban el segíthetik a hallgatók tanulási autonómiáját és a vizsgákra való önálló felkészülést. A korpusz alapján összeállított gyakorisági- és kulcsszó listák, amennyiben a hallgatók szabadon hozzáférhetnek az Anglisztikai Intézet honlapján, a tanórán kívüli tudatos szókincsfejlesztés fontos elemeivé válhatnak.

- A korpuszelemzés egyes eszközeinek alkalmazása, például konkordanciák futtatása a CORES szövegein, lehet séget nyújt a hallgatók számára arra, hogy azonosítsák, és egyéni megoldást találjanak nyelvi problémákra az írott nyelvhasználat minden szintjén. A korpusz referenciaként szolgálhat a hallgatók számára az írásfolyamat különféle fázisaiban, például azáltal, hogy ellen rizhetik benne a kérdéses szavak területükre jellemz kollokációit, melyre a korábban kifejlesztett egyszer gyakorisági szólisták nem adnak lehet séget. Ezáltal a CORES el segítheti a választékosabb, helyes szóhasználatot és a tudatos szókincsfejlesztést.

- Jöv beli terveim között szerepel egy olyan szoftver kifejlesztése a CORES korpuszra, mely további gyakorlási és szókincsfejlesztési lehet séget jelenthet

az angol szakos hallgatók számára a tanórán és azon kívül egyaránt. A számítógépes program elképzelésem szerint a felhasználó választása szerinti nehézségi- vagy szógyakorisági szinteken *C-* vagy *cloze* teszteket generál majd a CORES korpusz és/vagy bármely alkorpusza alapján, szintén a felhasználó igényei szerint. Az eszköz azután minden PTE angol szakos hallgatónak, s t esetleg országos érdekl d knek is, szabadon elérhet lehetne, bár ehhez valószín leg további anyagi és technikai forrásokat kell majd találni.

Értékelési vonatkozások

- A kvantitatív és kvalitatív szókincsmérési módszerek egy tesztformátumban történ ötvözésér l ad számot a bemutatott két el tanulmány, mely egyben kísérlet a szakirodalomban e téren mutatkozó hiány pótlására.

- Az empirikus vizsgálatok eredményei azt látszanak alátámasztani, hogy a szókincs és a grammatikai ismeret egymással szorosan összefügg és a nyelvi mérés terén egy konstruktumként kezelhet .

- A kutatás eredményei fontos információval szolgáltak a szókincsmérés területén arról, mely szavak ismeretének mérésére kell hangsúlyt fektetni, ha a mérés célja az angol szakon sikeres tanulmányok folytatásához nem megfelel receptív szókinccsel rendelkez hallgatók kisz rése.

- Az angol szakos hallgatók speciális lexikális igényeinek figyelembe vételével a CORES-alapú szókincsmér eszköz hatékonyabb, egyedi kontextusra szabott szókincsmérést tesz lehet vé.

A jöv titka még, hogy a célközönség mennyire tartja majd érdekesnek és hasznosnak a fenti javaslatokat. Én azt remélem, hogy az értekezésben felvetett gondolatok és kérdések hozzájárulnak az angol szakos hallgatók lexikális ismereteinek és igényeinek mélyebb megértéséhez, és pótolják a szókincskutatás e terén mutatkozó hiányt.

Hivatkozások

Alderson, J. C. (2000). *Assessing reading.* Cambridge: Cambridge University Press.

Alderson, J.C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation.* Cambridge: Cambridge University Press.

Bloomfield, L. (1933). *Language.* New York: Holt, Rinehart and Winston.

Brown, C., Sagers, S.L., & LaPorte, C. (1999). Incidental vocabulary acquisition from oral and written dialogue journals. *Studies in Second Language Acquisition, 21* (2), 259-283.

Caramazza, A. (1999). Hány feldolgozási szint van a lexikai hozzáférésben? In Z. Bánréti (Ed.), *Nyelvi struktúra és az agy. Neurolingvisztikai tanulmányok* (pp. 249-286). Budapest: Corvina.

Caron, J. (1992). *An introduction to psycholinguistics.* New York: Harvester and Wheatsheaf.

Carter, R. (1998). *Vocabulary. Applied linguistic perspectives.* London: Routledge.

Cobb, T. (2000). Compleat Lexical Tutor. [site Web] http://www.lextutor.ca.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213-238.

Ellis, R. (1994). *The study of second language acquisition.* Oxford: Oxford University Press.

Ellis, R., & He, X. (1999). The roles of modified input and output in the incidental acquisition of word meanings. *Studies in Second Language Acquisition, 21* (2), 285-301.

Gleitman, L., & Landau, B. (Eds). (1994). *The acquisition of the lexicon.* Cambridge, Massachusettes: The MIT Press.

Goulden, R., Nation, P., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics, 11* (4), 341-363.

Grabe, W., & Stoller, F. L. (1997). Reading and vocabulary development in a second language: A case study. In J. Coady, & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 98-122). Cambridge: Cambridge University Press.

Hazenberg, S., & Hulstijn, J. H. (1996). Defining a 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.

Hirsch, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language, 8* (2), 689-96.

Horváth, J. (2001). *Advanced writing in English as a foreign language. A corpus-based study of processes and products.* Pécs: Lingua Franca Csoport.

Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language. *Studies in Second Language Acquisition, 21* (2), 181-193.

Hughes, A. (1991). *Testing for language teachers.* Cambridge: Cambridge University Press.

Hulstijn, J.H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrance of unknown words. *The Modern Language Journal, 80*, 327-339.

Jarvis, S. (2002). Short texts, best fitting curves and new measures of lexical diversity.

Language Testing, 19 (1), 57-84.

Kempen, G., & Huijbers, P. (1983). The lexicalization process in sentence production and naming: Indirect election of words. *Cognition, 14*, 185-209.

Krashen, S. (1989). We acquire vocabulary and spelling by reading: additional evidence for the input hypothesis. *The Modern Language Journal, 73,* 440-464.

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In P. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 126-132). London: Macmillan.

Laufer, B. (1997). 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 (Eds.), *Second language vocabulary acquisition* (pp. 20-34). 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. (2005a). Focus on form in second language vocabulary learning. *EUROSLA Yearbook, 5*, 223-250.

Laufer, B. (2005b). Instructed second language vocabulary learning: The fault in the 'default hypothesis'. In A. Housen & M. Pierrard (Eds.), *Investigations in instructed second language learning* (pp. 311-329). Berlin: Walter De Gruyter**.**

Laufer, B., & Nation, P. (1995). Vocabulary size and use: lexical richness in L2 written production. *Applied Linguistics*, *16* (3), 307-322.

Laufer, B., Elder, C., Hill, C., & Congdon, P. (2004). Size and strength: Do we need both to measure vocabulary knowledge? *Language Testing, 21* (2), 202-226.

Martsa, S. (2007). *English morphology. An introduction.* Budapest: Nemzeti Tankönyvkiadó.

McCrostie, J. (2007). Investigating the accuracy of teachers' word frequency intuitions. *RELC Journal, 38*(1), 53-66.

Muncie, J. (2002). Process writing and vocabulary development: comparing lexical frequency profiles across drafts. *System, 30,* 225-235.

Nagy, B.C. (2008). "*To will or not to will" Exploring advanced EFL learners' willingness to communicate in English*. Unpublished PhD thesis. University of Pécs.

Nation, I. S. P. (2001). L*earning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review, 63*(1), 1-12 .

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.

Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching.* Oxford: Oxford University Press.

O'Keefee, A., McCarthy, M., & Carter, R. (2007*). From corpus to classroom: language use and language teaching.* Cambridge: Cambridge University Press.

Paribakht, T. S., & Wesche, M. (1999). Reading and "incidental" L2 vocabulary acquisition. An introspective study of lexical inferencing. *Studies in Second*

Language Acquisition, 21 (2), 195-224.

Pawley, A., & Syder, F.H. (1983). Two puzzles for linguistic theory: native-like selection and native-like fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191-226). London: Longman.

Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: an assessment perspective. *Language Learning, 52* (3), 513-536.

Qian, D. D., & Schedl, M. (2004). Evaluation of an in-depth vocabulary knowledge measure for assessing reading performance. *Language Testing, 21* (1), 28-52.

Read, J. (2000). *Assessing vocabulary.* Cambridge: Cambridge University Press.

Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition, 21*, 589-619.

Schmitt, N. (2008). Instructed second language vocabulary learning. Review article. *Language Teaching Research, 12* (3), 329-364.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing, 18* (1), 55-88.

Scott, M. (1999). *Wordsmith Tools.* Software. Oxford: Oxford University Press.

Singleton, D. (1999). *Exploring the second language mental lexicon.* Cambridge, Cambridge University Press.

Singleton, D. (2000). *Language and the lexicon. An introduction.* London: Arnold.

Szabó, G. (2008). *Applying item response theory in language test item bank building.* Frankfurt am Main: Peter Lang.

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics, 28* (1), 46-65.

Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition, 30* (1), 79-95.

West, M. (1953). *A general service list of English words.* London: Longman.

Wode, H. (1999). Incidental vocabulary acquisition in the foreign language classroom. *Studies in Second Language Acquisition, 21* (2), 243-258.

Wray, A. (2002). *Formulaic language and the lexicon.* Cambridge: Cambridge University Press.

Zareva, A. (2005). Models of lexical knowledge assessment of second language learners of English at higher levels of language proficiency. *System, 33*, 547-562.

A témához kapcsolódó saját publikációk

Lehmann, M. (2007a). Is intentional or incidental vocabulary learning more effective? *International Journal of Foreign Language Teaching, 3* (1), 23-28. http://www.tprstories.com/ijflt

Lehmann, M. (2007b). The lexical diversity of short texts: Exploring the productive-receptive continuum of lexical knowledge. In J. Horváth & M. Nikolov (Eds.), *UPRT 2007: Empirical studies in English applied linguistics* (pp. 293-305). Pécs: Lingua Franca Csoport.

Lehmann, M. (2007c). 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 Theoretical and Applied Linguistics* (pp. 25-38).

Sifakis, N., Oder, T., Lehmann, M. & Bluma, D. (2007). Aspects of learner autonomy in the national curricula of four European countries. In A-B. Fenner, & D. Newby (Eds.), *Coherence of principles, cohesion of competences: Exploring theories and designing materials for teacher education* (pp. 132-151). Council of Europe Publishing: Strasbourg.

Lehmann, M. (2006a). Vocabulary as a filter with first-year English majors. In M. Nikolov, & J. Horváth, *UPRT 2006. Empirical studies in English Applied Linguistics*, (pp. 121-137). Pécs: Lingua Franca Csoport.

Lehmann, M. (2006b). Magoljunk vagy olvassunk?: El tanulmány a tudatos és járulékos szótanulás összehasonlítására. *Modern nyelvoktatás, 12* (1), 44-58.

Lehmann, M. (2005). Bárdos Jen : Él nyelvtanítás-történet. Könyvismertet . *Magyar Pedagógia, 14* (3), 377-378.

Lehmann, M. (2004). Az idegennyelvi szókincs elsajátítás és mérés innovatív irányzatai. *Alkalmazott nyelvtudomány, 4* (1), 41-57.

Lehmann, M. (2003a). Range. Szoftver-kritika. *Modern nyelvoktatás, 9* (4), 88-89.

Lehmann, M. (2003b). 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,* (172-182). Pécs: Lingua Franca Csoport.

J. G. Gebhard, M. Fodor & M. Lehmann. (2003c). Teacher development through exploration: Principles, processes and issues in Hungary. In J. Andor, J. Horváth, & M. Nikolov (Eds.), *Studies in English Theoretical and Applied Linguistics,* (250-262). Pécs: Lingua Franca Csoport.

Konferencia el adások

Lehmann, M. (2009a). Korpusz-alapú szókincsmérések angol szakos hallgatókkal. 2009. október 26-27. Új módszerek az alkalmazott nyelvészeti kutatásban cím konferencia. Budapest, ELTE-Kodolányi János F iskola.

Lehmann, M. (2009b). Free productive vocabulary in student essays. 2009. június 12. University of Pécs Round Table conference.

Lehmann, M. (2009c). Issues of student vocabualry: quantity, quality, purpose. 2009. január 22-24. Pécs, HUSSE konferencia.

Lehmann, M. (2007a). Lexical diversity of short texts: Exploring the receptive/productive continuum of lexical knowledge. 2007. június 2. Pécs. University of Pécs Round Table Conference.

Lehmann, M. (2007b) Angol szakos hallgatók produktív szókincse. 2007. április 19-21. Kodolányi János F isloka, Siófok. XVII. Magyar Alkalmazott Nyelvészeti Kongresszus.

Lehmann, M. (2006a). Az idegennyelvi szókincs és az olvasott szövegértés kapcsolata. Mérések angol szakos hallgatókkal. XVI. MANYE Kongresszus, Szent István Egyetem, Gödöll . 2006. április 10-12.

Lehmann, M. (2006b). Is intentional or incidental vocabulary learning more effective in the long-term retention of words? University of Pécs Round Table conference. 2006. január 20. Pécs.

Lehmann, M. (2005). EFL vocabulary as a predictor of success in academic reading. 2005. szeptember 14-17. EUROSLA 15 Conference, Dubrovnik.

Lehmann, M. (2004). Az idegen nyelvi receptív szókincs nagysága. IV. Országos Neveléstudományi Konferencia 2004. október 20-22. Budapest, MTA.

Lehmann, M. (2003). Korpusz alapú szógyakorisági listák alkalmazása a szókincsmérésben. III. Országos Neveléstudományi Konferencia. 2003. október 9-11. Budapest, MTA.

Lehmann, M. (2002). The relationship between the lexis of L2 learners' writing and the size of their receptive vocabulary. XII. MANYE konferencia, 2002. március 27-29. Szeged, Juhász Gyula Tanárképz F iskola.