A MODEL OF LEXICAL KNOWLEDGE ASSESSMENT OF ADULT NATIVE AND NON-NATIVE SPEAKERS OF ENGLISH. STRUCTURE OF THE MENTAL LEXICON OF ADVANCED AND INTERMEDIATE NON-NATIVE SPEAKERS OF ENGLISH

by

ALLA PETROVA ZAREVA

(Under the Direction of Paula Schwanenflugel)

ABSTRACT

The study focused on adult second language (L2) learners' structure of the mental lexicon and the dimensions on which the overall state of their vocabulary knowledge can be examined. A model of studying lexical competence was proposed, which distinguished within the receptive and the productive domain two levels of assessment analysis: a microlevel, knowledge of the *grammar* of individual words, and macrolevel, the overall state of a learner's lexicon as reflected by vocabulary size, quality of lexical knowledge, and metacognitive awareness. The experiment designed to test the proposed dimensions was an attempt to conduct a model-based research that explored lexical competence of adult native speakers (NS) of English, L2 advanced, and intermediate learners with regard to their language proficiency and varying degree of familiarity with words. The quantitative analyses confirmed that the intermediate learners differed significantly from the other two groups in all factors used in the study but metacognitive awareness, whereas the advanced learners' overall state of the lexicon closely resembled that of NS across all variables but one – nativelike typicality of associations. An in-depth analysis of this variable revealed that just as most NS maintain common word association networks, L2 learners also have stable patterns of commonality of associations that do not, however, resemble the ones of NS. It was also argued that a more constructive way of studying how speakers of two or more languages organize their mental

lexicon would be to consider the stability of their meaning connections rather than the degree of nativelike commonality of their word associations. Next, by using several methods of regression analysis, a smaller set of factors that reflected the participants' language proficiency was identified as reliable predictors of lexical knowledge for assessment purposes. Similarly, a model comprising of the "best" set of predictors was proposed, which was as practically efficient as the full model. Finally, the interpretation of the lexical relations between the word associations and the stimulus words used in the experiment complemented meaningfully the overall qualitative and quantitative examination of the differences between the way NS and non-NS organize their mental lexicon.

INDEX WORDS: Lexical Competence, Vocabulary size, Quality of Word Knowledge, Metacognitive Awareness, Model of Lexical Knowledge, Lexical Assessment, Language Proficiency, Word Association Tests, Structure of the Mental Lexicon, Second Language Acquisition

A MODEL OF LEXICAL KNOWLEDGE ASSESSMENT OF ADULT NATIVE AND NON-NATIVE SPEAKERS OF ENGLISH. STRUCTURE OF THE MENTAL LEXICON OF ADVANCED AND INTERMEDIATE NON-NATIVE SPEAKERS OF ENGLISH

by

ALLA PETROVA ZAREVA

B.A., College of International Tourism, Bulgaria 1984

M.A., Veliko Turnovo University, Bulgaria 1992

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2004

A MODEL OF LEXICAL KNOWLEDGE ASSESSMENT OF ADULT NATIVE AND NON-NATIVE SPEAKERS OF ENGLISH. STRUCTURE OF THE MENTAL LEXICON OF ADVANCED AND INTERMEDIATE NON-NATIVE SPEAKERS OF ENGLISH

by

ALLA PETROVA ZAREVA

Major Professor:	Paula Schwanenflugel
Committee:	William Provost Marlyse Baptista Betsy Rymes

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2004

To my mother, Rozalia Subeva

Without her, I would not have become what I am today.

ACKNOWLEDGEMENTS

This dissertation wouldn't have been possible had it not been for the support, love, patience, and generosity of many people, who have helped me in their own incredible way to deal with myself while trying to find my voice in what I do.

I would especially like to acknowledge and thank Dr. Paula Schwanenflugel, who as a major professor and director of my dissertation opened my eyes to see the details in the "big picture". I highly admire her scholarship, dedication, efficiency in decision-making, valuable advice, and guidance. Thanks to her, I got to know Dr. Stephen Olejnik, who magically opened up my mind for a science that I never imagined my mind was capable of comprehending. Special thanks must go to Dr. William Provost, who as a Director of the Linguistics Program has enormously helped me to keep on moving. Had it not been for his constant administrative and academic support, I doubt I would have finished the Program and my dissertation in such a timely manner. There isn't a quantifier in the English language that can express how grateful I am for everything that he has done for me during the past four years. I would also sincerely like to thank Dr. Marlyse Baptista, who has always been full of new ideas and contagious enthusiasm and who has encouraged me to be respectful to the papers that I've written. I have a special thank for Dr. Rymes as well, whose energy and resourcefulness have been inspiring. Finally, I a big "thank you" should go to the participants in this study and especially to a good friend and colleague of mine at Bourgas Free University in Bulgaria, Yordanka Nikolova, who graciously helped me with collecting the data for the study.

My greatest thanks go to my nearest and dearest – my family and my friends - whom I selflessly and unconditionally love, and I know they love me back. I thank my parents Rozalia Subeva and the late Peter Subev for giving me life, love, values, and belief that I can accomplish anything I set out to do. Thank you, Mum, for sacrificing so much to make this

possible for me! All I have ever wanted was to make you proud of me! And thank you to my brother, Nikolay, who has always been my most reliable ”little brother”. I also don’t think it would have been possible to write this dissertation without my husband, Dimitar Zarev, and my son, Ivo, being lovingly and loyally by my side, even from a distance. My husband’s patience and high level of tolerance to all my academic queries, intellectual quests, and unpredictable mood swings have been more than highly appreciated. My son’s well-being, love, and encouragement have been my guiding light. Finally, I am entirely indebted to all my friends for their spiritual generosity and for being my friends. Thank you!

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

LEXICAL KNOWLEDGE

(INTRODUCTION AND LITERATURE REVIEW)

Introduction

Vocabulary has been recently acknowledged as one of the most important components of language proficiency in both first language (L1) and second language (L2) acquisition. The last couple of decades noticeably brought to the fore the need for re-evaluation of the place of lexical knowledge in L2 acquisition, which, consequently, revived L2 researchers' interest in studying various aspects of knowledge of words and their relationship to language proficiency. For a long time, though, lexical researchers have primarily been concerned with estimating only one dimension of lexical knowledge, vocabulary size, or *breadth* of vocabulary knowledge, paying little attention to any other aspects. However, as interest in vocabulary growth developed, researchers started to show increasing awareness that a measure of vocabulary size alone can no longer provide a satisfactory description of L2 lexical knowledge (e.g., Schmitt, 1999; Wolter, 2001; Greidanus & Nienhuis, 2001; Meara, 1978; Meara, 1996; Nation, 1993; Read, 1993; Wesche & Paribakht, 1996) because knowledge of words is multidimensional, regardless of whether lexical items are tested in context or in isolation. In other words, there is growing realization in the field that, on the one hand, language users differ in how much they know about the lexical features of each individual item that comprises their mental lexicon and, on the other hand, those differences are mirrored to a great extent by the overall state of their lexical competence. Therefore, one of the goals of vocabulary research began to be seen in

designing a relevant theoretical framework of studying lexical knowledge. Such that would make it feasible to practically examine as many aspects of word knowledge as possible and draw relevant to the researched level conclusions, i.e. conclusions that would not underestimate or overgeneralize what learners know about words. Similarly, the relationship between various aspect of lexical competence and L2 learners' language proficiency also started to attract a lot of attention because of findings that pointed to a strong correlation between vocabulary and certain language skills, such as reading and writing. Consequently, L2 lexical researchers began to put forth a great amount of effort in finding out more about the potential of using vocabulary tests to predict the proficiency of language learners.

In this chapter, I will present an overview of the literature concerning L2 lexical knowledge and the dimensions of studying lexical competence. I will, first, discuss the place of vocabulary in L2 acquisition and research, in particular, the shift of focus from structure to lexis, in light of current linguistic theory and psycholinguistic research. I will argue that the current heightened interest in the lexicon is the meeting point of three bodies of research, i.e. applied linguistic, linguistic, and psycholinguistic, which are frequently thought of as being part of completely different research traditions. In the context of this discussion, I will also try to delineate the place the experimental study that will be presented in the next chapters in current L2 lexical research. Secondly, I will briefly touch upon some problems related to the notion of "word" and the ways they are resolved in L2 lexical research. The discussion on the notion of a word is particularly relevant to the discussion on how knowledge of words is studied. Next, I will elaborate on models of lexical knowledge as proposed by various L2 lexical researchers by examining several separate trait proposals and some global trait models. Finally, I will outline the scope of the study that will be presented in this thesis, the issues of research interest, as well as the methodology that is used in the experiment.

Lexis, language structure, and lexical access: general considerations

There is a common feeling in the field that until recently L2 research has been synonymous with "grammar" research with a primary focus on understanding the acquisition of syntactic structures, while the investigation of the lexicon remained a largely overlooked area. In the last couple of decades, though, many researchers realized that "using the right word is the most important aspect of language use" (Politzer, 1978, p. 258), that lexis should become a "major learning priority" in L2 acquisition (Jones, 1994, p. 441), that lexical errors far outnumber grammatical ones as language proficiency increases, and that L2 learners themselves feel it is vocabulary that causes most of their language problems, not syntax (Meara, 1978). Unfortunately, not so long time ago, the most striking dilemma in English language teaching and syllabus design was the dilemma between the role of syntax and the role of lexis in language teaching and learning. In this regard, O'Dell (1997) rightly points out that "the words *lexis* and *vocabulary* are remarkable by their absence from either chapter headings or indexes in the major writers on syllabus of the 1970s and 1980s" (p. 258). And it was only since the second half of the 1980s that syllabus theorists and designers have began to direct attention to lexis, challenged by the growing realization that possessing a good vocabulary stock is what enables language learners to use their knowledge of the language in a way that meets their needs (e.g., Dubin & Olshtain, 1986; O'Dell, 1997). Regrettably, I believe that one of the major reasons behind L2 teaching practices failing to acknowledge the importance of developing learners' vocabulary for such a long time is linked to the overall reluctance of linguistic, psycholinguistic, and applied linguistic bodies of research to mutually inform each other's work with the findings of each other's research. The traditionally given explanation to that problem maintains that these three research paradigms belong to different research traditions; hence, they tend to think and explore language in fundamentally different ways (which sounds more like an excuse rather than a resolution to the problem). However, this should not necessarily be the case considering the fact that the three research traditions very often approach the study of a linguistic

phenomenon from commonly shared theoretical perspectives. And I believe that the shift of focus in L2 teaching practices from “grammar” mainly to including lexis as well is a perfect example of promoting successful L2 learning by incorporating in ESL/ EFL instruction linguistic, psycholinguistic, and applied linguistic ideas concerning L2 acquisition. As a result, the relationship between grammar and lexicon has been fundamentally re-evaluated and the previously artificially drawn dividing line between the two domains has been seriously called into question (Singleton, 1999). In the following paragraphs I will succinctly present my views on the relationship between linguistic theory, psycholinguistic research, and applied linguistic research and try to show how these three bodies of research inform each other. In doing so I will also attempt to determine the place of the experiment that will be discussed in this thesis and its relevance to current ideas about the lexicon and recent models of lexical production. Finally, I will provide an outline of the main research questions that will be investigated in the present experimental study and the way they contribute to L2 lexical research.

One of the principal questions about the role of theory in the growing subfields of interdisciplinary studies and applied linguistic research concerns how theoretical concepts are translated into empirical procedures (McLaughlin, 1987). In this regard, current ideas about the place of the lexicon in the scheme of language acquisition offer a very productive point of departure for both theoretical models and empirical studies in this field. For example, a central principle of the “Government and Binding” version of Chomskyan generative grammar model is the so-called “Projection Principle”, which underscores the idea that the lexical information determines to a large extent the syntactic structure (Haegeman, 1994, p. 55). Cook further comments on this principle thus:

> The lexicon is not a separate issue, a list of words and meanings; it plays a dynamic and necessary part in the syntax. …GB does not segregate syntactic and lexical phenomena. Consequently, many aspects of language that earlier models dealt with as “syntax” are now handled as idiosyncrasies of lexical items; the syntax itself is simplified

by the omission of many rules, at the cost of greatly increased lexical information.(Cook, 1988, p. 11).

Recently, the recognition of the grammatical role for the lexicon in linguistic theory has evolved even further. Current hypotheses coming out of the Chomskyan tradition, hold the view that “syntax is invariant; languages differ in their lexicons” (Cook, 1995, p. 63), when it comes to the universal principles governing languages, and that “language acquisition is in essence a matter of determining lexical idiosyncrasies” (Chomsky, 1989, p. 44). These ideas have been further developed by the Minimalist Program, which came as a logical outgrowth of Chomsky’s theory of generative grammar. Having reduced the levels of representation to two, PF (Phonetic Form - an abstract representation of sound), and LF (Logical Form - an abstract representation of meaning), the Minimalist Program (MP) puts the lexicon in the center of syntactic structure. In other words, the lexicon is assumed to be “an arguably irreducible component of the grammar expressing what we know when we know the words of a given language” (Epstein, Thránsson, & Zwart, 1996, p. 8), which can potentially be described as being the source of all variation between languages (Smith, 1999, p. 50). On the whole, the generally abstract ideas that MP advances come mainly to attract attention to the hypothesis that language-specific structures are encoded in the lexicon. Consequently, how much L2 learners know about the lexical items comprising their L2 lexicon determines how well they will be able to manipulate syntactic structures in the new language.

Some recently proposed models of language production are fully consonant with the ideas advanced by MP and the trend towards blurring the distinction between lexicon and grammar. One such model that has also received a wide recognition in L2 research is the lexical model proposed by Levelt (1989). Even though the model was initially developed to describe the production of L1 speakers, De Bot (1992) convincingly argues that a single model that describes both production and comprehension of both L1 and L2 speakers should be preferred to separate models, especially in light of the fact that monolingualism is more the exception than

the rule. At present, psycholinguistic research in lexical access has been mostly interested in the kind of processing mechanism that governs the skill of accessing words and Levelt's model proposes two component processes to lexical access. The first one is called lexical selection, i.e. retrieving the appropriate lexical item from among several alternatives, and the second one is referred to as phonological encoding, i.e. computing the phonetic shape of the selected item's phonological code, as stored in the mental lexicon (Levelt, 1993). In this view, the mental lexicon consists of lemmas (semantically and syntactically specified forms) and lexemes (morphologically and phonologically specified lemmas) (Kempen & Huijbers, 1983). So, in the first phase of language production, i.e. semantic selection, lemmas are retrieved when the semantic conditions of the message are met which, in turn, activates the syntactic specifications of the selected item. In the second phase of lexical access, i.e. phonological encoding, the selected item is given phonetic shape. The model also assumes that the grammatical form, the argument structure, and the semantic form are related types of information that constitute the lemma of an entry. The phonetic form and the graphemic form belong to the system of lexemes. It is important to note that this proposal has essential consequences for the status of the different types of lexical items in the mental lexicon because, on the one hand, it determines the combinatorial properties of lexical items; on the other hand, it defines the contribution of an item to the articulatory organization.

Overall, the two step lexical access models, in one guise or another, have been common for quite some time to most modern views of lexical access (e.g., Morton, 1969; Levelt, 1989, 1993; etc.). I focused on Levelt's blueprint of lexical access here because it has the advantage, among all other advantages, of representing the lexicon as mediator between conceptualization and grammatical and phonological encoding. Furthermore, the view that lexical selection drives grammatical encoding is in concert with current views of the relationship between lexis and syntax advanced by Minimalism and one cannot but notice the resemblance between the hypothesized two step lexical access model and two levels of representation hypothesized by

the minimalists. On a final note, I believe that the interest in the lexicon in L2 psycholinguistic and applied linguistic research has been revived by the new theoretical developments in linguistic theory, which did need empirical validation. I am also convinced that lexical research can be the most constructive meeting point of three bodies of research - linguistic, psycholinguistic, and applied linguistic - regardless how different their research traditions are. On the contrary, it might well be the case that it is the difference in tradition that would bring them closer together, rather than draw them apart.

The experimental work that will be discussed in this thesis is an attempt to bridge the aforementioned traditions. My interests in the lexicon were inspired by linguistic theory. My proposal of a model of lexical knowledge was motivated by the psycholinguistic model-based tradition of research. The experiment, which was designed to test the model, was an attempt to conduct a model-based applied linguistic research that would allow for tangible generalizations, not overgeneralizations. Regrettably, one cannot but notice that most of the L2 lexical research was carried out in, so to say, piecemeal fashion, i.e. the researched lexical phenomena and hypotheses, more often than not, were analyzed outside a specific model of studying the L2 lexicon, which consequently led to creating an incomplete, patchwork-like picture of lexical competence, often hinging on assumptions and metaphors about its organization borrowed from L1 lexical research. To make up for this deficiency in the L2 applied linguistic field, in this thesis I propose a model of studying and assessment of native speakers' (NS') and L2 learners' lexical knowledge, which is also empirically tested by an experiment specifically designed to try out the model. I also statistically examine the potential of each variable and combinations of variables used in the experiment to account for the greatest amount of variance in the lexical knowledge of NS and non-native speakers (non-NS) alike, when differences in language proficiency are taken into consideration. The underlying assumption behind this procedure is that such an analysis would allow me to identify the smallest set of variables that can be employed as practically efficient predictors of adult NS' and L2 learners' lexical knowledge. Next, by

examining the word associations (WA) generated by the participants in the experiment I offer analyses of the structure of the mental lexicon of NS and non-NS, considering the fact how little research has been done on the role of language proficiency in the associative patterning of L2 learners' lexical knowledge and the nature of the differences between the structure of the NS' and non-NS' lexical knowledge, i.e. whether these differences are predominantly qualitative, quantitative, or both in nature. Finally, after having examined issues related to the qualitative and quantitative patterns of associative responses of NS and L2 learners, I offer an in-depth analysis of possible influences stemming from particular SW characteristics, i.e. lexical category and frequency of occurrence of SW, on the features of NS' and L2 learners' WA domains. I believe that the answers to the questions raised in the study will bring us closer to understanding the nature of the factors that have impact on the way NS and non-NS organize their lexical knowledge, how language teaching can facilitate the development of better structured L2 mental lexicon, and how the assessment of adults' L2 lexical competence can be used to predict their overall level of proficiency. Therefore, I see this study addressing two sets of questions, one -- applied linguistic --related to lexical knowledge assessment and another one –psycholinguistic -- concerning lexical knowledge organization, which on the surface seem relatively independent; however, when looked into in depth, the relevance of their interdependence becomes unambiguously clear.

The notion of "word"

Any discussion concerning the lexicon, the dimensions of lexical knowledge, or the structure of the mental lexicon is preceded by an overview of what a word is. In this section, I will outline how the notion of word is treated in the relevant literature and how some of the problems associated with defining a word have been resolved in L2 vocabulary research. First and foremost, beyond an interest in the concept of word at a theoretical level, a definition of a word is of greatest importance to specifying what it means to know a word for the practical

purposes of studying language users' lexicon. Based on an operationalized definition of a word, for example, the first step for most researchers would be to identify what lexical aspect they will be investigating, i.e. whether their focus will lie in the lexical features associated with individual words or whether their research interest will concentrate on the overall state of learners' vocabulary. Next, dependent on how a word is defined, a decision would have to be made whether vocabulary is comprised of single words only, or whether larger lexical items, e.g. idioms, collocations, compounds, etc., would be considered as well. Finally, a satisfactory working definition of a word would also be required in order to determine the nature of the construct "vocabulary knowledge" (Read, 2000), which is the key to understanding what this knowledge involves. The issues that will be discussed in the following paragraphs are an inseparable part of the problems lexical researchers face in making decisions regarding what a working definition of a word should include, what words should be tested, how they should be selected for testing, etc. Here, I only briefly touch upon some of the difficulties associated with defining a word and how L2 lexical researchers go about the problem. In Chapter 2 I further elaborate on some of these issues in the context of stimulus word (SW) selection from a dictionary source.

By and large, defining a word is considered to be a difficult task because a definition needs to include specifications at several levels, i.e. at the level of abstraction as well as at the linguistic levels (Singleton, 1999). Additionally, a definition will also need to account for the extent to which semantic content is used as a criterion in defining what a word is (ibid.). At the level of abstraction, there are several basic points that need to be explained at the start considering their relevance to SW selection from a dictionary for assessment purposes. One of the distinctions that has to be made is between words counted in terms of tokens (i.e. the actual occurrence of an item) or as types (i.e. items with different identity). When words are counted as tokens, every time a word is used in a source it is counted as a separate item, regardless of whether the word or its inflected forms occur just once or several times. On the other hand, the

number of types is the total number of different word forms, which are counted only once in a dictionary source or a text. For example, on a count of tokens the verb *play* whose paradigm consists of four forms (*play, plays, played, playing*) will be considered to comprise four words if they occur in a running text, whereas on a count of types, the four forms will be taken to be just one word. Overall, L2 lexical researchers, who select words from a dictionary to test vocabulary knowledge, treat words as types rather than tokens. This procedure is based on the assumption that inflectional morphology is not linked to a particular lexical item but to a specific lexical category (e.g., nouns, verbs, adjectives, etc.), which allows inflectionally affixed forms to be treated as different realizations of a single word (Sandra, 1994).

When attention is restricted to content words, a working definition of a word should necessarily address the question of how inflected and the derived forms will be treated. At a morphosyntactic level, the status of derivational morphemes is different from that of inflectional morphemes; hence, they should receive a different treatment in discussions on the structure of the mental lexicon. A major argument in support of this assumption is that inflections can be added to all kinds of words, regardless of whether they have affixes or not and whether they are compounds or not. But as soon as an inflectional morpheme is added, no further suffixes can be added (e.g., **play**s**ful*) and it is not possible to use the word with an attached inflectional morpheme in a compound (e.g., **play**s**mate*) (Levelt, 1989). The difference in treatment of inflectional and derivational morphemes is also supported by psycholinguistic research on the organization of the morphology in the mental lexicon of language users, which has convincingly argued that inflectional suffixes should be treated outside the lexical domain because they are linked to a particular class form (e.g., verbs, nouns, adjectives, etc.) and produce, in fact, different realizations of a single word rather than different words (Sandra, 1994 for an overview). Therefore, much of the research investigating the structure and the organization of the mental lexicon ignores inflectional morphology, as being unproductive, and focuses on the lemma as a unit of analysis. In vocabulary studies, the base form and its inflected forms are collectively

known as a lemma. A lot of research studies investigating the lexicon and language users' lexical organization take the lemma as a unit of analysis that holds a potential to reveal the organization within and between lexical entries (Levelt, 1989). Therefore, in lexical studies that involve counting the number of words, e.g. the number of words in a written or spoken text, the number of words in a dictionary, or estimating the vocabulary size of learners, one of the first steps that needs to be taken is to lemmatize the tokens so that the inflected forms are counted as instances of the same lemma as the word base (Read, 2000).

In general, an issue that seems to be of a more complicated nature is the treatment of certain derivatives because there doesn't seem to be agreement among researchers regarding which affixes should be excluded from word counts and, respectively, from testing and why they should be excluded. More specifically, the issue in question is the concept of a word family which, even though intuitively attractive and widely used, does need further specification in order to be applied in a unified way across lexical studies. While Nation (2001) argues that "when we talk about knowing a word, we should really be talking about knowing a word family" (2001, p. 47), the definition of a word family that he proposes leaves too much latitude in the choice of what to be included in a word family. In his words "a word family consists of a head word, its inflected forms, and its closely related derived forms" (Nation, 2001, p. 8). While inflectional morphemes are a clear case for treatment (included in the concept of lemma), the definition potentially leaves a lot of latitude in the decisions concerning which derivatives are closely related and which are not. Thorndike and Lorge (1944), for example, when compiling their frequency of occurrence count, listed all regularly pluralized nouns with *–s* morpheme, regular comparative (*-er*) and superlative inflections (*-est*), verb forms in *–s, -ed, -ing*, past participles formed by adding *–n*, and adjectives ending in *–ly* under the main word, thus treating all these forms as comprising a word family. In other words, the researchers included morphemes that can be both inflectional and derivational (e.g., *–ed*, *-ing*, *-n*) as well as only derivational (e.g., *-ly*) as forms of one and the same word without any justification of this

inclusion. In vocabulary studies, different researchers propose different sets of affixes to be included in a word family by putting forward different reasons for inclusion. Some researchers (e.g., Nation, 1993, 2001) suggest that morphological forms of the base word plus *–s*; *-ed*; *-ing*; *-ly*; *-er*; and *-est* should not be counted as separate items but as a single word family because they are used systematically in the language and, thus, significantly reduce the learning burden of the resultant words. Other researchers (e.g., Meara, 1996; Nation, 2001) propose that this list of morphemes should be enriched with some common derivational affixes such as *un-*; *-ness*; *-ment*, etc. because these affixes can easily be understood by L2 learners, without having to learn each form separately. Unfortunately, regardless of the argumentation underpinning the aforementioned proposals, none of the researchers has offered linguistic or psycholinguistic reasons that would justify the inclusion of the proposed morphemes in the word family. Ease of recognition and reduced learning load are practically meaningful reasons but they do not explain why some of the most productive affixes (e.g., *–ness, -ly, un-, -ment*) should be included in the word family, while other equally productive, frequent, and regular affixes (e.g., *–ize, -ate, in-*)(cf. Baayen & Renouf, 1996; Plag, 1999) should not. Therefore, the decision about the criteria that would be used to group words into word families is one of the most important decisions that a researcher has to make. In Chapter 2 I discuss in great detail how the SW used in this study were selected for testing and I also offer justification of the selection criteria applied.

The linguistic levels of describing a word play an important role in lexical research in that it focuses on examining the acquisition of various word features by L2 learners. For example, as an orthographic entity the verb *plays* consists of a sequence of five letters, i.e. p + l + a + y +s; as a phonetic entry, it consists of a sequence of sounds with specific phonetic properties; as a phonological entry, it represents a sequence of phonemes, i.e. /p + l+ e + ɪ + z/ which combine in a pattern that English phonology allows; at a morphological level, it consists of two morphemes – one free and one inflectional bound morpheme (play + -s); at a syntactic level, the three parameters specified in the lemma (the 3rd person + singular + present tense) are realized

together in the suffix *–s*; and at a semantic level the verb can be treated, for example, as a synonym of *entertain oneself*. Sometimes, at a semantic level, this form is called a lexeme, or a semantic word that is specified for a sense (Seed, 1997). In other terminologies, a lexeme is a form already specified for semantic and syntactic features to which morphological and phonological codes are added (cf. Levelt, 1989). Thus, knowledge of word features can be studied at several linguistic levels dependent on the feature(s) a researcher may choose to study.

In a nutshell, defining a *word* is one of the most controversial issues in theoretical linguistics as well as in lexical research because one can think of a word in many different ways, for example, word types, word tokens, phonological words, syntactic words, lexical items, etc. Therefore, the most practically useful approach to defining of a word for research purposes is to operationalize the concept in a way that would make it clear what aspects of the multifaceted notion are being examined and what aspects have been left out of the analysis.

Vocabulary knowledge

Determining what lexical knowledge entails is closely associated with the definition of a *word*. By and large, conceptualizing vocabulary knowledge is directly linked to the answers to two questions: 1) what do language users need to know about a word in order to say that they know it? and 2) what measurable dimensions can be distinguished in the notion of vocabulary knowledge? While the answers to the first question provide the framework of characteristics of word knowledge as related to the notion of word, the answers to the second question outline the dimensions of word knowledge as related to the lexicon as a whole. This distinction corresponds to what Read (2000) refers to as microlevel of word knowledge, i.e. the knowledge associated with the descriptive features of an individual item, and the macrolevel, i.e. the level of the overall state of a learner's vocabulary (p. 248). It should be noted here that the micro- and the macrolevel are so intrinsically interconnected that it is hardly possible to look at them as totally

independent levels. However, it is also important to point out that test instruments designed to measure, let's say, the macrolevel say little about the microlevel of lexical competence and vice versa. Therefore, in light of the suggested distinction, specific attention should be paid not only to whether a test instrument is suitable for the lexical aspect chosen for investigation but also to the extent to which the generalizations made over one of the levels hold true for the other level as well.

Several researchers (e.g., Meara, 1996, Henriksen, 1999, Read, 2000) have stressed the complexity and the enormity of the task to measure all aspects of lexical knowledge. Meara (1996) notes that while theoretically it is possible to specify what it means to know a word and identify all aspects of word knowledge, practically, it is only possible to design a test that can measure a handful of aspects related to a limited number of items. He also suggests that a distinction should be made between assessing how well individual words are known and assessing the overall state of a learner's vocabulary (p. 46). Read (2000) takes this comment a step further by arguing that the distinction between knowledge of individual lexical items and the general state of a person's lexicon should be reflected in the selection of specific types of measures that account for this distinction. That is, when specific vocabulary items are the focus of assessment, what he calls "selective vocabulary measures" (p. 10) should be used. On the other hand, when the focus is on the overall state of vocabulary, a comprehensive measure should "take account of all vocabulary content" (p. 11). In his dichotomous distinction, a selective measure concerns vocabulary tests based on a set of selected test items, whereas, a comprehensive measure applies primarily to language use tasks, e.g. listening, reading, writing, or speaking tasks, where judgements about the quality of learners' vocabulary knowledge are made on the basis of their performance on these specific tasks. In my view, a test instrument that is designed to evaluate the macrolevel of vocabulary knowledge of language learners, regardless of whether it is task-based or individual-item-based, is comprehensive in nature in that it aims at describing the overall size, organization, degree of connectivity between lexical

items, quality of lexical knowledge, etc., i.e. it aims at capturing characteristics that apply to a learner's lexicon as a whole. This understanding rests on the assumption that items tested in isolation, actually, provide information about learners' previous knowledge of words, while with task-based instruments it is difficult to distinguish previous lexical knowledge from strategic skills of using context to infer the meaning of unfamiliar words. Saying this another way, task-based instruments are assessing vocabulary use, whereas word-based instruments are assessing existing knowledge of vocabulary. This is not to say that competence is independent from performance in an absolute sense. It is only to point to their *relative* independence on which the distinction between vocabulary use and vocabulary competence is based.

There are several theoretical proposals of frameworks of lexical knowledge that are widely cited in the field which try to cover in their specifications as many aspects as possible of what lexical knowledge encompasses. They are sometimes referred to as "separate trait" models of lexical competence (Henriksen, 1999, p. 334). Central to these proposals is the question what criteria should be used to fully capture the various aspects of word knowledge. In this regard, researchers have put forward various criteria relating to what learners should know about a word. Aitchison (1987), for example, suggested that a language user should know a word's meaning, its role in a sentence, and what it sounds like. This proposal offers a very layperson's framework of word knowledge and has been further elaborated and refined by other lexical researchers. In this regard, Meara (1996) noted that most of the researchers, who have tried to specify the components of lexical knowledge (e.g., Nation, 1990; Wesche & Paribakht, 1996; Gass & Selinker, 2001) more or less have re-iterated Richards' (1976) proposal of seven main aspects of word knowledge that the scholar identified as follows:

1) to know the probability of encountering a word in speech or writing;
2) to know the limitations of word use according to function and situation;
3) to know its syntactic properties;
4) to know the word's underlying form as well as its derivations;

5) to know the associations between the word and other words in the language;

6) to know the semantic value of the word; and

7) to know many of the meanings associated with the word.

Nation (1990) refined Richard's proposal by suggesting classification criteria that make a distinction between receptive and productive vocabulary. The framework he suggested outlines the following criteria of knowing a word: to know its *form* (spoken and written), its *position* (grammatical patterns and collocations), its *function* (word frequency and appropriateness), and its *meaning* (conceptual content and associations). This framework is widely cited in L2 research but, in my opinion, it raises several fundamental questions if it were to be considered for application in L2 research. For example, since we compare non-NS' vocabulary knowledge to NS', do NS know all the aspects of a word when they say they know it? Is it practically feasible to design a test that can accommodate all outlined criteria? Can we identify a small number of measurable dimensions that reflect properties of the L2 lexicon as a whole? The first two questions often receive an answer in the negative (e.g., Aitchison, 1987), while the third question is usually answered positively. Nonetheless, these issues need to be better addressed in lexical research because we need to know what the overall state of NS' lexical knowledge is before making judgements about L2 learners' lexicons and, more importantly, we need test instruments that will provide a reliable measure of <u>both</u> NS' and non-NS' lexical competence.

Dimensions of lexical knowledge

A second view of the structure of lexical competence is associated with looking at the lexicon as a whole, in terms of dimensions of lexical knowledge. By and large, discussions about the dimensions of lexical knowledge are usually incorporated either in discussions on vocabulary ability or lexical competence. One such effort to describe vocabulary ability was made by Chapelle (1994), who defined vocabulary ability based on a definition about the more general construct of language ability proposed by Bachman (1990). Like Bachman's construct,

Chapelle's definition included "both knowledge of language and the ability to put language to use in context" (p. 163); hence there are three components to vocabulary ability: 1) the context of vocabulary use; 2) vocabulary knowledge and fundamental processes; and 3) metacognitive strategies for vocabulary use. Overall, Chapelle's definition is comprehensive and can be very useful in task-based research on lexical knowledge. However, when lexical items are tested in isolation, only its second component seems to be of some value as a framework of criteria. Chapelle outlines four dimensions of this component:

1) *vocabulary size*, i.e. the total number of words that a person knows, though, not in absolute sense, but rather in relation to a specific context of use;
2) *knowledge of word characteristics*, i.e. the varying degree of knowledge of a word's characteristics (from partial to precise), again, in a context of use;
3) *lexicon organization*, i.e. the way words and other lexical items are stored in the brain; and
4) *fundamental vocabulary processes*, i.e. the processes language learners apply to gain access to their lexical knowledge, both for receptive and productive purposes.

In the context of L2 research, the dimensions suggested by Chapelle (1994) are, in general, theoretically meaningful. However, they seem to have less potential for direct application for practical assessment purposes, especially when vocabulary is tested in isolation (cf. Read, 2000). Also, Chapelle's proposal sounds like an attempt to reconcile lexical competence with lexical performance in one model of lexical ability. However, from a linguistic point of view, the functional dichotomy between competence and performance represents the distinction between our knowledge of language and our use of that knowledge, even though "this autonomy of knowledge from the exercise of that knowledge does not alter the fact that our performance usually provides much of the evidence as to what our competence is" (Smith, 1999, p. 28). Nonetheless, I believe that for an approach to the description of lexical knowledge to be consistent with current linguistic theory it should examine lexical knowledge in light of lexical

competence, rather than lexical ability. In this regard, I support Meara's (1996) view on lexical knowledge which states that "despite the manifest complexities of the lexicon, lexical competence might be described in terms of a very small number of easily measurable dimensions. These dimensions are not properties attached to individual lexical items: rather they are properties of the lexicon considered as a whole" (p. 37). As to what these dimensions might be, there are several proposals that have been extensively discussed in the L2 literature. In the following paragraphs, I will concisely present some of the suggested dimensions in order to set the theoretical background stage for the three-dimensional model that will be proposed in Chapter 2.

Vocabulary size

There seems to be unanimous agreement in the field that one of the dimensions that can well describe lexical knowledge is vocabulary size, or quantity of lexical knowledge, also referred to as *breadth* of vocabulary. There is documented interest in estimating what the vocabulary size of a NS of English might be that dates back to 1891. By evaluating his own size of vocabulary, Kirkpatrick (1891, cited in D'Anna et al., 1991) estimated that educated college graduates most probably have a vocabulary size of 20,000 to perhaps 100,000 words, while ordinary US citizens are likely to have about 10,000 words in their vocabularies. Following this attempt, there have been a host of other studies (e.g., Seashore & Eckerson, 1940; Hartmann, 1941; Oldfield, 1963; Diller, 1978; Goulden et al., 1990; D'Anna et al., 1991; Zechmeister et al., 1995) which obtained quite different estimates of vocabulary size. Nonetheless, all these studies have been motivated by finding answers to several major to the language acquisition questions, such as: what is the size of the vocabulary learning task facing each language learner? Is there a place for direct instruction in increasing vocabulary size? How does vocabulary size affect the academic performance of language users? In the context of L2 teaching and learning, these questions are slightly modified and the emphasis of vocabulary studies is on finding out what minimum number of words L2 learners need to know in order to cope with the language

demands of their studies (Read, 2000)? How close does non-NS' vocabulary size come to that of NS? How many words should L2 learners be expected to know at different levels of language proficiency? For the most part, researchers who aim at estimating the vocabulary size of L2 learners have done so in the understanding that knowing an adequate number of words is a prerequisite for effective language use; that learners whose vocabularies are below a specific threshold are likely to struggle through academic tasks; that examination of L2 learners' lexical deficits should inform teaching practices. With all these considerations in mind, vocabulary size, or *breadth* of vocabulary knowledge, has been for a long time the only dimension of lexical competence that has received considerable research attention. Nonetheless, in the last couple of decades L2 researchers began to realize that knowledge of words should not be studied unidimensionally, i.e. with respect to size only, but other possible dimensions should be attempted to account for.

Depth of vocabulary knowledge

A second dimension of the lexicon that is often posited in the literature is quality, or *depth*, of lexical knowledge. By and large, this is the most controversial dimension primarily because there is little agreement among L2 researchers what quality, or *depth*, means. This dimension has been first suggested by Dolch and Leeds (1953) in L1 research and later has been adopted in L2 research and modified by several researchers (e.g., Anderson & Freebody, 1981; Read, 1993, 2000; Wesche & Paribakht, 1996; Henriksen, 1999; Wolter, 2001; Vermeer, 2001). By studying the vocabulary sub-tests of several major reading and achievement test batteries for American school children, Dolch and Leeds (1953) expressed concern about the potential of the existing tests to measure children's knowledge of word meaning. They found that the tests probe only for the most common meaning of each SW, so the researchers concluded that these tests had no potential to assess whether subjects knew any other SW meanings because they were required to simply identify the synonym of each SW, without showing any evidence that they knew the meaning of the synonyms or the difference between

the synonymous meanings. So, in an attempt to remedy the situation, Dolch and Leeds (1953) designed test items that were intended to measure what they called *depth of meaning*. They defined *depth* thus:

> The homonyms, derived meanings, and figurative meanings just mentioned may be thought of as degrees of meaning beyond the most common meaning. But the most common meaning has also degrees of meaning, or different depths of meaning. Our question is, how well do vocabulary tests measure depth in the case of this most common meaning? (Dolch & Leeds, 1953, p. 184)

The excerpt shows that Dolch and Leeds' understanding of depth of vocabulary knowledge included several aspects of word meaning as studied by lexical semantics, i.e. homonymy, derived meanings (synonymy, polysemy, hyponymy), and figurative meaning. Likewise, the researchers argued that these aspects can successfully provide a measure of depth of knowledge of the commonest meaning, which was usually tested by the existing tests.

Anderson and Freebody (1981) further elaborated on this proposal and introduced in the literature the metaphorical dichotomy "breadth and depth of word knowledge". In their words:

> It is useful to distinguish between two aspects of an individual's vocabulary knowledge. The first may be called "breadth" of knowledge, by which we mean the number of words for which the person knows at least some of the significant aspects of meaning… [And] a second dimension of vocabulary knowledge, namely the quality or "depth" of understanding. We shall assume that, for most purposes, a person has a sufficient deep understanding of a word if it conveys to him or her all of the distinctions that would be understood by an ordinary adult under normal circumstances. (Anderson & Freebody, 1981, p. 92-93)

The authors further explained that the proposed dichotomy was motivated by psychological research about how children accumulated features of meaning as a function of age, i.e. the idea of *depth* of meaning was based on research findings that there is progressive differentiation of

word meanings as a result of increasing age and experience. While from developmental point of view reasonable, the description of *depth of meaning* was, apparently, not intended to serve the practical purpose of examining *depth* in measurable terms. As a matter of fact, the working definition of *depth* barely sketches what it refers to and doesn't say much about the factors, besides age and experience, that are contributive to deepening understanding of word meanings. Overall, the authors were convinced that knowledge of words continues to deepen throughout lifetime; though, as they admitted, "there is no hard data to support this conjecture" (Anderson & Freebody, 1981: 94).

Nevertheless, the metaphor of *breadth and depth* of word knowledge largely appealed to the L2 lexical research community and many researchers began to use this distinction in their writings (e.g., Read, 1993, 2000; Wesche & Paribakht, 1996; Henriksen, 1999, Wolter, 2001; Vermeer, 2001, Greidanus & Nienhuis, 2001). Unfortunately, what has not been clearly spelled out yet is what exactly *depth* of word meaning encompasses and how *depth* can be measured. One suggestion of what *depth* of vocabulary knowledge entails comes from Wesche and Paribakht (1996). They proposed a Vocabulary Knowledge Scale (VKS) as an instrument that has the potential to capture the *depth* of learners' lexical knowledge in terms of varying degree of knowledge of words. Another proposal comes from Henriksen (1999), who distanced herself from Wesche and Paribakht's VKS of assessment of *depth* of knowledge on the basis of partial - precise knowledge of words. Instead, she explained that her understanding of *depth* of knowledge is more closely linked to the aspects of knowledge as outlined by Richards (1976) and Nation (1990).

Recently, most of the attempts in examining quality of word knowledge have employed association tests as a means of tapping the intentional aspects of L2 learners' meaning representation. One of the basic assumptions underpinning word association tests is that NS have remarkably stable patterns of word associations, which reflect the complex lexical and semantic networks in their mental lexicon (Read, 1993). On the other hand, L2 learners tend to

have more diverse patterns of responses than NS maintain, though, there is evidence that as learners' proficiency increases, their patterns of responses develop towards NS' norms (e.g., Meara, 1978; Vermeer, 2001; Greidanus & Nienhuis, 2001). So, researchers using association tests in their studies tend to share common understanding about the value of the associations provided by non-NS as revealing of their associative patterns of meaning. However, a major difference between the association studies concerns the interpretation of the association patterns as indicators of quality of vocabulary knowledge. Schmitt (1998), for example, proposed that the attribute on which quality of L2 responses should be judged is the degree of *native-likeness* their associations share with those of NS. In his analysis, the most common response pattern provided by his norming group was three (54%) and two (34%) group responses, which made him argue that this high level of commonality among NS' responses (88%) is an indicator of *native-likeness of association behavior* against which non-NS' associations should be matched. In the main, his interpretation of association responses follows an earlier traditional of interpretation of associative behavior as an ability to produce native-like associations to L2 words (e.g., Meara, 1982; Kruse, Pankurst, & Sharwood-Smith, 1987).

A second way of interpreting patterns of association responses is strongly linked to association studies conducted by Meara (1978, 1984). Overall, Meara (1996) finds the term *breadth and depth of vocabulary knowledge* confusing in the context of assessment of lexical knowledge. Instead, he proposes to label the two dimensions simply as vocabulary size and quality of vocabulary knowledge, where quality should be linked to the networks of associations that language users develop, which are revealing of the organization of their mental lexicon. In his view, it is important to examine how words are organized in L2 learners' mental lexicon with respect to whether they are organized in semantic networks that are similar to the ones that NS have or whether L2 learners simply tag L2 words into their L1 lexicon. In my view, looking at associations this way allows us to find out whether the patterns of word associations produced by non-NS are broadly comparable with the ones produced by NS, for example. Such an

analysis can also reveal whether there are structural overlaps between NS and non-NS as proficiency increases or whether non-NS' lexicon is "fundamentally different" (Wolter, 2001) from NS'. Several researchers (e.g., Read, 1993; Wolter, 2001; Vermeer, 2001; Greidanus & Nienhuis, 2001) adopted this point of view in their research and associated quality or *depth* of lexical knowledge (both terms are very often used interchangeably without any further specification) with the study of the patters of paradigmatic, syntagmatic, or phonological (clang) relationships that L2 learners build to organize their mental lexicons. In the experiment that will be discussed in the next chapters, I will examine the validity of both points of view to show how the analysis of native-likeness of associations as well as the study of the qualitative and quantitative patterns of associations contribute to our better understanding of the organization of the L2 mental lexicon.

Receptive-productive dimension

The distinction between receptive and productive vocabulary is very often proposed in the literature as a third dimension of vocabulary knowledge. In general, L2 researchers agree that there is a difference between receptive and productive vocabulary, i.e. between vocabulary used for comprehension and vocabulary used for production. Whether the ability to use a word is described as knowledge (e.g., Nation, 1990) or as control (e.g., Henriksen, 1999), researchers are, generally, in one mind that word comprehension does not automatically predict correct use. The number of words that can be recognized and understood in both L1 and L2 is considered to be larger than the number of words that are used in speech or writing. So, the distinction between receptive and productive vocabulary is generally acknowledged by lexical researchers, who also refer to it as the distinction between *passive* and *active* vocabulary (e.g., Laufer & Paribakht, 1998). It is also generally assumed that words are first known receptively and only later become available for productive use (e.g., Read, 2000). Also, most researchers have assumed that passive vocabulary is larger than active (e.g., Aitchison, 1989). However, it is not clear what specific criteria can account for the difference between both types of

knowledge, especially since researchers propose differing criteria. As pointed out by Read (2000), the difficulty with conceptualizing the distinction between reception and production lies in the different status that words have in one's lexicon, i.e. some words have just a receptive status, while others are part of a person's productive vocabulary. However, it is important to note that the receptive - productive distinctions not a dichotomous one; rather, it refers to a receptive - productive continuum, which represents an increasing degree of receptive and productive control of word knowledge (e.g., Melka, 1997; Henriksen, 1999; Read, 2000). The problem is to find the threshold at which a word passes from receptive to productive status (Read, 2000). In this regard, Melka (1997) suggests *word familiarity* to be the threshold for the receptive - productive distinction, i.e. receptive knowledge of vocabulary can become productive dependent on a learner's familiarity with a word. Importantly, in her account, this is not an either-or situation, i.e. vocabulary knowledge is not viewed as either receptive or productive; rather, some aspects of word knowledge are assumed to become productive, while others remain at a receptive level only. Other researchers see the distinction between receptive and productive vocabulary differently. Meara (1996), for example, proposes *automaticity* for a criterion to be used to distinguish between productive and receptive vocabulary, where the degree of automaticity that has been developed determines the distinction between receptive and productive knowledge. Laufer and Paribakht (1998), on the other hand, identify productive vocabulary with vocabulary that can be actively used, and receptive vocabulary with "passive" use, where *word frequency* is the factor that motivates movement of lexical items from the passive to the active domain. Read (2000) offers a narrower distinction by arguing that reception and production are too broad terms; therefore, researchers should be specific about what is actually studied, i.e. whether it is recognition, recall, comprehension or use. He further comments that recognition and recall are often used in experimental research as criteria to study whether learners are able to recognize a word and whether they are able to recall it when

prompted. However, he emphasizes that these types of techniques should not be identified with assessment of comprehension or use.

Conclusion

In sum, the discussion about lexical competence comes to highlight that vocabulary knowledge is not a simple construct to examine. Lexical knowledge is a complex phenomenon that falls on a continuum in all of the above discussed dimensions, where we are dealing with varying degrees of knowledge of words rather than with either-or instances of knowledge. Similarly, "knowing a word" is a multifaceted task for all language users and, consequently, examining what speakers know about every lexical item they have in their lexicons is a "mammoth task" for a test designer and "completely unmanageable for a testee" (Meara, 1997: 46). Therefore, having a model of lexical knowledge would allow researchers to study lexical competence in a model-based framework and design test instruments that would adequately measure its aspects.

In L2 research, there are two traditions of modeling the lexicon: a separate trait tradition, which proposes studying knowledge of words with regard to the lexical aspects of describing a word, and global trait tradition, which takes a more comprehensive approach to examining lexical competence in its entirety. Researchers within the separate trait paradigm put forward a set of descriptive criteria of what it means to know a word, whereas researchers favoring the global approach propose two (*breadth and depth*) or three global dimensions (quantity, quality, and control) for describing the overall state of L2 learners' vocabulary knowledge. According to Meara (1996), the first two dimensions offer a rich framework for describing different aspects of lexical competence and suggest some interesting relationships between vocabulary growth and connectivity of the lexicon. He also argues that the size dimension becomes less important as the size of vocabulary gets larger and that the quality dimension takes precedence after a threshold of 5,000 – 6,000 words is passed. Therefore, given that L2 learners' vocabulary size is considered to be generally smaller compared to the vocabulary size of NS, the importance of it

being structured considerably increases. In this regard, Meara (1996) hypothesizes that L2 learners whose vocabularies are structured would be better performers on most real life tasks than learners with less structured vocabularies of an equivalent size. Many researchers have adopted the two-dimension model consisting of quality and quantity of lexical knowledge. Henriksen's (1999) proposal for a third dimension, in addition to quality and quantity of lexical knowledge refers to the distinction between receptive and productive knowledge, i.e. it reflects how well a learner can access and use a lexical item. However, the three-dimensional model is largely at the stage of a hypothesized framework, whereas the two-dimensional model has received some empirical support. In my thesis I also propose a three-dimensional model of studying the lexicon, which, however, differs from Henriksen's framework in that it recognizes the distinction between receptive and productive vocabulary to apply to both knowledge of the lexical features of individual words as well as to the lexicon as a whole. A full description of the model is provided in Chapter 2.

Overview of the thesis

Overall, in my view, an approach to exploring L2 lexical competence in its entirety makes studying lexical knowledge more attainable a goal; therefore, the approach taken in this research is within the global trait tradition. The study bridges two traditions of examining the lexicon, applied linguistic and psycholinguistic tradition, by investigating the lexicon from practical and structural (psycholinguistic) point of view. The overall goal of this experimental research is examining adult NS' and L2 learners' lexical knowledge within a model-based framework of the lexicon and finding evidence that with an increase in language proficiency the L2 mental lexicon stabilizes and starts to resemble NS' lexical knowledge. No other study in the field, to my knowledge, has previously investigated what the quantitative, qualitative, and metacognitive differences are between non-NS' and NS' lexical knowledge as a function of proficiency. In addition, I will try to challenge some metaphors and assumption about L2 lexical

competence, directly borrowed from L1 research, that have become so convenient to use in the L2 research field that they do not seem to be questioned any longer. Yet, metaphors are powerful comparisons that may obscure or even change our insights about phenomena that are much more complex than a metaphor can express, so they need to be scientifically tested and re-defined, if necessary.

Chapter 2 and 3 offer an applied linguistic perspective on studying the lexicon by focusing on the dimensions that have the potential to capture the general state of NS' and L2 learners' lexical knowledge and the factors that can best account for the greatest variance among language users at different levels of language proficiency. In Chapter 2, I introduce my proposal of a model of lexical knowledge, which distinguishes within the receptive and productive domain two levels of lexical knowledge, a microlevel and macrolevel. The distinction that is made in the model serves the single purpose of distinguishing between what assessment of knowledge of individual words entails and how it differs from assessment of the overall lexicon. Moreover, while the microlevel has received ample research attention, there are still many questions concerning the macrolevel of L2 lexical competence that need to be addressed in L2 research. That is why I chose to examine the macrolevel of vocabulary knowledge in my research and the rest of the chapter discusses an experiment specifically designed to test the significance of the proposed macrolevel dimensions by involving participants at several levels of language proficiency. The results of the study indirectly challenge the metaphor that "breadth and depth" of lexical competence cover it all.

Chapter 3 examines the interaction between several factors used in the experiment to measure the lexical competence of both NS and L2 learners in an attempt to identify the smallest set of factors that can account for the greatest amount of variation in the lexical knowledge of adult NS and non-NS of English. I use three procedures of regression analysis, i.e. stepwise method of variable selection, direct search on "t" method, and all possible regression method to find out which combination of factors has the greatest potential to predict

the overall state of the vocabulary of participants at different levels of language proficiency. The analyses generally aims at identifying which variables associated with specific dimensions of lexical competence do not add any meaningful increments in the variation explained, while dramatically increasing the bias of the models. By comparing the results of the three methods of regression analysis, I propose models consisting of a small set of factors that are meaningful for assessment purposes in addition to being cost and time-efficient for testing. To my knowledge, such a global analysis of combinations of factors for predictive purposes has not been conducted in the field.

Chapter 4 takes a different tack by shifting the focus from applied linguistic issues to psycholinguistic issues. By and large, I believe that associative behavior holds a great value for learning more about how non-NS' qualitative and quantitative patterns of associative meaning compare to those of NS. Analyzing WA data by using "nativelikeness" of associative commonality as a measure on which judgements about the quality of non-NS' lexical knowledge are made has been a commonly used approach in L2 research. However, the analyses in the previous chapters reveal that it is an unproductive approach to investigating WA behavior which is not sensitive to degrees of language proficiency but only to the broad distinction between NS and non-NS. Therefore, in this chapter I look at the WA data generated by the participants in the study in terms of patterns of associative connections that language users tend to build in their mental lexicons to organize meaning relationships among the words they have at their disposal. More specifically, I compare the quantitative as well as qualitative characteristics of NS' and non-NS' WA domains with regard to the effects of participants' increased language proficiency and varying degree of familiarity with words.

Chapter 5 focuses on whether or not characteristics of the SW, in particular lexical category and frequency of occurrence, affect the qualitative and quantitative features of the WA domain of NS and non-NS. I decided to examine these effects because, on the one hand, I believe that knowing more about how word characteristics affect meaning connections would

raise researchers' awareness of the importance of a careful SW selection in empirical research. On the other hand, to my knowledge, these effects have not been researched in L2 studies, with the notable exception of Söderman (1993), who only looked at the effects of high and low frequency SW on the qualitative association patterns of NS and advanced learners of English. Therefore, the purpose of this analysis is to shed light on issues concerning the impact of lexical category and frequency of occurrence of SW on the features of the WA domain of NS and non-NS considering participants' level of proficiency and degree of familiarity with words.

Finally, Chapter 6 summarizes the findings of the studies and their contribution to the field.

CHAPTER 2

RELATIONSHIP BETWEEN MACROLEVEL OF LEXICAL COMPETENCE AND LANGUAGE PROFICIENCY

Introduction

The study of the dimensions of L2 word knowledge and their relationship to vocabulary growth have recently attracted a lot of interest among L2 researchers. As interest in vocabulary growth developed, researchers have began to show increasing awareness that knowledge of words is not unidimensional and should be studied at several dimensional plains. There is also growing realization in the field that the role of each dimension in the lexical knowledge framework strongly depends on the proficiency of the L2 learners, among all other factors. Along these lines, one of the goals of vocabulary research began to be seen in capturing the importance of each lexical dimension with respect to the language proficiency of L2 learners in order to develop test instruments that would reliably measure different aspects of word knowledge. In this regard Read (2000) points out that an aspect that needs to be further researched is the role of different measures in making decisions about L2 learners' lexical needs (p. 248). In his view, an issue that has not received sufficient attention in L2 lexical research yet is the relationship between language proficiency and the overall state of a learner's vocabulary. It is generally assumed that the higher the proficiency level, the better the general state of lexical knowledge of L2 learners is. However, little is known for example at what proficiency level L2 learners' organization of the mental lexicon stabilizes; where the differences and similarities between NS' and non-NS' mental lexicons should be looked for; what proficiency level can be identified as a cutoff level, at which variation in the lexical knowledge

among L2 learners, compared to NS, affects all dimensions. These are only few of the questions that deserve to be further examined in light of a well-conceptualized framework of levels and dimensions of studying lexical competence. Such an approach would allow for more clarity in the interpretation of the results of L2 vocabulary studies and their compatibility across L2 lexical research.

Conceptualizing lexical knowledge is not an easy task. In Chapter 1 I outlined some recent proposals of criteria of studying lexical knowledge that are widely cited in the literature. They are all based on the understanding that "knowing a word" is a multifaceted task for language users and the assessment of that knowledge is a "mammoth task for the test constructor" (Meara, 1996, p. 46). The enormity of the task resides in the large number of aspects associated with the notion of a "word" and, respectively, with the various aspects of what knowing a word entails. Some researchers (e.g., Richards, 1976; Nation, 1990; Wesche & Paribakht, 1996; Gass & Selinker, 2001) have proposed models of lexical knowledge consisting of separate traits (also known as *separate trait models*). This idea is best summarized by Nation (1990), who listed the following aspects of what knowing a word encompasses:

- *Form:* 1) spoken form
  2) written form
- *Position:* 3) grammatical behavior
  4) collocational patterns
- *Function:* 5) word frequency
  6) appropriateness
- *Meaning:* 7) conceptual content
  8) word associations

The problem with the proposed descriptive criteria is mostly practical. As noted elsewhere in the literature, while theoretically it is possible to describe what it means to know a word, practically,

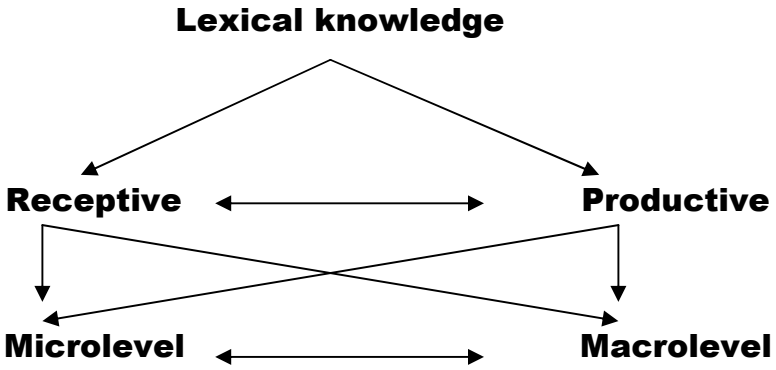
it is hardly possible to design a test that would measure all traits, as mentioned above, for every word an individual knows. Therefore, this practical concern calls for a distinction to be made between assessment of how well a particular set of words are known and assessment of the overall state of a learner's vocabulary. in this regard, having a small number of measurable dimensions that reflect properties of the lexicon as a whole, rather than properties of individual items, would make it possible for researchers to examine language users' general state of the lexicon in its entirety (cf. Meara, 1996).

Currently existing models that promote several global characteristics (dimensions) for description of the lexicon as a whole are referred to as *global trait models*. In L2 research, global trait models of vocabulary knowledge propose two or three dimensions of examining L2 learners' lexicon. As for the amount of support each dimension receives from the research community, there seems to be a great deal of agreement among researchers on the importance of vocabulary size (or *breadth*) as a dimension of lexical competence. There also seems to be agreement that a second essential aspect of lexical competence is quality (or *depth*) of word knowledge; though, L2 researchers have proposed several different interpretations of what quality or "depth" entails. Finally, there is a third dimension that is sometimes posited (e.g., Henriksen, 1999), receptive-productive dimension, which seems to build the bridge connecting lexical competence with performance.

In sum, the distinction between global trait and separate trait models of lexical competence is made essentially for practical reasons. For exactly the same reason, I shall propose a different model of studying lexical knowledge, which would be also empirically tested in an experiment specifically designed to investigate the advantages of the model. My major concerns behind the proposal relate to the realization that it is not enough to only consider how suitable a test design is for the lexical aspect chosen for investigation, but it is equally important to draw realistic conclusions, within a specific model, about the extent to which the generalizations made over one level would hold true for the other. Also, a model-based

investigation allows for looking at relationships between dimensions that, otherwise, may remain completely unaccounted for. Finally, I believe that empirical research should be model-based, because a model, even an imperfect one, shows what has been included in the analysis and what has been left out.

The following diagram summarizes the major premises of the proposed model for studying and assessment of lexical knowledge.



(*"grammar"* of individual words) (overall state of L2 learners' lexicon)

- as a phonetic/ an orthographic entry
- as a phonological entry
- as a morphological entry
- as a syntactic entry
- as a semantic entry

- quantity
- metacognitive awarensss
- quality

The model outlines two major distinctions that need to be made in the study of lexical knowledge. The first major distinction is between receptive and productive vocabulary, i.e. examining word knowledge with regard to learners' ability to recognize a word, as well as with respect to their ability to use it productively. The second major distinction that is shown in the model is the distinction between microlevel and macrolevel of word knowledge, which applies to

both the productive and the receptive domains. As the diagram shows, the receptive and the productive domain, as well as the micro- and the macrolevel are not completely independent of each other. On the contrary, they constantly interact, exchange information, and inform each other. Therefore, the distinction that is being made here serves the single purpose of distinguishing between what assessment of knowledge of individual words entails and how it differs from assessment of the overall lexicon.

The microlevel refers to what has been called *grammar* of individual words, where *grammar* is used in its broadest sense. Assessment at a microlevel (receptive or productive) can successfully be applied to a limited number of lexical items in research designed to focus on examining how well learners know or have learned a particular set of words. It should be emphasized, though, that a selective measure of this kind does not hold the potential to reveal much about the general state of a learner's vocabulary as a whole. The macrolevel of lexical knowledge in the model is the level that describes the overall state of an individual's lexicon within a framework of three dimensions: quantity, quality, and metacognitive awareness. While the role of quantity and quality of word knowledge has been extensively discussed in the literature, little mention has been made about the place of metacognition in lexical competence. Bialystok and Sharwood-Smith (1985), for example, draw a distinction between *knowledge* and *control* of vocabulary. Knowledge is defined as "the way in which the language system is represented in the mind of the learner", whereas control is defined as "the processing system for controlling that system during actual performance" (p. 104). Gass (1988) rightly points out that this distinction is very useful when applied to vocabulary "since it crosses the boundaries of more traditional notions of productive and receptive knowledge. Both production and reception include information regarding knowledge and control" (p. 95). I need to note here that I use the term *metacognitive awareness* as a more general term that refers to "one's knowledge concerning one's own cognitive processes and products or anything related to them" (Flavell, 1976, p. 232 cited in Gombert, 1992, p. 7) or, put even more succinctly, "cognition about

cognition" (Gombert, 1992, p. 8). I should also add that, in my view, this dimension performs a bridging function between the micro- and the macrolevel of knowledge, i.e. it allows language users to monitor what aspects of the grammar of every word they know when they say they know a word.

While the microlevel has received ample research attention there are still many questions concerning the macrolevel of lexical competence that need to be addressed. Thus, the present study has been designed to empirically test the significance of the proposed macrolevel dimensions of lexical knowledge assessment at several levels of language proficiency. The study will aim at finding out whether with an increase in language proficiency, the macrolevel of L2 lexical knowledge stabilizes and starts to resemble NS' macrolevel of lexical competence. Moreover, to my knowledge, no previous study has investigated that matter with regard to more than one proficiency level; though, the general assumption has always been that language proficiency is positively related to L2 learners' overall state of vocabulary. In several studies examining the vocabulary size of L2 learners, it has been found that the number of words learners know is linked to their level of proficiency (e.g., Meara & Jones, 1988; Read, 2000). However, vocabulary size accounts for only one dimension in the proposed framework of lexical knowledge, which, in fact, says little about what L2 learners know about lexical items when they claim they know them. Also, getting better understanding of how the lexical knowledge of L2 learners at various stages of language proficiency compares to that of NS of English on the three posited dimensions will give us valuable insights into the features of their general state of vocabulary. To this end, several research questions were of interest to the researcher: i.e. 1) How is the quantity of L2 lexical knowledge different or similar to the L1 knowledge as a result of an increase in language proficiency? 2) How does the quality of L2 learners' vocabulary compare to that of NS as a function of language proficiency? 3) How does the metacognitive awareness of language users at different levels of proficiency relate to the overall state of their lexicons? 4) Do the three dimensions fully capture the general state of

language users' knowledge of words? The answers to some of these questions will also provide us with valuable insights into the nature of multicompetence and the way it influences lexical competence.

The rest of the chapter presents an experiment designed to test the three-dimensional model of macrolevel assessment of the lexical competence of adult NS and non-NS of English. In the section that follows, I describe the methodological background of the experiment. I discuss problematic issues linked to dictionary word counts and stimulus word (SW) sampling procedures in great detail in order to justify the decisions I made concerning specific inclusion and exclusion criteria. I also offer a comprehensive analysis of the selected SW and the test format used in the study. Finally, the section presenting the results of the statistical analyses is followed by a discussion and comments on the importance of the findings of the experiment and their relationship to the research questions.

Method

Participants

Sixty-four adults, both NS and non-NS of English, representative of a sample of normal educated adults, participated in the study. The NS of English (n = 30) were undergraduate students from different majors at the University of Georgia, who were enrolled in an introductory course in linguistics at the time of the experiment. Both male (n = 9) and female (n = 21) NS of English, age 18 and above (M = 19.7) participated in the study. All participants in the group of NS were asked whether English was their first language and only the results obtained from NS were included in the data. They were also asked to rate their knowledge of a second language, if they had any, on a 5-pont scale (0 = no knowledge, 1 = beginners', 2 = lower intermediate, 3 = intermediate, 4 = upper intermediate, 5 = advanced). The mean ratings ranged from 0 to 1.65, which showed a low level of knowledge of another language that should not be expected to influence their performance on the test.

The group of non-NS consisted of Bulgarian L2 learners of English at two levels of language proficiency: advanced (n = 17) and intermediate (n = 17), who at the time of the experiment were attending respectively an advanced and intermediate certificate course in English at Pharos Private School for Foreign Languages in Bulgaria. Their group mean age (M = 21.5) was non-significantly higher compared to the mean age of the NS. Both genders, males (n = 14) and females (n = 20), were represented. All participants were NS of Bulgarian who have completed their high school education in Bulgaria and have received formal education in English in Bulgaria.

The proficiency level of the participants was determined on the basis of their performance on two tests, which Pharos School regularly uses to assess the progress of the students who attend their certificate preparation classes, i.e. Cambridge Certificate in Advanced English (CAE) test and Cambridge First Certificate of English (FCE) test. CAE is an advanced exam, provided by Cambridge ESOL, a department of the University of Cambridge in England, which is linked to the Council of Europe's Common European Framework for modern languages. The exam is set at Level 4 of an international five level scale, established by the Association of Language Teachers in Europe (ALTE), and was used in the present study to identify the proficiency level of the advanced group of participants. CAE recognizes the ability of fully operational command of the language in a wide range of real life situations and the certificate is accepted by almost all universities in the United Kingdom, as well as by a growing number of employers world-wide. The exam is based on realistic tasks aiming at the assessment of test-takers' abilities to successfully perform in a variety of reading, writing, use of English, listening, and speaking tasks. The second exam used in the experiment to identify the language proficiency of the intermediate L2 learners was FCE. It is an exam that is set at Level 3 of the 5-level ALTE scale and is recognized as an (upper) intermediate level exam, designed for learners whose command of English is adequate for many practical everyday purposes. The structure of the exam is similar to that of CAE; however, the degree of difficulty of the tasks is

appropriate for a lower level of language proficiency. Both CAE and FCE offer a very detailed specification of the quality of the skills a successful candidate should be able to demonstrate in order to be awarded a certificate of language proficiency by Cambridge ESOL. So, given the international recognition of the exams and their reliability as an overall proficiency measure, I believe that the language proficiency level of L2 participants was established dependably. Official TOEFL scores were also accepted as a proficiency level indicator for several of the participants.

The scores on the *Reading* and *Use of English* sections of CAE and FCE were only taken into consideration, with a lowest passing grade of D (between 50% and 60% correct answers) on both sections cumulatively. For the advanced group the mean on the CAE *Reading section* was 70.1% (range 52% - 85%) and the mean on the CAE *Use of English* section was 62.7% (range 42% - 80%). The results of the intermediate group were as follows: FCE *Reading* mean 72.5% (range 66% - 82%) and FCE *Use of English* mean 56.2% (range 44% - 67%). Since lexical knowledge in written form was going to be tested, the researcher did not take into account participants' scores on the *Speaking* and *Listening* sections of both tests. In addition, seven participants were placed in the L2 advanced group based on their official TOEFL scores. Five participants had a TOEFL score over 560 (M = 581, range 561 – 620) on a pencil-and-paper test format and two participants had a mean score of 259 on a computer-based test format. All tests for determining language proficiency were taken in controlled settings.

The three groups completed the test used in the experiment as a take-home test. By not limiting the time for test completion, the researcher wanted to avoid the effects of fatigue by allowing the participants to work at their own pace. The instructors of the three groups took special care of explaining how the test should be completed by going over the test instructions in class and giving several examples of how test items should be approached. The instructors also made sure that the participants understood they were expected to complete the test honestly, without using a dictionary or any other reference material of that nature. The group of

NS served as a control group against which the results of the L2 advanced and intermediate learners of English were compared.

Materials

Sampling Procedure

Word Source. The three groups of participants completed a discrete test of vocabulary knowledge based on the assumption that for assessment purposes vocabulary knowledge can be treated as an independent construct, separate from the other components of language competence (Read, 2000, p. 8); yet, not unrelated. In this study, I evaluated the lexical knowledge of non-NS by using a methodology for the selection of SW that is frequently used in L1 lexical research (e.g., Anglin, 1993; Johnson & Anglin, 1995) as well as in L2 studies (e.g., Goulden, Nation, & Read, 1990) known as a spaced sampling procedure. The procedure involves selecting items from a randomly determined starting point in a dictionary, taking words at a specific interval.

The words were selected from *Oxford Student's Dictionary of Current English* (*OSDOCE*) (Hornby, 1978), which the publisher claims to contain 35,000 words and phrases and 50,000 example phrases and sentenses. This particular dictionary offers several advantages for vocabulary research involving non-NS of English. First, the dictionary is a fairly recent adaptation of the *Oxford Advanced Learner's Dictionary of Current English* (Hornby, Gatenby, & Wakefield, 1963) and reflects contemporary British English and American English usage. This was important to the study because the English language taught in Bulgaria is predominantly British English, though American English is fast gaining a place in the language instruction. Secondly, the dictionary is compiled to be used by learners of English as a second or foreign language and it contains both British English and American English examples of words, definitions, derivations, compounds, and common collocations. Thirdly, the dictionary is a product of the British lexicography, which has a long history and tradition in compiling ESL

and EFL dictionaries, not to mention that the author's name has for a long time been synonymous with dictionaries for English language learners (Landau, 2001, p. 74). The latter comments can be taken as an implicit recommendation for a good quality dictionary, which is particularly relevant for use at advanced and intermediate level of language study. Next, though the dictionary is concise, compared to larger unabridged dictionaries, it is comprehensive and inclusive of a wide range of vocabulary usage, e.g. formal, informal, slang, poetic, technical, foreign, etc., besides being useful and valuable as an educational tool. Finally, choosing a dictionary that has not been previously used in lexical research for assessment of vocabulary knowledge will allow for comparison of the results from this study with other studies, which also used a spaced sampling procedure for their word selection but from other dictionaries.

Following D'Anna, Zechmeister, and Hall (1991), I obtained information regarding the comprehensiveness and usefulness of the dictionary for academic purposes by determining the number of words in two lists of English vocabulary that appeared in the dictionary. One of the lists was Barron's *Pocket Guide to Vocabulary* (Brownstein, Weiner, & Green, 1996) designed to be used for preparation for TOEFL and SAT, thus, compiled primarily with low-frequency words that appear in those tests. The second list was *The Teacher's Word Book of 30,000 Words* (Thorndike & Lorge, 1944), which is a widely used reference source of frequency of occurrence of words in approximately a 4.5-million-word corpus. I randomly sampled 100 words from each of those word lists in order to determine the percentage of the sampled words or their variant forms that appear in *OSDOCE*. Of the words that were sampled from Barron's *Pocket Guide to Vocabulary,* 78% appeared in the same form in *OSDOCE* and 3% were found in a different form (e.g., *penurious* vs. *penury*, *indubitably* vs. *indubitable*, etc.). From *The Teacher's Word Book of 30,000 Words,* I obtained a sample from the list of words that occurred at least once per 4 million. It was found that 76% of those words had the exact same listing in *OSDOCE* and 8% were listed in a variant form. These results allowed me to conclude that *OSDOCE* offers

a meaningful definition of a corpus of contemporary English vocabulary on which the assessment of lexical knowledge of well-educated adult non-NS can be based.

Word count exclusion criteria. As Nagy and Anderson (1984) convincingly argued, the sampling and evaluation of tests that aim at measuring vocabulary size, among other variables, largely depends on the answers to three questions, namely:

1) What words are included or excluded from the count?
2) What words are grouped together and what words are treated as separate entries?
3) What is the size of the dictionary from which the sample is obtained?

With these fundamental questions in mind, I performed an independent count of the dictionary entries based on specific criteria of exclusion of entries from the count and, respectively, from testing. The following entries were excluded from the dictionary count when listed as separate boldface main entries:

1. letters of the alphabet and names of letters (e.g., *A, B, alpha, beta,* etc.);
2. affixes (e.g., *a-, ambi-, bio-, co-, -ary, -an, -ly,* etc.);
3. capitalized proper words and their derivations (e.g., *Baptist, Catholic – Catholicism, Arab – Arabic, America – American – Americanism,* etc.);
4. abbreviations, including those not followed by a period (e.g., *asp , amp, chimp, mag, doc*, etc.);
5. contracted forms (e.g., *aren't*, *isn't, won't, shan't* etc.);
6. interjections (e.g., *eh, ha, hello, heck, eureka*, *oops*, etc.);
7. words with an alternate spelling (e.g., *burr = bur, color = colour, center = centre*, *milage = mileage*, *cullender = colander*, *disc = disk*, etc.);
8. regularly inflected forms (e.g., *cry - cried - cries, dry – drier - driest*, etc.).

These exclusion criteria are based on the proposals made by Goulden at al. (1990) about the use of spaced sampling procedures for vocabulary testing. The authors argue that one way of avoiding the striking differences in the estimates of vocabulary size that exist across

studies is to consistently apply unified criteria to the choice of items that will be included in the dictionary word count and, respectively, for testing. Attractive as this proposal may sound, it is very slowly finding its way in the sampling methodologies of lexical research studies. Nonetheless, it has certainly raised researchers' awareness that obtaining a reliable estimate of the number of words in the dictionary from which the test items will be selected is crucial to the interpretation of the final results of a study. This becomes particularly important in light of the fact that decisions related to how words are counted in a dictionary become a major factor in determining the absolute vocabulary size of test-takers (Nagy & Anderson, 1984). In addition, as Goulden et al. (1990) underscore, a principal consideration in test sampling from a dictionary should be that, size-wise, the sample of words selected for testing should be as closely as possible representative of the words in the dictionary. In other words, the larger the sample, the more representative it will be. Finally, when words are selected from a dictionary by means of a spaced sampling procedure, it is of paramount importance that high frequency words are not over-represented in the sample, which, again, can be achieved by maintaining consistency in the word count and item selection (ibid.).

With these considerations in mind, a word listed as a headword in *OSDOCE* was excluded from the count, without any exception, if it fell into one of the above-mentioned categories. By and large, these exclusion criteria seem to be adopted by most lexical researchers who use dictionaries for SW selection. However, there are certain problematic categories of dictionary items, such as foreign words, slang, old usage, compounds etc. that have received various treatments, very often without sufficient justification, by different researchers. In this connection, I felt it is important to briefly discuss those problematic categories and their treatment in previous lexical research in order to (1) show how their handling in this study differs from previous research and (2) justify my decisions and their consequences for the final estimates of vocabulary size.

Problematic issues. One of the problematic issues concerns the treatment of slang, foreign words and old usage. D'Anna et al. (1991) and Zechmeister et al. (1995), for example, in studying the vocabulary size of educated adults NS of English based their SW selection on a count from *Oxford American English* (Ehrlich, Flexner, Carruth, & Hawkins, 1980), which excluded hyphenated words, slang, foreign words, and old usage. In my view, the exclusion of these types of lexical items is not lexically justifiable. To begin with, in some estimates 65% of the English vocabulary is of non-Germanic origin, which makes the identification of *foreign words* very controversial. Moreover, excluding from a count foreign words, slang, and words identified as *old use*, in effect, means completely discarding the possibility that educated adults, both NS and non-NS of English, would know words such as *design, pizza, spaghetti, éclair, gender, hero, intellect, radio, piano, rhythm, lad, crap, terror, de luxe* etc., which all fall in one of those categories. Yet, undoubtedly, all of these words have become an inseparable part of English speakers' day-to-day experience and communication. Not to mention words such as *analysis, basis, synthesis, crisis, physics, cliché, format, genre, habitat, interfere, metabolism, nostalgia, ode, plight, radius, simpleton, sire, thou, vague,* etc., which can frequently be found in textbooks, works of literature, in the media, etc., i.e. in information sources that educated adults use on daily basis. Therefore, I consider the exclusion of entries, which represent any of those uses and registers not to be a meaningful criterion for exclusion of words from a dictionary count. In the present study I included foreign words, slang, as well as words identified as *old use* in the understanding that a test sample should not only have as many words as possible but should include as wider as possible a range of vocabulary uses that a dictionary allows.

Compound words present another serious problem for sampling from a dictionary. On the whole, this problem is stemming from the difficulty to define a compound. To begin with, it is not always possible to make a clear-cut distinction between compounding and derivation, since both processes result in the formation of new words. On the other hand, the distinction between compounding and collocating is not always straightforward, since the difference between single

and double stress does not always correlate with a semantic difference (Bauer, 1983). Bauer (ibid.) proposes a definition of a compound that seems to cover most of the cases where other definitions have problems explaining the identification of a compound. According to him, a compound is a form that contains two potential stems (not two words as often posited) which can explain instances such as *fishmonger* and *scandalmonger* that are clearly felt to be compounds, not derivations, by NS. However, these examples do not fit a definition suggested by Anglin (1993), following Bloomfield (1933), according to which a compound is "a lexical entry that consists of two or more words" (p. 18), since *monger* is used only in compounds and never as an independent word (Bauer, 1983).

Even a more nettled issue is the distinction between compounds written as two words and collocations. One of the reasons behind the ambiguity surrounding the distinction between both lies in the lack of definitive criteria for differentiating between the two structures. A major controversy about compounds relates to the claim that stress patterns in English should be considered as the defining criterion for a compound. That is, an English compound is more "word-like", hence, it has only one major stress on the leftmost element (e.g., '*point-blank,* '*high life,* '*midsummer madness,* '*hen-party*, etc. [Hornby, 1978]), while a collocation is a syntactic phrase and as such it is double stressed (e.g. '*powerful* '*engine*, '*strong* '*tea*, '*tarnished* '*reputation*, etc.) (Saeed,1997). These assumptions of stress patterns are largely based on the presumption that any given speaker consistently assigns a particular stress pattern to a given compound, regardless of whether the compound is pronounced in isolation or in a sentence. However, Bauer (1983) argues that this is a slippery assumption and his research shows that both in actual use or in experimental conditions NS are far from consistent in their stress pattern assignment. Moreover, there are double stressed compound structures in which the compounding seems to be semantically conditioned (e.g., *re*'*turn* '*ticket*, '*olive* '*oil*, '*civil* '*engineering*, etc.); hence, compounds do not represent a unitary group of stress patterns. Along these lines, the author suggests that it makes more sense to think of single and double stressed

noun + noun compounds, rather than oppose compounds to collocations on the basis of stress patterns (Bauer, 1983, p. 109).

Lexical researchers approached the problems associated with compound words and derivations with various solutions. Some researchers (e.g., Goulden et al., 1990; D'Anna et al.,1991; Zechmeister et al., 1995) excluded derivatives and compounds from their counts and from testing. Others (e.g., Nagy & Anderson, 1984) made a strong case about including those categories on the basis of relatedness among words. In view of the whole discussion, I decided to include derivatives and compounds, both solid written and multi-entries, in my dictionary count. I based the decisions about inclusion of multi-lexical headwords on lexico-semantic criteria, i.e. a multi-word entry was treated as a compound when it could function as a unit whose meaning can be derived from the whole expression rather than from its separate elements (e.g., *guinea pig* vs. *large pig*, *blackboard* vs. *black board*, etc.). Accordingly, the main entries that met these criteria were included in the count, e.g., *cross-examine, inasmuch as, line-up, nota bene, zip code*, *eau de Cologne*, etc. However, multi words listed under a head word as run-on entries (e.g., *blood bank, bloodhound, blood relation, bloodshed, blood-sports, blood pressure, blood poisoning* etc. listed under the main entry *blood*) were not included in the count because the distinction between compounds and collocations was completely blurred in the microstructure of entries.

Another issue that needed to be handled in a principled fashion in this study was the treatment of certain derivatives, since there doesn't seem to be agreement among researchers which affixes should be excluded from word counts (and testing) and why they should be excluded. Dictionary policies are often not very helpful in this regard, either. By and large, lexicographers traditionally take into account some theory of relatedness among lexical items when dealing with certain affixes. Therefore, they tend to treat regularly inflected variants as forms of the same words, i.e. as members of the word family (Nagy & Anderson, 1984). But, at the same time, many dictionary compilers list otherwise regularly inflected forms as separate

entries for purely orthographic reasons. For example, in *OSDOCE* along with an entry for *cry* as a noun, *dry* as a verb, and *dry* as an adjective, the plural form *cries,* the past and past participles *dried*, as well as the comparative and superlative *drier, driest* are listed as separate boldface entries, cross-referencing dictionary users to the main uninflected forms of the words. Apparently, these separate listings are motivated by plainly orthographic reasons, which is completely understandable in view of the purpose of a dictionary as a reference source. This observation only comes to remind again that along with the necessity of conducting an independent count of the dictionary used for test item sampling, researchers need to establish clear criteria for what would comprise a word family and whether words or word families would be the basic unit of vocabulary estimates.

Justification of established criteria. Defining a word family. In light of the preceding discussion which outlined pitfalls in using a dictionary in lexical research and in an attempt to overcome some of the problematic issues of previous research, I carefully examined the organization of *OSDOCE.* Special attention was paid to the consistency the lexicographer maintained throughout the dictionary in order to determine the extent to which the dictionary policy was in agreement with my linguistic and research criteria. I believe that such an analysis would support the decisions I made about what to include and what to exclude from my dictionary count. Finally, by comparing the lexicographer's criteria with my research criteria I tried to find additional support for the usefulness of dictionaries, in particular of *OSDOCE,* for vocabulary testing.

The examination of the organization of the dictionary information showed that the macrostructure of the dictionary contains lemmatized[1] headwords, letters, affixes, proper words, abbreviations, cross-references of contracted and irregular forms, interjections, and multi-word entries. I included only the lemmatized headwords in the count and the multi-word entries, which

[1] A *lemma* can be defined as a canonical form of a word that is chosen to represent a paradigm (Landau, 2001, p. 98).

were granted a headword status. Goulden et al. (1990) support this selection principle by arguing that dictionaries considerably differ in the way they deal, for example, with proper words and their place of inclusion and capitalization, letters and names of letters, abbreviations, etc. Proper words very often acquire either some connotative and cultural value or can take their place among the ordinary words of a language (e.g., Lorge & Chall, 1963; Goulden et al., 1990). Letters of the alphabet, names of letters, abbreviations, contracted forms, and alternative spellings have been found not to place a significant extra learning burden on language learners, so they can be safely excluded from testing. Affixes are not words and interjections function more as sentence emotional modifiers rather than as word types (Kolln & Funk, 2002); hence, both categories should not be included in a word count and testing.

A thorny issue that dictionary compilers have to deal with at the level of macrostructure is homonymy, in particular homographs[2], and polysemy. *OSDOCE* approaches these phenomena in a consistent way by granting homonyms separate headword status, whereas definitions of polysemous entries are listed under the headword entry. This means that homonymy is treated as part of the macrostructure, whereas polysemy is assumed to affect the microstructure (the organization of a dictionary article) of the dictionary organization. My analysis revealed that divergence of meaning, regardless of grammatical form, seems to be the overriding factor in determining homonymy. For example, the word *mess* is listed with two homographs in *OSDOCE*: *mess¹* provides information about its meaning as a noun (i.e. "*a state of confusion, dirt, or disorder"*) and as a verb (i.e. "*to put into disorder or confusion*"). However, the subsequent homograph *mess²* is defined again as a noun with a meaning that largely diverges from the meaning listed in the previous entry, i.e. "*company of persons taking meals*

[2] A homograph is one of two or more words that are spelled identically but differ in meaning and derivation. For example, *fair* as a noun and *fair* as an adjective are homographs because they have the same spelling but different meanings. In general, lexicographers are interested only in those homonyms that are spelled alike, not in homophones that are spelled differently (e.g., *there* vs. *their*), and commonly list them as separate entries, with a numerical superscript to distinguish them (Anglin, 1993; Landau, 2001).

together (esp. in the Armed Forces); these meals; the room, etc. in which these meals are taken". This approach is consistently maintained not only with homographs from the same lexical category, but also with homographs that belong to different parts of speech. For example *hard* as an adjective differs in most of its senses from *hard* as an adverb. So, even though there is some overlap of senses between both, this overlap has not evidently been found to be sufficiently strong so that to list, let's say, the adverb as a run-on entry[3] of the adjective. In a like manner, irregularly inflected forms are given separate entries in the dictionary, which is very much in line with experimental evidence that supports the notion of obscured morphemes, e.g. suppletive forms like *go* and *went*, having a separate representation in the mental lexicon (Sandra, 1994, p. 238).

At the microstructure level, the dictionary entries contain definitions, examples, run-on derivatives, and other forms pertinent to the headword. The analysis of the microstructure of an entry is important for, at least, two reasons: first, it reveals what derivatives the lexicographer has included in the word family and, second, it gives out information about the extent to which the publisher's policy is in agreement with current linguistic and psycholinguistic research in affixation. In general, the inclusion of derivatives in the microstructure is based on the presumption that if dictionary users know the meaning of the main word and the meaning of an affix, they will have no difficulty understanding the derived word (Landau, 2001, p. 101). The analysis of the dictionary entries in *OSDOCE* made it clear that the microstructure of an entry includes information about the seven types of regular inflectional suffixes in English: the plural inflection (e.g., the *–s* in *dogs*), the possessive inflection (e.g., the *–'s* in *father's*), the third person singular verb inflection (e.g., the *–s* in *plays*), the present participle verb inflection (e.g., the *-ing* in *playing*), the past tense/ past participle inflection (the *–ed* in *played*), the comparative inflection (the *–er* in *quicker*), and the superlative inflection (the *–est in quickest)*. These suffixes

[3] Run-on entries can be defined as the canonical forms of grammatically related words which are listed at the end of a main entry (Landau, 2001)

are assumed to produce mere paradigmatic variants of the base. Along these lines, psycholinguistic research in modeling the mental lexicon has convincingly argued that inflectional suffixes should be treated outside the lexical domain because they are linked to a particular class form (e.g., verbs, nouns, adjectives, etc.) and produce different realizations of a single word rather than different words (Sandra, 1994, p. 231). In fact, experimental research in this direction largely supports a two-stage lexicalization system of language speakers, where during the first stage syntactically and semantically specified items (i.e. lemmas) are accessed to which morphological and phonological codes are added during the second stage (e.g., Kempen & Huijbers, 1983; Levelt, 1989). Therefore, research investigating the structure and the organization of the mental lexicon ignores inflectional morphology, considering it unproductive, and focuses on the lemma as a unit that is revealing of the connections within and between lexical entries (Levelt, 1989).

Finally, with regard to the treatment of derivational affixes, the observed practice in *OSDOCE* seems to be based on the understanding that if a derivative has a meaning that is not covered by the senses or the definitions listed under the main entry to which it is appended, it should be entered and defined as a separate entry. This practice, though quite reasonable, prevents, to a great extent, from a systematic treatment of commonly used derivational affixes, such as *-ly*; *-ness; -ment*; or *un-*, proposed by some researchers to be excluded from dictionary counts and from testing (e.g., Meara, 1996; Nation, 2001). However, I decided to adopt the lexicographers' policy as explained above and not to exclude from the count any boldface main entries derived by means of any derivational affixation. This decision rests on the understanding that excluding from the count some commonly used affixes would result in a failure to acknowledge unpredictable differences in the meaning of words which have been derived by otherwise predictable affixation processes. For example, the entries *high*[1] (adj.), *high*[2] (adv.), *high*[3] (n.), as well as *highly* (adv.) and *highness* (n.) are listed as separate entries in *OSDOCE* and were, thus, counted as separate entries in the word count, though there are two nouns and

two adverbs among them. The example only comes to illustrate that even though the derivations with *–ly* and *–ness* are among the most predictable in meaning, the affixation alone cannot account for the fact that *high* and *highly* as adverbs have different meanings, nor can it explain why *high* as a noun differs in meaning from *highness*. By contrast, *marked* (adj.), *markedly* (adv.) and *marking* (n.) are treated in the dictionary as run-on derivatives of the main entry *mark* (v.) because of the semantic overlap between all the forms. Accordingly, they were not counted separately.

In sum, the notion of a "word" adopted in this study was largely based on semantic rather than on purely morphological criteria. It was taken to include: 1) a headword (including its other variants, e.g. spelling, abbreviations, etc.), 2) its regularly inflected forms, and 3) its derived forms (including zero derivation), but only the ones covered by the senses or the definitions cited under the main entry. Thus, the understanding of a "word" in this study comes close to the understanding of a "word family" proposed by Nation (2001), but with the stipulations outlined above. So, following the suggested specifications, knowing the word *motor* (meaning the word family of the headword), for example, will include knowing the following forms: *motor* (n.) and its inflected forms (*motors*, *motor's*), *motor* (v.) and its inflected forms (*motors, motored, motoring*), as well as the derived words *motorist*, *motorize*, *motoric* listed under the main entry. But it will not include, for example, *motorcade*, *motorway*, or *motor nerve* because their meaning differs from the senses covered in the definition.

All words in the *OSDOCE* were counted, according to the above criteria, by the researcher. The total number of words, which occurred as boldface new main entries in the dictionary was found to be 16,045. Then, 100 pages of the 762-page dictionary were randomly sampled and a second researcher used the established criteria to make an independent count of the words on these pages. Of the 2,121 words that occurred on the 100 randomly selected pages, there were only 12 disagreements with the original count (less than 0.6%), which means that there was a high degree of agreement (99.4%) between the two independent counts. This

also suggested that the margin or error would be very small, practically insignificant, and that the count of words can reliably be used by other researchers, who agree with the proposed dictionary count specifications.

The obvious discrepancy between the stated by the publisher size of 35,000 words and phrases and the researcher's count of 16,045 entries is a result of different counting criteria set up by the publisher and the researcher. Landau (2001) provides an extensive list of the items that publishers usually include in their count, which explains the serious disagreements between researchers' counts and publishers' claims. Unfortunately, this discrepancy has led to enormous divergence of estimated vocabulary sizes dependent on what count researchers based their estimates. For example, the preface of *Webster's Third New International Dictionary* (1961) says that the dictionary contains over 450,000 words; however, Dupuy (1974) found that if the same line entries were excluded, the dictionary would contain 267,000 words. Nonetheless, Diller (1978) took the publisher's word for the number of words in the dictionary and, as a result, estimated the vocabulary size of high-school seniors to be 216,000 words - a size that has never been supported by any other lexical study. In this regard, instead of blaming discrepancies of research results on dictionary policies, I believe that it is a researcher's responsibility to be fully cognizant of what lexicographers include in their count and how their inclusion criteria are similar or different from the criteria established by the researcher. In this connection, Landau (2001) offers an extensive list of the items that lexicographers tend to include in their entry counts, which can be used as a good starting point of any analysis. Lexicographers' entry counts include the following:

1) the main entry (e.g., *mother* as a noun);
2) any other defined part of speech of the main word (eg, *mother* as a verb);
3) inflected forms that are shown in the entry (e.g., *mothered*);
4) run-on derivatives, such as *motherless*, *motherly*, *motherhood*;
5) idioms (e.g., *do/ think fit (to do sth)* meaning *to decide to*);

6) hidden boldface entries, i.e. alternative forms of an entry;

7) other variants, spelling, abbreviations etc. (e.g., *colour, amp,* etc.);

8) lists of words with a common affix following a main entry (e.g. lists of words prefixed by *un-* following the entry introducing the prefix).

Apparently, the above mentioned criteria attain a specific purpose, i.e. besides offering lexical information, they serve the purpose of making the reference source, what a dictionary is, user-friendly. However, the overall goal of lexical research differs from that of dictionary compilers in that lexical research aims at finding as much as possible about the lexical knowledge of language learners who, above and beyond being learners, are also dictionary users. The examination of the above mentioned set of lexicographer's count criteria in the context of *OSDOCE* made me conclude that the dictionary compiler has made every effort to accommodate in the dictionary findings of current linguistic and psycholinguistic research in that its organization and structure closely reflect recent findings about the organization and structure of the mental lexicon. Furthermore, the discussed framework of criteria can be said to illustrate the lexicographer's concept of a word family, which does not significantly differ from my specification of a word family. In conclusion, I should mention that based on the overall analysis of the dictionary, its policy, organization and structure I believe that dictionaries can be an invaluable test sampling source for vocabulary research purposes, given that a researcher is well-informed of the extent to which the compilers' policy is in concert with his/ her framework of count criteria.

Sampling of words. The test items were randomly selected from the dictionary by means of a spaced sampling procedure. The sampling interval was every first new boldface main entry from every 20th column, starting from the right-hand column on the first page of the dictionary, regardless of whether the entry was a homograph or not. Most of the criticism towards research based on samples that counted homographs in their sample selection (e.g., Seashore & Eckerson, 1940; Smith, 1941 cited in Anglin, 1993; Templin, 1957 cited in Anglin, 1993)

addresses the common problem of largely inflated estimates of absolute vocabulary size stemming from methodological decisions. In these studies, the researchers counted subsequent homographs (e.g., *high¹, high², high³*) as separate entries but then they accepted the word as known if a participant could recognize and express the meaning of any of the homographs (Anglin, 1993). Also, failure to adjust for homographs often results in a sample that is biased towards high frequency words since those are the words that tend to have multiple entries in a dictionary (ibid.). In the present study, homographs were treated as separate entries, when listed as separate boldface entries in the dictionary, but their lexical category was specified next to SW in the test sample. For example, one of sampled SW was the word *hard*, which was specified in the test to be responded as an adverb since it occurs as a first entry on a 20th column and is listed in *OSDOCE* as an adverb that is a subsequent homograph of the adjective *hard*. The lexical specification was intended to limit participants' correct responses only to the ones in which they provided an explanation showing recognition of the SW as lexically specified. Thus, the resulting list of 75 words in the test sample became representative of the 16,045 items counted as headwords in the dictionary in terms of word frequency, lexical categories, and morphological word types. The following criteria were consistently applied for a dictionary entry to be selected for a SW:

1) An item should be a word that occurs as a boldface main entry in the dictionary.
2) An item should be a content word (noun, verb, adjective, adverb, preposition).
3) An item could be a compound or hyphenated word that occurs as a main entry in the dictionary (e.g., *point-blank, drawback*).
4) A selected item is listed with its syntactic category, when it occurs as a subsequent homograph in the dictionary.
5) An item is listed with the syntactic category that is first listed in the entry by the lexicographer.
6) Irregular forms are treated as separate items.

7) Both British English and American English usage are included when such entries are listed as separate boldface headwords.

In sum, the suggested inclusion criteria for the selection of SW were consistently applied throughout the sampling procedure. The approach taken to the dictionary count, as well as to the sampling of the test items, was to provide a meaningful operational definition of a word that would describe the lexicon of educated adult non-NS of English in terms that are close to the ones most educated people would apply when referring to the content of their vocabularies. The test sample covered a broad range of lexical usage, including specialized vocabulary, foreign words, slang, formal and informal usage, old and current usage, as well as poetic and technical words from a wide range of word frequencies. Finally, the systematic application of very specific criteria left little room for subjective decisions or ambiguity, which traditionally surround dictionary-based sample selections.

Description of the test items. The selected SW were listed alphabetically in the test in the order they were selected from the dictionary. After the sample was completed, two words were excluded from testing (*nonagenarian,* and *stern* [v.]) because their frequencies were not found in *The Teacher's Word book of 30,000 words* (Thorndike & Lorge, 1944) nor were they listed in *The Educator's Word Frequency Guide* (Zeno, Ivens, Millard, & Duvvuri, 1995), where the word frequency count is based on over 14 million words. Therefore, I decided to exclude these two items from the test because of their apparently very low frequency of occurrence. The final version of the test contained 73 items as SW (see Appendix A) of different syntactic categories (see Table 1). Appendix B shows the classification of the SW according to their syntactic category.

Table 1

Syntactic categories of the SW used in the study

Syntactic categories		**Number of items**
I. Content words with specified lexical category	nouns verbs adjectives adverbs	42 16 13 2

A problem, previously briefly mentioned, that stems from using spaced sampling for vocabulary testing concerns the representation of different frequencies in the test sample. In other words, it has been suggested in the literature that a direct consequence of the application of this procedure is the likelihood of high-frequency words to be over-represented in the sample since they tend to occupy more dictionary space than the lower frequency words (e.g., Lorge & Chall, 1963; Goulden et al., 1990, Nation, 1993). Avoiding such a bias becomes even more important in light of the fact that there is a strong connection between frequency of occurrence of a word in a language and the likelihood of people knowing that word (Nation, 1993). Therefore, to ensure that several frequency bands were well-represented in the test sample, I examined the frequencies of the selected SW in *The Educator's Word Frequency Guide* (Zeno et al., 1995). *The Educator's Word Frequency Guide* is based on over 60,000 text samples (over 14 million words) from a wide range of materials that students in the United States are likely to encounter throughout their school and college years. The word frequency estimates are based on materials taken from a variety of textbooks, works of literature, fiction and non-fiction books used in schools and colleges, which allows for getting unbiased to a specific type of text estimates of word frequencies. The frequency of each word included in the test was identified by its Standard Frequency Index (SFI), which is a logarithmic transformation of the *u*-value (i.e. the frequency of a word per million tokens in a corpus of an infinite size) (Zeno et al., 1995). The authors claim that in addition to retaining all advantages of the *u*-value, the SFI provides a more

compressed range of values than the *u*-value. So, when, for example, a word has a SFI = 20.8 (e.g., *rigidity*), this means that the word has a frequency per million approximately ten times lower than a word with a SFI = 30.7 (e.g., *tuner*), twenty times lower than a word with SFI = 40.4 (e.g., *fathom*), and thirty times lower than a word with SFI = 50.6 (e.g., *beaten*). The frequencies of the test items, as well as their distribution are given in Appendix C.

The analysis of the word frequencies of the selected SW showed a relatively equal representation of lexical items from several frequency bands. There were 13 SW (18% of the test sample) from a frequency band with SFI between 20.0 and 30.0, 23 items (31% of all items) from the frequency band with a SFI between 30.0 and 40.0, 20 items (27% of the test items) with SFI from 40.0 to 50.0, and 18 items (24%) with SFI from 50.0 to 70.0+ (see Table 2).

Table 2

Frequency bands represented in the test sample

SFI band	number of items	% of representation
20.0+	13	18 %
30.0+	23	32%
40.0+	19	26%
50.0+	18	24%

Based on this distribution I can say that, firstly, low and high frequency words were fairly evenly distributed in the sample and, secondly, the sample was not biased towards any frequency band.

Description of the test format. The test format required the participants in the experiment to demonstrate in a verifiable way that they knew what each of the words means. To this end, I used a modified version of the Vocabulary Knowledge Scale (VKS) which was developed by Canadian ESL researchers Paribakht and Wesche (1993) for use in L2 lexical research. On the whole, the instrument can be successfully utilized for any set of words that a researcher is interested in testing (Read, 2000). It combines self-report with elicitation of demonstrated

knowledge of specific words in written form, which allows for comparison between self-perceived knowledge and actual knowledge of at least one sense of any SW. In effect, the instrument used in the experiment consisted of two scales: one, showing a degree of familiarity with the lexical items, and the other, revealing of patterns of associative responses. The two scales accompanied every SW and were presented simultaneously to the participants (see Appendix D).

The first scale has four categories. For each SW, the participants were asked to choose the category (from I to IV) that best reflected how well they knew the SW. Option I means that the participant doesn't recognize the word at all, while at option II, the word is recognized but the meaning cannot be retrieved (or is not known). In general, the first two steps rely on participants' honest self-report of whether or not they recognize the SW, whereas the next two steps require from them to verify that they know the stimulus. The distinction between option III and IV involves an element of judgement on the part of the test-takers about the extent to which they are sure what the SW means. In either case the participants were asked to demonstrate their understanding of the SW by writing down a synonym, a brief explanation, or a translation equivalent (for the non-NS).

Option V represents the second scale that moves the participants from retrieval of meaning to production of associations. Provided they chose either option III or IV and responded as required, the participants were asked to supply as many as three words that they associate with the SW. This is where the test format used in this study differs from the VKS proposed by Paribakht and Wesche (1993; Wesche & Paribakht, 1996). In their scale, following the recognition of the word and an explanation of its meaning (options III and IV), their test-takers were required to produce a sentence that would illustrate their ability to use the SW with certain syntactic and semantic accuracy in a sentence (Wesche & Paribakht, 1996). However, the focus of this study is different from the focus of the study for which this scale was originally developed. The original study examined incidental vocabulary acquisition of ESL students

during instruction, while the purpose of this study is to examine lexical knowledge of L2 learners as a function of language proficiency. So, the scale was slightly modified to serve the purposes of the present research.

Scoring procedures

Seven dependent variables, frequently associated with the posited three dimensions of lexical competence, were used to compare the macrolevel of participants' vocabulary knowledge. Quantity, also called *breadth* of vocabulary, is often studied by examining test-takers' vocabulary size and knowledge of words from different frequency bands. Metacognitive awareness can be measured by the extent to which language-users' self-reported knowledge reflects their actual knowledge of words. Finally, quality of participants' vocabulary is most often studied by looking at several quantitative features of their WA domain, such as degree of commonality of shared nativelike-ness, number of associations, within-group consistency of the WA domain, etc. The scoring procedures used for quantifying each variable are described below.

Self-perceived knowledge. As previously noted, the VKS procedure combines self-report with demonstrated knowledge of a set of SW in written form. In addition, it implicitly yields information about the test-takers' metacognitive awareness in that it allows for measuring their ability to engage in self-interrogation concerning their degree of knowledge of the SW. As Brown (1980) argues, self-interrogation concerning the current state of one's own knowledge during any problem-solving task is an essential skill in a wide variety of situations (p. 454). In other words, the capacity of monitoring task demands and finding ways to meet these demands constitutes the individual's metacognitive awareness which, in fact, lies at the heart of the distinction between knowledge and the understanding of that knowledge (ibid.).

It was believed that the difference between participants' self-perceived knowledge and their actual knowledge of the SW would reflect their degree of metacognitive awareness, i. e.

how much they actually know when they claim they know a word. The VKS has four options and, in effect, translates participants' responses to each SW into test scores. Therefore, each option in the VKS was assigned a numerical value (i.e., option I = 1 point, option II = 2 points, option III = 3 points, option IV = 4 points) and dependent on how a participant self-reported his/ her familiarity with the SW, a corresponding score was assigned to each stimulus.

Verified vocabulary knowledge. To capture how participants' self-perception of self-knowledge corresponded to their actual knowledge I used a scoring scale similar to the one proposed by Wesche and Paribakht (1996), with certain modifications. Options I and II yielded a score of 1 and 2, respectively. Option III could lead to a maximum score of 3, if a participant had provided an acceptable synonym, brief definition, or translation (for the non-NS) of the SW, or 2 if a participant claimed some familiarity with the stimulus but the response had shown that he/ she did not recognize it correctly. Option IV yielded maximum 4 points, if knowledge of the SW was demonstrated as required, or 2 points if the response revealed some sort of a misinterpretation of the SW or did not reflect its lexical category as specified. Table 3 presents the scoring system for all options.

Being part of the paradigmatic relationships between words, for a synonym or a translation of the SW to be considered a correct response, it had to be semantically and syntactically relevant to the SW. For example, for the SW *hard* specified as an adverb in the test, synonyms that could yield 3 or 4 points were *strenuously, laboriously,* etc. and synonyms that would lead to only 2 points would be *difficult, strong, tough, solid*, etc. because of the failure to respond correctly to the lexical category of the SW. The requirement that a response should adequately reflect the semantic and the syntactic properties of the SW was applied to the translations, as well. When a brief definition/ or explanation was provided as a response in category III and IV, the decision about its acceptability was based on the criteria, outlined by Landau (2001).

Table 3

Meaning of the VKS scores

Self-report options	Possible scores	Meaning of sores
I.	**1**	The participant doesn't recognize the SW at all.
II.	**2**	The participant recognizes the SW but cannot retrieve its meaning.
III.	**2** or **3**	The participant provides a synonym, brief definition or translation that is incorrect or does not correspond to the specified lexical category of the SW, which renders the word familiar but not known. The participant shows a degree of uncertainty about the meaning of the SW but provides a correct synonym, brief definition, or translation.
IV.	**2** or **4**	The participant provides a synonym, brief definition or translation that is not correct or does not correspond to the specified lexical category of the SW, which renders the word familiar but not known. The participant is sure about, at least, one of the senses of the SW and provides a correct synonym, brief definition, or translation.

These criteria are used in lexicography for defining words by part of speech. For instance, the explanation definition of a noun should immediately answer the question, "What is it?" and in order to answer that question it has to contain a noun, either quantified or not, in the first part of the definition that, actually, shows the part of speech. The definition of adjectives was expected to contain some of the introductory words and phrases used for defining adjectives, e.g. *able to, being, belonging to, full of, having the quality of, pertaining to*, etc. The definition of verbs was anticipated to begin with another verb, with or without infinitival particle "to". Adverbs were considered correctly defined if another adverb was included in the definition or if the definition contained the phrase with the meaning "in the manner of".

Vocabulary size. Estimates of vocabulary size were derived by multiplying the number of known words (obtained from the correct responses to option III or IV in the test) by the number of words counted in the dictionary, divided by the number of SW used in the test, i.e.:

$$\frac{\text{Number of known words x Number of dictionary items (word families)}}{\text{Number of SW in the test}}$$

The estimates were derived on the basis of the notion of "word" as specified earlier by the researcher.

Nativelike typicality of associative responses. By and large, there are three principal procedures for elicitation of WA responses within the free association method: a) *discrete free associations*, where participants are asked to respond with the first word that comes to their mind; b) *continued associations*, i.e. a SW is presented a certain number of times and participants are asked to give as many responses as the number of times the stimulus is presented; and finally c) *continuous associations*, where the stimulus is presented only once and serves as a point of departure for the production of a chain of associations. Each of these procedures has been found to significantly affect the characteristics of the association domain. Therefore, following the proposals of several L2 researchers (e.g., Schmitt & Meara, 1997; Meara, 1996; Schmitt, 1998; Read, 2000; Wolter, 2002), I decided to ask my participants to give as many as three responses, rather than just one or as many as they can think of, in order to obtain a reasonable number of associative performance that would form a sufficiently large association domain for every SW.

The traditional procedure used for quantifying non-NS' associative responses is to match them to a norm list of associations elicited from a group of NS. Schmitt (1998) accordingly notes that when comparing non-NS' to NS' associations, the attribute on which L2 responses should be judged is the degree of "native-likeness" of associations rather than the opposition "*correct*

vs. *incorrect* or *is an association* vs. *is not an association*" (p. 390). The quantification procedure I used was an attempt to give the non-NS credit for producing responses that were given by NS, both typical and idiosyncratic. Interestingly, the distribution of common responses (an association generated by at least 2 participants, when responding to a given SW) as well as idiosyncratic associations in the NS' data was very surprising. Out of a total of 950 cases of three responses generated to the SW by the NS, it was found that only 17% of the responses (159 cases) were three common associations, 33% (315 cases) were two common responses, 32% (301 cases) included only one common response, and 18% (168 cases) contained three idiosyncratic associations. In the context of total number of associations (3,304) in the norming list, 52% (1,709) of the responses were common and 48% (1,595) were idiosyncratic. These figures suggested that, on the one hand, there was a small 2% difference between the number of NS' common and idiosyncratic responses, and, on the other hand, giving three common responses or three idiosyncratic responses was not a typical behavior for this norming group. Rather, including one or two common responses was the most common pattern. Interestingly, this distribution of responses was quite different from the one reported by Schmitt (1998) in his study of three associations given by 100 university students to 17 SW. In his analysis, the most common response pattern was three (54%) and two (34%) common responses, which made him argue that this high level of commonality among NS' responses (88%), is an indicator of *native-likeness of associative behavior* against which non-NS' associations should be matched. My figures did not support such a conclusion. They indicated a degree of *native-like typicality* (65%) which made it reasonable to take NS' associative behavior as a criterion in quantifying the non-NS' responses. At the same time, it was easily noticeable in the non-NS' association data that they also showed a high degree of typicality of different nature, which prompted me to explore the notion of typicality of associations in two ways: first, by comparing the three groups on typicality of their associations based on the NS' norming list and, second, by comparing the consistency of NS' WA domain with the non-NS' degree of commonality of associative behavior.

The first step in quantifying the associative responses was to compile a norming list from the associations provided by the NS. The elicitation process from the control group resulted in a total of 2,084 different associations, with few illegible responses, which were not considered in the count. The norming responses were lemmatized and tallied on a list. The following items were listed as separate associations:

1) base words and some of the inflections were lemmatized and combined as one item, i.e. inflection *–s* for nouns (e.g., *car* and *cars*); inflection *-s* for verbs (e.g., *forget* and *forgets*); inflections *–er* and *–est* for adjectives (e.g., *tall – taller - tallest*) and adverbs (e.g., *fast – faster - fastest*);
2) multiword responses (e.g., *auspicious situations, good opportunities, marching line, give up, spread out,* etc.) were listed and scored as one item;
3) all derivations were treated as separate items (e.g., *cleaning, cleaned, breathless, immoral, resignation, disagreement, scrutinize, unfamiliar*, etc.);
4) irregularly inflected forms (e.g., *children, better, went*, etc.) were treated as separate forms.

For an association to be considered in the analysis, the participants should have demonstrated an acceptable degree of familiarity with the SW by responding to option III or IV in the test. For example, responses to the SW *gambol* that were tallied included associations such as *frolic, fun, laugh, play, skip, lively* etc. because the participants responded acceptably to option III (*I **think** this word means* _____) or IV (*I **know** that this word means* _____). Clang associations, i.e. responses given as a result of misinterpreting the SW to mean *gamble*, such as *money, casino, dogs, horses, gambling*, etc. were not included in the norming list.

As a result, a list of all associative responses generated to each SW was obtained with a tally how often each response was given. For example, for the SW *advantageous*, the top three responses were *helpful* (7), *beneficial* (4), and *good* (4). So the best performance, for a NS or non-NS, would be to produce the three most frequently given associations, which would

translate into a score of 15 points (7 + 4 + 4 = 15). As mentioned earlier, few of the participants gave the three most common associations; nonetheless, each response was assigned a numerical value of typicality dependent on its frequency of occurrence in the data. Responses that occurred only once in the NS' data were assigned a score of 1.

Within-group consistency of associative responses. To measure the typicality of non-NS' responses and compare its degree of commonality to NS' associative domain, the same procedure was applied to compiling a list of the associations supplied by the non-NS as a group. The list was used to quantify non-NS' responses once again, this time with regard to their within-group typicality. The procedure allowed for a comparison between the degree of consistency of each group's WA domain, in addition to comparing shared nativelike-ness of associative meaning between NS and L2 learners.

Word frequency. The SW in the test represented six word frequency bands, which were defined by reference to *The Educator's Word Frequency Guide* (Zeno et al., 1995). The assumption behind considering frequency of occurrence of words as a factor that influences knowledge of words was that vocabulary knowledge is cumulative across several frequency bands. To examine the relationship between word frequency and vocabulary knowledge, a numerical value from 1 to 6 was assigned to each SW the participants knew, dependent on its frequency band, i.e. the lower the frequency of occurrence of a word, the higher the numerical value. As a result, responses supported by demonstrated knowledge of a SW with a SFI 70+ (e.g., *back*) were assigned 1 point, responses to a SW with SFI = 60+ (e.g., *hard*) were awarded 2 points, responses to a SW with SFI = 50+ (e.g. *diamond*) were marked with 3 points, etc.. A major consideration in the quantification of the responses with respect to word frequency was to give more credit to participants who knew words from the lower frequencies in addition to the high frequency words.

Number of associations. A numerical value reflecting the absolute number of associations generated to any SW by a participant was assigned to every SW.

Reliability

Since the SW were not selected from commonly used word lists, e.g., *Nation's Vocabulary Levels Tests* (Nation, 1990), the *University Word List* (Xue & Nation, 1984), *General Service List* (West, 1953) etc., which are organized around word frequency counts (e.g., Thorndike and Lorge, 1944), it was of paramount importance to assess the internal consistency reliability of the test as a measuring instrument. In other words, it was important to find out the extent to which the test showed consistency in the performance of all participants from one set of items to another across all administrations of the instrument (Frankel & Wallen, 2000). Reliability for internal consistency of the test was calculated for each group by using Kuder-Richardson 21 (*K-R* 21) formula. The results were as follows: for the group of NS a value of *K-R* 21 = .88 was obtained, for the group of advanced learners of English *K-R* 21 = .80, and for the L2 intermediate group *K-R* 21 = .85. These results were comparable with the internal consistency reliabilities of other widely used instruments for assessment of lexical knowledge as, for example, Forms A and B from the University Word Level Test (Xue & Nation, 1984) which have reliability coefficients of .85 and .84, respectively (Beglar & Hunt, 1999, p. 139). In this regard, some researchers (e.g., Vierra & Pollock, 1992 cited in Beglar & Hunt, 1999) suggest that reliability coefficients between .80 and .90 represent an acceptable reliability range, while others (e.g., Popham, 1990) argue that there are no hard and fast rules how large a respectable reliability coefficient should be in view of the presumption that the reliability of a test is very much situation-dependent. For the purposes of this study, I followed the assumption that the internal reliability coefficients simply indicate the extent to which the SW used in the test were internally consistent (Popham, 1990) and I found the results satisfactory.

Results

A series of One-Way ANOVA was conducted to evaluate the relationship between language proficiency and each of the variables associated with lexical knowledge of educated

adults. The independent variable, language proficiency, had three levels: NS of English, who served as a control group, L2 advanced learners, and L2 intermediate learners. The dependent variable was lexical knowledge, measured by 1) verified vocabulary knowledge, 2) self-perceived knowledge of vocabulary, 3) vocabulary size, 4) knowledge of words from different frequency bands, 5) native-like typicality of associations, 6) within-group consistency of the WA domain, and 7) number of associations. The responses were transformed into numerical values as described in the preceding section. Several null hypotheses were tested, H_o: $\mu_{NS} = \mu_{advanced} = \mu_{intermediate}$, which stated that there were no differences among the three groups of English language users in their mean scores on the dependent variables.

I took the strategy of doing seven independent ANOVA rather than a single multivariate ANOVA because I had substantial interest in each of the dependent measures. Means and standard variations are presented in Appendix E. The analyses yielded a significant group effect on all dependent variables indicating statistically significant mean differences among the NS, L2 advanced, and L2 intermediate learners on verified vocabulary knowledge, $F(2, 61) = 23.82$, $p = .000$, $\omega^2 = .43$, self-perceived lexical knowledge, $F(2, 61) = 29.68$, $p = .000$, $\omega^2 = .48$, vocabulary size, $F(2, 61) = 13.16$, $p = .000$, $\omega^2 = .29$, number of associations, $F(2, 61) = 14.86$, $p = .000$, $\omega^2 = .32$, knowledge of words from different frequency bands, $F(2, 61) = 10.91$, $p = .000$, $\omega^2 = .25$, native-like typicality of associations, $F(2, 61) = 61.75$, $p = .000$, $\omega^2 = .66$, and within-group domain consistency of the WA domain, $F(2, 61) = 11.83$, $p = .000$, $\omega^2 = .27$. The calculated practical significance of the measured relationships (ω^2), i.e. the proportion of the total variability in each of the dependent variables that could be explained by the language proficiency of the participants, revealed that the dependent variables provided practically meaningful measures of the overall state of participants' lexicons when proficiency was used as a grouping variable.

The results of the analyses supported a conclusion that the NS', L2 advanced learners', and intermediate learners' lexicons were different on all dimensions of their lexical competence.

However, it was of much greater interest to explore the nature of these differences and the way they related to the language proficiency of the L2 learners. To this end, post hoc pairwise comparisons were conducted to compare the three groups on all measures used in the study. The post hoc comparisons, in which the significance level was adjusted to .05 based on Bonferroni rationale, revealed that there were statistically significant differences in all subsets of comparisons between the intermediate learners' overall state of vocabulary and the NS or advanced learners. However, no statistically significant differences were found between the advanced learners' lexical knowledge and the NS' state of vocabulary on all dependent variables but one – nativelike typicality of associations. The two groups were found to significantly differ only in the degree of their shared native-like typicality of associations. In other words, even at an advanced level of language proficiency, the non-NS have not reached the extent of typicality of associative characteristic of NS. On the other hand, the fact that the advanced learners were not statistically different from the NS on any of the other variables, such as actual knowledge of words, self-perceived lexical knowledge, vocabulary size, knowledge of words from various frequency bands, and number of generated associations, was a thought-provoking result because it directly challenged the usefulness of WA tests, as a measure of the degree of shared "nativelike-ness" of associations, on which conclusions about the lexical knowledge of non-NS could be based. Moreover, when the within-group consistency of the advanced group's WA domain was compared to the NS' typicality of responses, it was found that there were no statistically significant differences between the two groups. This came to suggest that just like NS tend to show a high degree of typicality of associative connections, L2 learners at an advanced level of proficiency also show stable patterns of commonality that does not necessarily resemble the one of NS. I shall further explore this idea in the discussion section of the chapter.

In sum, the examination of the macrolevel of lexical knowledge, across its three dimensions, led to the logical conclusion that the overall state of L2 learners' vocabulary

crucially depended on L2 learners' level of language proficiency. The analyses revealed that at the intermediate level of proficiency L2 learners' quantity, quality, and degree of metacognitive awareness of lexical knowledge showed features of instability, which would inevitably influence L2 learners' ability to productively use their knowledge of words. By contrast, advanced learners' overall state of vocabularies showed insignificant differences from that of L1 speakers across all variables but native-like typicality of associations. These results were interpreted to challenge the usefulness of WA tests as instruments on which judgements about the quality of L2 learners' word knowledge should be made. Instead, the finding was found to be more in agreement with Meara's (1996) proposal of looking at association data as revealing of patterns of connections that NS and non-NS tend to build among words in their mental lexicon, rather than examining associative behavior with respect to degree of shared nativelike-ness of meaning between NS and non-NS. Besides, it has become axiomatic in the L2 research field that "most non-NS never reach a point of being indistinguishable from NS of a particular community" (Gass, 1990, p. 37), though, it is questionable whether this is the ultimate goal of L2 learning. Nonetheless, the conclusion itself presupposes detectable differences between the two broadly defined groups, especially when the measuring stick is nativelike-ness of commonality of associations. Therefore, it is more meaningful to invest research resources in studying whether, for example, non-NS' strength and diversity of relationships among words in their mental store compare to those of NS, instead of exhausting a research area that does not hold a potential for novel findings.

General Discussion

The research questions examined in the study concerned three dimensions of L2 vocabulary knowledge posited to account for its macrostructure. The dimensions were measured by seven variables frequently used by L2 rechers to explore the lexical competence

of L2 learners. It was argued that quantity or "breadth" of vocabulary, quality or degree of connectivity among lexical items in language users' mental lexicon, and learners' metacognitive awareness were the dimensions that fully capture the overall state of L2 learners' lexicons. By and large, breadth of vocabulary is most often estimated by obtaining a measure of learners' knowledge of lexical items from different frequency bands or by their size of vocabularies. The degree of connectivity of learners' lexicon is usually established by examining the features of their WA domains, for example, its size (determined by the number of responses generated to a SW in a continuous association test) and its consistency (determined by the degree of typicality of associative responses). Finally, it is logical to assume that when we are dealing with adult learners' lexical competence, we are also dealing with their metacognitive awareness, i.e. their consciously self-perceived degree of familiarity with lexical items, which may or may not be dependent on their level of language proficiency. Therefore, seven variables associated with the three dimensions were used in the experiment in an attempt to investigate their relationship to the language proficiency of the participants in the study.

Three groups of participants were involved in the experiment. The group of NS served as a control group against which the performance of the L2 advanced and intermediate learners was compared. Even though the two non-NS' groups acquired English through formal instruction in Bulgaria, the distinction between L2 and FL was not considered to be of importance to the conclusions of the study because I share the view that FL research can "serve as the testing ground for some of the hypotheses in SLA" (VanPatten, 1990, p.18). Moreover, the relationship between FL and L2 research is no longer unidirectional, i.e. with a flow of theory, hypotheses, and research information only from the SLA to the FL context. Rather, it is one that is going both directions and is actively contributing to answering questions pertinent to both contexts of language acquisition. Not to mention that with the enhanced exchange of information and the ample opportunities to travel and study in English speaking countries, as well as to communicate in English outside the classroom, the boundaries between ESL and EFL

learning have become blurred to an extent that makes it difficult to draw a clear dividing line between L2 and FL. Gass (1990) further points out that “what on the surface appears to differentiate the two learning situations may in fact only be a difference of degree rather than absolute difference in which productive competence is possible in one situation, but not in another” (p. 38). Therefore, by comparing the three groups of participants it was assumed that the L2 learning situation would not obscure the results of the experiment. In the following paragraphs, I will briefly discuss the most important conclusions derived from the analyses. The discussion will focus on the way the results supported a three-dimensional model of studying lexical competence as a model that holds the potential to capture the variation in the vocabulary knowledge among language users when proficiency is taken into consideration. To this end, the major findings of the experiment will be discussed within the framework of the proposed model.

Breadth of vocabulary knowledge. One of the dimensions of lexical competence, and probably the best researched one, is *breadth* or quantity of word knowledge. The quantity of participants’ vocabulary was examined by measuring their size of vocabulary and their knowledge of words from a wide range of frequencies. To estimate the vocabulary size of the participants I used a method of test sampling from a dictionary, which is frequently used in L1 and L2 lexical research. The procedure is referred to as a *spaced sampling procedure* and involves selecting SW from a dictionary at a specific interval. Thus, the sample becomes representative of the content of the dictionary. This procedure is always accompanied by several caveats which relate to what lexical items should be included or excluded from a dictionary count and what items should be selected as SW for testing purposes (e.g., Nation, 1990, 1993; Goulden et al., 1990). On the whole, the way researchers deal with these issues largely depends on the goals of their vocabulary sampling and the overall goal of their studies. However, the research community has been recently showing growing awareness of the importance of handling some problematic issues in a consistent manner primarily because all

estimates of vocabulary size crucially depend on the treatment they receive in experimental work.

As far as this study is concerned, special attention was paid to analyzing the structure and the organization of the dictionary from which the SW were sampled, which the compiler claimed to contain 35,000 words and phrases. With respect to the size of the dictionary from which items should be selected, Wesche et al. (1996), for example, advice that a reasonable dictionary size for L2 research is one that contains at least 30,000 words. A problem frequently discussed in the research literature relates to the dilemma whether a researcher should take the lexicographer's word about the size of a dictionary or whether he/ she should conduct an independent count. Unfortunately, relying on lexicographers' estimates has usually led to obtaining highly inflated vocabulary sizes simply because dictionary compilers approximate their estimates based on sets of criteria that are consumer-driven rather than research-driven. That is why conducting an independent dictionary count has become a must if a researcher's goal is obtaining an estimate of his/ her subjects' vocabulary size. Therefore, after setting up explicit inclusion and exclusion criteria for what counts as a word, I conducted an independent count of *OSDOCE* and arrived at an estimate that was quite different from the lexicographer's claim. Based on my specific research criteria, it was found that the dictionary contains 16,045 words and this number served as a baseline lexicon for estimation of participants' vocabulary size. In the forthcoming figures I shall present the findings in numbers in words, with the meaning of "word" as specified earlier, not base words.

Even though providing an estimate of NS' vocabulary size was beyond the focus of this study, a brief mention needs to be made about how the results obtained in the experiment compare to L1 studies that specifically focused on finding out the number of words NS have in their lexicons. Such a comparison is necessary because it will show whether the estimated size of the NS' lexicon in the present study provides a reasonable baseline against which non-NS' quantity of word knowledge can be matched. By and large, estimates of how many words NS

have in their lexical repertoire abandon, ranging from unbelievably high (e.g., Diller [1978] found the vocabulary size of high-school seniors to contain 216,000 words) to more conservative current estimates. Recent studies (e.g., Goulden et al., 1990; D'Anna et al., 1991; Zechmeister et al., 1995) indicate that educated adult NS of English have vocabulary knowledge of approximately 14,000 to about 20,000 base words or in the range of 17,000 word families (Meara, 1996). Other studies (e.g., Nusbaum, Pisoni, & Davis, 1984 cited in D'Anna et al., 1991), using a smaller baseline dictionary count of 19,750 words suggest that the mental lexicon of undergraduate college students contains on average less than 15,000 words. My results indicated that based on a 16,045-baseline lexicon, educated adults NS of English, what undergraduate college students are, know on average approximately 9,500 words (range 7,392 – 11,501 words). As a proportion, this number showed that the native speaking participants in this study knew on average 59% of the estimated total size of the dictionary from which the SW were selected. I should mention here that I used conservative criteria for what counted as "knowing a word", i.e. 1) participants had to provide verified responses; 2) a response was considered acceptable only if it accounted for the lexical category of the SW in addition to, at least, one of its senses; 3) partial responses of the SW were not counted in the number of words rated as familiar by the participants. Given these stringent criteria, the proportion of known by the NS words to the estimated size of the dictionary compared very well with other L1 studies that also used abridged dictionaries (e.g., D'Anna et al., 1991; Zechmeister et al., 1995). For example, based on a dictionary count of 26,091 words and a 5-point scale of rated SW familiarity, D'Anna et al. (1991) found that college students know the meaning of 16,785 words, which was 62% of their dictionary count. Several years later, using the same dictionary, Zechmeister et al. (1995) felt that even though the earlier obtained size was relatively small compared to other estimates in the literature, it was very likely that it overestimated participants' vocabulary size because the test-takers simply rated their knowledge on a scale from 1 to 5 without being required to verify it. In an attempt to overcome some of the shortcomings of the

previous study, the participants also took a multiple-choice test of the words they had rated as known to verifying their rated degree of familiarity. Under those conditions, the researchers estimated that undergraduate college students on average knew 15,872 words, which was 59% of the of the overall dictionary count. Most probably, the small 3%-difference between the size obtained by rated responses only in D'Anna et al. (1991) experiment and the size derived by verified responses in Zechmeister et al. (1995) study can be explained by the generally heightened metacognitive control that college students have over their knowledge. On the whole, it can be said that, proportion-wise, the estimated vocabulary size of the NS in this study is in line with estimates derived by other researchers (e.g., Zechmeister et al.,1995), who specifically focused on measuring the quantity of vocabulary knowledge of educated adult NS of English. In addition, when the means derived from this experiment were compared with the average sizes across other studies, the results revealed that they were compatible, though non-significantly smaller. Thus, the comparison supported a conclusion of a meaningful estimate of NS' vocabulary size against which non-NS' vocabulary could be dependably matched.

Studies conducted in L2 lexical research have generally supported the notion that the L2 mental lexicon is "fundamentally different" (Wolter, 2001) from that of NS. By that most researchers until recently meant that L2 learners have significantly lower size of vocabulary, which causes most of their problems with dealing successfully with L2 language tasks. The results of the present study only partially agree with that claim because the experiment has the advantage of involving L2 learners at two proficiency levels, which, in turn, allowed for finer generalizations to be made concerning the quality of their vocabulary knowledge. It was found that the average vocabulary size of the advanced learners did not differ in any significant way from the average vocabulary size of NS. The average lexicon size for the advanced learners was estimated at 8,737 words (range 5,990 – 11,484 words), which indicated that advanced learners could recognize and define/ or translate approximately as many words as NS could. However, the intermediate students' vocabulary size differed greatly from that of the NS and the

advanced learners alike. The average lexicon of an intermediate student was found to contain 6,033 words (range 4,127 – 7,939 words), which suggested a strong relationship between level of language proficiency and quantitative differences in L2 learners' lexicon size. The same differences among the three groups were found when their knowledge of words from different frequency bands was compared, which only confirmed that variation in the quantity of vocabulary knowledge is reflected by the overall proficiency level of L2 learners. Along these lines, it has been widely accepted that for learners at higher levels of language proficiency, syntax is no longer a problematic area (e.g., Laufer, 1997). Instead, it is vocabulary that they themselves tend to identify as a major source of their language problems (Meara, 1978) or as a primary obstacle to successful comprehension (Laufer, 1997). In this regard, Meara (1996) points out that vocabulary size is probably the most powerful measure when dealing with what he calls "small lexicons", i.e. lexicons that for the English language consist of 5,000 – 6,000 words. He further suggests that beyond this "critical threshold", vocabulary size becomes less important and proposes that organization of the mental lexicon, or a measure of how well words are known, takes its place (p. 45). The estimates in this study suggest that on average, the intermediate students' vocabulary size falls more or less within the suggested by Meara "small lexicon" size. On the other hand, in the upper bound of its range, it closely approximates the average size of the advanced learners' lexicon, which means that the intermediate level can be thought of as a threshold level of language proficiency where growth in learners' lexical size, most probably, essentially contributes to their successful dealing with language tasks. On a final note, I should say that the conclusions regarding the size of the participants' lexicon are strictly limited to recognition of words in isolation, which renders the lexical task receptive in nature. However, as far as productive vocabulary or performance on communicative tasks are concerned, the estimated quantity of lexical repertoire may not be the same. The best prediction I can make, based on Laufer (1997) and the derived vocabulary size approximations, is that the intermediate learners should be expected to have reached the lexical threshold that, for

example, would allow them to successfully transfer L1 reading strategies onto L2 reading tasks. Laufer (1997) further suggests that knowledge of 6,000 word families is likely to result in 77% text comprehension, which is to a great extent supported by the reading scores of the L2 participants in this study. However, in terms of text coverage, especially regarding work with texts for academic purposes, the proposed figure of 3,000 word families being sufficient for comprehension of 95% text coverage (e.g., Nation, 1990; Laufer, 1992) may need serious revision.

Quality of lexical knowledge. For a long time, most L2 lexical research has primarily focused on estimating the vocabulary size of L2 learners, or *breadth* of their lexical knowledge, rather than on examining the quality of their vocabulary, or *depth* of their vocabulary knowledge. However, as interest in vocabulary growth developed, researchers began to show increasing awareness that a measure of vocabulary size alone can no longer provide a satisfactory description of L2 vocabulary knowledge; that they should also try to measure the quality of word knowledge (e.g., Schmitt, 1999; Wolter, 2001; Greidanus & Nienhuis, 2001; Meara, 1978; Meara, 1996; Nation, 1993; Read, 1993; Wesche & Paribakht, 1996). On the whole, there is recognition in the field that knowledge of words is multidimensional, i.e. the words that we have at our disposal are also connected in a variety of networks, e.g., thematically, phonologically, conceptually, socio-linguistically, culturally, etc. Moreover, the general assumption is that the denser the networks of a given word, the greater the knowledge of that word; hence, the greater the *depth* of vocabulary knowledge.

Various attempts have been made to explore the idea of measuring the *depth* of lexical knowledge. Many lexical researchers (e.g., Schmitt, 1999; Wolter, 2001, 2002; Greidanus & Nienhuis, 2001; Meara, 1978, 1996; Nation, 1993; Read, 1993; Wesche & Paribakht, 1996; Kruse et al., 1987) seem to agree that using WA tests to probe L2 learners' *depth* of vocabulary knowledge can be a valuable instrument for collecting data that are revealing of the way language users organize their mental lexicon. However, most of the attempts have been

devoted to studying the extent to which non-NS' associative behavior resembles the degree of commonality of associations that NS maintain. This line of research has been largely promoted by L1 lexical studies which have repeatedly found that NS tend to cluster their responses to any SW around a small number of associations. Consequently, this high degree of associative commonality has become to be interpreted as indicative of the stability of semantic connections that NS build among the words in their mental store, which, in turn, reflects the way they organize the meaning networks. Similarly, some L2 researchers (e.g., Schmitt, 1999; Wolter, 2001, 2002; Kruse et al., 1987, etc.) adopted the approach of comparing the degree of non-NS' associative commonality to that of NS' associative behavior in an attempt to account for the effects of language proficiency on the semantic stability of the L2 organization of the mental lexicon. This line of research was further encouraged by findings suggesting that L2 learners tend to have much more diverse and unstable patterns of responses, though, as their proficiency increases, some aspects of their associative connections develop towards NS' norms (e.g., Meara, 1978; Vermeer et al., 2001). By the same token, as pointed out by Kruse et al. (1987), WA tests started to be used by "L2 researchers who are interested in the developing lexicon as a measure for establishing baselines for various experimental investigations" (p. 141). Consequently, several relatively distinct lines of L2 WA research developed, which differ primarily in the perspective from which L2 association patterns are viewed, for example, from a socio-cultural perspective, from a language proficiency point of view, as an indicator of *depth*, or quality, of vocabulary knowledge, etc. Unfortunately, the relationship between these lines of research with regard to the implications they have for each other has never been commented in the L2 literature. The present study offers such an opportunity, which I will explore in the context of the discussed experiment as far as the data allow me to do.

In my study I used three variables that are commonly associated with the analysis of the quantitative features of language users' WA domain, i.e. native-like typicality of responses, number of associations, and within-group consistency of the WA domain. The rationale behind

using three variables, where L2 researchers usually use just one, was to find out the variables that would account for, both, the effects of language proficiency on the quality of participants' lexicon organization, as well as the stability of connections that they build among the words they know. Figure 1 summarizes the results of the analyses.

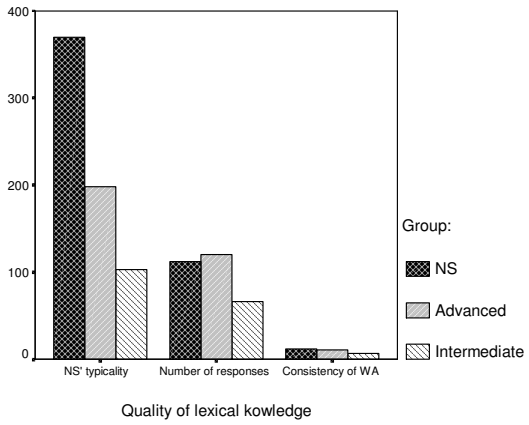


Figure 1. Quality of lexical knowledge measured by native-like typicality of associations, number of responses, and within-group consistency of the WA domain.

As the graph shows, the comparison among the three groups of participants on their shared native-like commonality of associative responses revealed that neither the advanced, nor the intermediate learners come even close to the relatively high degree of associative commonality that the NS maintained. Similar findings are traditionally interpreted in the L2 WA literature to mean that L2 learners produce diverse and unstable associations, which implies a lesser degree of connectivity among the words they know that, in turn, indicates a lower quality of their meaning connections and lexicon organization. Had I used just this one variable to analyze the quality of non-NS' lexical knowledge, I would have arrived at the same conclusion. However, a closer look at the association data, as well as at the socio-cultural WA research, prompted me to challenge the usefulness of this approach to analysis of WA data on which conclusions about the quality of L2 learners' organizations of their lexical knowledge can be

based. To begin with, positing that the quality of lexical organization depends on the degree of shared commonality of associative meaning is in effect positing that semantic organization is entirely and solely linguistically-based. However, more recent WA studies from the socio-cultural line of research have reached a different conclusion. For example, Yoshida (1990) has noticed that Japanese college students tend to produce language-dependent associative responses to SW from the culture category but not so much to culturally-neutral stimuli. Other researchers from the cultural paradigm (e.g., Szalay, 1984; Szalay & Brent,1967; Szalay & Windle, 1968) further argued that the acquisition of cultural concepts is accompanied by cognitive restructuring of the conceptual system. Szalay et al. earlier research (1968, 1972), for example, uncovered differences in the response patterns of Korean participants traceable to the influence of the culture in which the L2 was acquired and not so much to the language of response itself. Thus, Korean participants who acquired English in an American setting and experienced the culture of the L2 in the process of acquisition produced responses that were very similar to NS' associations. By contrast, there was a very high incidence of intergroup association patterns for the Korean learners who didn't have a cultural experience, regardless of whether the participants responded in their L1 or in English. This finding was interpreted to indicate that L2 learners' conceptual system was not very likely to reorganize without extensive immersion experience in the L2 culture, compared to the conceptual system of participants who had such an exposure. In the main, the conclusions drawn by the researchers working with WA data in the socio-cultural paradigm convincingly show that meaning organization is not only linguistically-driven but it is also strongly influenced by a host of extra-linguistic factors. This general conclusion was also strongly supported by numerous examples in my data which showed that using native-like typicality of associative connections to measure quality of L2 learners' knowledge is an approach that, firstly, does not have the potential to distinguish between different levels of proficiency because it is only sensitive to the broad distinction between NS and non-NS of English. Secondly, in my view, it is an unproductive line of research

not only because for more than two decades it has not arrived at any novel findings, but also because in its approach to the analysis of WA data it fails to acknowledge findings of other WA research paradigms.

The results of my analysis definitely suggested that even at an advanced level of language proficiency, L2 learners do not and probably cannot reach the extent of typicality characteristic of NS' associative behavior, which does not necessarily mean that their degree of lexical connectivity is of a lesser quality than that of NS. This conclusion was further supported by the comparison of the number of associations that each group of participants generated to the SW used in the study. The analysis revealed that on average the advanced learners generated as many associations (M = 121) as the NS did (M = 112), even a bit more, whereas the intermediate learners were able to produce a significantly smaller number of responses (M = 65). Apparently, there is a strong relationship between the number of different things and ideas that an individual may associate with any word and the size of his/ her lexicon. This hypothesis was additionally supported by examining the bivariate Pearson correlation between vocabulary size and number of associative responses (r = .83, p = .000), which indicated a strong interdependence between the two variables.

Finally, a closer look at the association data prompted me to examine how the degree of within-group consistency of the WA domain compared across the three groups. It was easily noticeable in the data that just like most NS tended to maintain common word association networks, the L2 learners also had stable patterns of commonality of associative responses that did not, however, resemble the ones of NS. Therefore, I compared the degree of each group's typicality of associative responses as a measure of how well a word was integrated in the lexicon of the NS, advanced, and intermediate learners, i.e. the underpinning assumption was that degree of commonality of associative behavior, rather than degree of nativelike commonality, could be a more sensitive index of language proficiency. This assumption was confirmed by the analysis of each group's commonality of associations. Moreover, as an

approach to the analysis of WA data, it accounted equally well for the difference in the level of proficiency of the L2 learners, as well as for the linguistic and the extra-linguistic factors that influence associative relationships. Not surprisingly, the advanced learners' consistency of WA responses was as strong as the NS' commonality of responses (though of different nature), whereas the intermediate learners' strength of the WA domain was noticeably less consistent compared to NS' and advanced learners' semantic stability of associations. This finding was in full agreement with the comparison of the size of the associative domain (number of responses) of the three groups of participants, which led to the conclusion that these two variables provide much more meaningful estimate of *depth* of lexical knowledge than native-likeness of associations.

Metacognitive awareness. The third dimension of lexical competence that was proposed in the model was metacognitive awareness. It was argued that this dimension performs a bridging function between the microlevel and the macrolevel of lexical knowledge in that it accounts for language users' ability to constantly monitor how much they know about a word (as a phonetic/ or orthographic entry, phonological, morphological, syntactic, and semantic entry) when they say that they know it. One of the advantages of using a rating scale and verified responses is that, on the one hand, such a test design allows the participants to rate their knowledge of words on a scale that reflects partial as well as precise lexical knowledge. On the other hand, comparing participants' self-perception of the words they think they know with the amount of words they actually responded correctly makes it possible to draw a number of conclusions about their degree of metacognitive awareness, regarding their overall state of vocabulary.

In discussing a metacognitive framework for the development of language skills, Bialystok and Ryan (1985) suggest that it encompasses two components, analyzed knowledge and cognitive control, which represent two interrelated aspects of metalinguistic performance; nonetheless, they should be distinguished from each other. The authors also argue that

information that appears in the form of unanalyzed knowledge is not subject to intentional manipulation because it is used routinely, whereas analyzed knowledge is used creatively with deliberate attention resources devoted to the task. The scholars further suggest that our linguistic knowledge varies in the degree to which it is analyzed, dependent on the demands of the linguistic task itself. By and large, there is an inverse relationship between the availability of context in the task and the amount of analyzed knowledge that is required, i.e. the more context is available, the less pressure there is on language learners to exercise analyzed knowledge since context provides alternative routs to meaning extraction. On the other hand, the cognitive control dimension of the metalinguistic framework of linguistic skills serves an information coordination function where monitoring procedures are required to oversee the manipulation of several facets of a specific task.

The proposed framework, which the authors argue is necessary for understanding the nature of language proficiency and development, has a specific relevance to the present study. In the experiment, the participants were required to exercise both analyzed knowledge, since the SW had to be recognized in isolation, as well as cognitive control, since they were required to coordinate several bits of information regarding the features of the SW before responding with a synonym, brief explanation, or translation. Therefore, it was assumed that the point of intersection between these two metalinguistic skills can be accounted for by comparing participants' self-reported knowledge with their actual knowledge of the SW. In addition, it was important to find out whether L2 learners' language proficiency would have specific effects on their ability to exercise adequate control procedures and perform well on the lexical task to the extent permitted by the degree of their analyzed knowledge. Interestingly, the results revealed that in spite of the different levels of language proficiency and, respectively, various quality and quantity of lexical knowledge, all participants were equally skilled in making relatively difficult metalinguistic decisions when responding to the SW (see Figure 2).

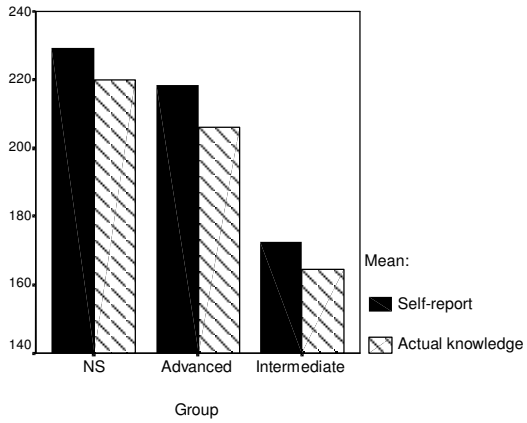


Figure 2. Metacognitive awareness of NS, L2 advanced, and intermediate learners of English.

In other words, as shown in the graph, the three groups of participants showed heightened degree of metacognitive awareness, regardless of their level of language proficiency. This can be, to a large extent, explained by their overall cognitive development (inasmuch as all participants were adults), which, respectively, triggers metalinguistic awareness. In addition, this finding is in agreement with the hypothesis advanced by Bialystok and Ryan (1985) that if metalinguistic skill is taken to be the control required to analyze a structured representation of a language, then adult L2 learners, who have already had an experience of acquiring one language, should find it relatively easy to solve metalinguistic problems in the new language. The only restriction that the authors suggest to apply is the extent to which learners have analyzed knowledge. Apparently, it should be expected that L2 learners at an intermediate, or higher, level of language proficiency can exercise as heightened degree of analyzed knowledge and cognitive control in the new language as NS can in their native language.

The model. The last research question that was investigated in the experiment concerned whether the three-dimensional model had the potential to fully capture the variation of lexical knowledge among language users, when their level of proficiency was taken into account. This question will be addressed in great detail in Chapter 3 but, at this point, suffice it

to say that the full model accounted for virtually all the variance, over 90%, among the group of NS, advanced learners, and intermediate learners of English. This analysis added strong empirical support to the proposed three-dimensional framework of studying and assessment of lexical competence, which outlined the theoretical basis for the model, and proved its relevance and usefulness for practical application. However, the present experiment also raised some interesting questions that also need to be addressed, for example:

1) Is it necessary to obtain information about all measures associated with the three dimensions of lexical competence or is it possible that some measures are more revealing of the overall state of learners' vocabulary than others?
2) Is it plausible to identify a smaller set of predictors of lexical knowledge, instead of using seven variables, which can predict almost the same amount of variation among the three groups of participants and, at the same time, would be easier to work with?
3) Given the wealth of WA data collected in the experiment, do NS and non-NS connect words in their mental lexicon along similar qualitative and quantitative patterns or is the organization of their mental lexicon "fundamentally different", as often claimed in the L2 literature?
4) Do the qualitative and quantitative patterns of lexical organization depend on language users' degree of familiarity with words?
5) Do specific features, such as lexical category or frequency of occurrence of words, influence in any way the type of connections among words that NS and non-NS develop in their mental store or is it the level of proficiency and degree of familiarity with words that are the driving force behind lexical organization?

These questions will be addressed in great detail in the following chapters.

CHAPTER 3

MODELS OF LEXICAL KNOWLEDGE ASSESSMENT OF EDUCATED NATIVE AND NON-NATIVE ADULT SPEAKERS OF ENGLISH

The results of the analysis presented in Chapter 2 challenged me to look further into the relationship between the factors used to measure the three dimensions of vocabulary knowledge and their predictive power. In other words, I decided to examine the potential of each variable or combination of variables to account for the greatest amount of variance in the lexical knowledge of NS and non-NS taking into consideration differences in their language proficiency. The underlying assumption was that such an analysis would allow me to propose a model for lexical knowledge assessment of adult NS and non-NS of English by identifying for each group the smallest set of variables that can be employed as practically efficient predictors of their lexical knowledge. I should hasten to add here that in the context of this chapter, a "model" only means a "set of predictors" of lexical knowledge, not an explanatory framework of lexical competence, such as the one proposed in Chapter 2. In addition, it was believed that this approach would also allow for examining the proposed models comparatively with regard to whether the lexical knowledge of NS, advanced, and intermediate learners can be studied by means of a common set of predictors or whether difference in language proficiency make certain models more cost and time efficient for one group but not for the other.

Therefore, the analysis that will be presented in this chapter is predictive in nature, i.e. it will serve the practical purpose of identifying the smallest set of predictors of lexical knowledge

that would, ideally, be as good as the full model. In Chapter 2, I proposed a three-dimensional model of studying lexical competence of language users at different levels of language proficiency. The model was also empirically tested by an experiment specifically designed to examine whether the proposed dimensions (quantity, quality, and metacognitive awareness) fully capture the overall state of lexical knowledge of the participants involved in the study. Each dimension was measured by variables that are commonly used in lexical research to investigate knowledge of vocabulary. Quantity was studied by examining participants' size of vocabulary and their knowledge of words from different frequency bands. Quality was examined by looking at several quantitative features of participants' WA domain, such as degree of nativelike commonality of their associations, number of associations, and within-group consistency of the WA domains. Finally, metacognitive awareness was measured by the extent to which language-users' self-reported knowledge reflected their actual knowledge of words. The analysis reported in Chapter 2 raised several interesting questions that, to my knowledge, have never been addressed in L2 lexical research. For example, 1) for the practical purpose of assessment, is it necessary to obtain information about all measures associated with the three dimensions of lexical competence or is it possible that some measures are more revealing of the overall state of learners' vocabulary than others? And 2) can we identify a smaller set of predictors of lexical knowledge, instead of using as many as seven variables, which can predict almost the same amount of variation among the three groups of participants and, at the same time, would be easier to work with?

To address these issues I used regression analysis as a statistical procedure that, dependent on the goals of a study, can be applied for predictive as well as for explanatory purposes. By and large, there are different opinions about the relationship between predictive and explanatory research, as well as numerous attempts to delineate their status in research studies. However, the distinction that most researchers seem to agree on is that predictive

research puts emphases on practical application, whereas explanatory research focuses primarily on understanding phenomena (Pedhazur, 1997, p. 196). Consequently, despite the apparent relationship between the two types of research, the importance of distinguishing between both becomes acutely relevant when it comes to the interpretation of the implications of the regression analysis results. The decision to use this procedure in the study was motivated solely by practical reasons. In other words, for the purposes of assessment of vocabulary knowledge, I aimed at identifying the smallest set of variables that could predict the overall state of lexical knowledge of various proficiency groups, primarily, because working with many variables, though rewarding, was both demanding and time-consuming. In my previous analysis I used two variables to study each of the proposed dimensions; however, it is not logical to assume that each variable contributed equally well to the explained variation among the participants' vocabulary knowledge. Therefore, in this study it was hypothesized that certain combinations of factors could be as good predictors as the full model. In addition, I was interested in finding out whether the lexical knowledge of language users at different level of proficiency could be studied by means of a common set of predictors or whether the difference in language proficiency would make certain models more efficient for one group but not for the other. Finally, it was expected that the predictors identified by the regression analysis as "best" would also point to the dimension(s) that might occupy a privileged position in the overall state of language users' lexical knowledge when proficiency is taken into consideration.

The chapter is organized as follows: first, the method section introduces the research questions that will be examined in the chapter. Next, the results of several statistical procedures of variable selection are presented. Finally, the discussion section offers conclusions drawn from the application of the methods of variable selection and comments on the predictor models obtained through the statistical procedures.

Method

I set up to examine the relationship between the lexical knowledge of educated adults, NS and two groups of non-NS of English, and several predictor variables. Vocabulary knowledge, measured by participants' verified responses to 73 SW, was selected for a criterion. The predictor variables were as follows: (1) self-perception of vocabulary knowledge, (2) knowledge of words from various frequency bands (also referred to in the chapter as word frequency effects), (3) number of associations, (4) native-like typicality of associations, and (5) vocabulary size. These variables are frequently used in L2 lexical research to investigate different aspects of language users' knowledge of words; hence, the choice of predictors was based on previous research evidence of the relationship between the overall state of lexical knowledge and the selected variables. I was interested in finding out how each variable was related to participants' vocabulary knowledge, the extent to which the variation in the lexical knowledge could be predicted by each of the factors or combination of factors, and which variables, or rather set of variables, could be identified as the "best" predictors of the vocabulary knowledge of educated adult NS, L2 advanced, and intermediate learners.

Results

Multiple linear regression analysis was conducted to examine the relationships between the criterion and the predictors, with all variables being entered into the regression simultaneously. Using Cohen's guidelines for estimating the strength of the relationship between vocabulary knowledge and the explanatory variables, the number of NS participating in the study ($n = 30$) was sufficient to detect a large correlation among the variables ($\rho = .60$; $\alpha = .05$; $1 - \beta = .80$). The number of L2 advanced ($n = 17$) and intermediate learners ($n = 17$) was also satisfactory to detect large correlation and avoid type I error ($\rho = .75$; $\alpha = .05$; $1 - \beta = .80$). The distributions of the scores on all variables for the three groups was examined and was found to be normal. The examination of Cook's D and DF Beta revealed one outlier (Cook's D = 2.666;

DF Beta = 4.952) among the NS and one outlier among the L2 advanced learners (Cook's D = 2.44; DF Beta = 2.38) that deserved to be further examined with respect to their overall effect on the results of the analysis. However, the outliers did not show to have significant influence on the results of the analysis, so I did not consider removing those scores from the data. The descriptive statistics of the analysis of vocabulary knowledge, self-perception of vocabulary knowledge, knowledge of words from different frequency bands, number of associations, native-like typicality of associations, and vocabulary size for the three groups is presented in Table 4.

Table 4

Means and standard deviations for the groups of NS, L2 advanced, and intermediate learners of English

Source	*N*			*Mean*			*SD*		
	NS	Adv	Intrm	NS	Adv	Intrm	NS	Adv	Intrm
verified vocabulary knowledge	30	17	17	220	206	164	27.00	31.74	19.61
self-perception of vocabulary knowledge	30	17	17	229	218	172	26.90	26.77	19.85
vocabulary size	30	17	17	9447	8737	6033	2053.52	2746.55	1906
knowledge of words from various frequencies	30	17	17	170	157	106	43.92	56.70	38.91
native-like typicality of associations	30	17	17	369	198	103	85.10	88.53	70.01
number of associations	30	17	17	109	120	65	25.94	35.38	39.03

The bivariate relationship between the variables was studied by using Pearson product-moment correlations. The analysis revealed that for the NS all of the predictor variables were significantly correlated with verified vocabulary knowledge, i.e. self-reported knowledge of words

(r = .95, p = .000), size of vocabulary (r = .88, p = .000), knowledge of words from various frequency of occurrence (r = .85, p = .000), number of associations (r = .52, p = .003), and native-like typicality of associations (r = .44, p = .015). These relationships were even stronger for the advanced learners, with very high bivariate correlations between verified vocabulary knowledge and self-reported knowledge (r = .98, p = .000), vocabulary size (r = .98, p = .000), word frequency effects (r = .97, p = .000), native-like typicality (r = .74, p = .000), and number of associations (r = .96, p = .000). The correlations between the criterion and the predictors for the intermediate learners were also high, i.e. self-reported vocabulary knowledge (r = .96, p = .000), vocabulary size (r = .87, p = .000), knowledge of words from various frequency bands (r = .86, p = .000), native-like typicality (r = .79, p = .000), and number of associations (r = .77, p = .000). All correlations were significant, with the exception of the correlation between native-like typicality and self-reported of vocabulary knowledge for the group of NS (r = .33, p = .077), which suggested that for the purposes of lexical assessment, degree of associative commonality was weakly related to NS' self-reported knowledge of words. Pearson product-moment correlations for all groups are shown in the correlation matrix (see Table 5).

Table 5

Correlation matrix for the groups of NS of English (n = 30), L2 advanced (n = 17), and intermediate learners of English (n = 17)

Pearson Correlations	Verified vocabulary knowledge			Self–reported vocabulary knowledge			Vocabulary size			Knowledge of words from various frequency bands			Native-like typicality of associations			Number of association		
	NS	Adv	Intrm	NS	Adv	Intrm	NS	Adv	Intrm	NS	Adv	Intrm	NS	Adv	Intrm	NMS	Adv	Intrm
Verified vocabulary knowledge	1	1	1															
Self-reported vocabulary knowledge	.95*	.98*	.96*	1	1	1												
Vocabulary size	.88*	.98*	.87*	.84*	.93*	.76*	1	1	1									
Knowledge of words from various frequency bands	.85*	.97*	.86*	.80*	.92*	.75*	.99*	.99*	.98*	1	1	1						
Native-like typicality of associations	.44*	.74*	.79*	.33	.69*	.71*	.46*	.78*	.87*	.48*	.80*	.80*	1	1	1			
Number of associations	.52*	.96*	.77*	.43*	.91*	.66*	.55*	.97*	.90*	.55*	.97*	.85*	.86*	.86*	.93*	1	1	1

*p < .05

The regression analysis pointed to a statistically significant relationship between vocabulary knowledge and the predictor variables for the NS, $F(5, 24) = 67.90$; $p = .000$, for the advanced learners, $F(5,11) = 216.31$, $p = .000$, and for the intermediate group, $F(5,11) = 80.49$, $p = .000$. The results suggested that the five predictors used in the analysis, i.e. self-perception of vocabulary knowledge, vocabulary size, knowledge of words from various frequency bands, number of associations, and native-like typicality of associations could explain virtually all the variation in the vocabulary knowledge among the NS (Adj. $R^2 = .92$), the L2 advanced learners (Adj. $R^2 = .99$), and intermediate learners (Adj. $R^2 = .96$). However, when the other variables were held controlled, the only significant relationship for the group of NS was between their verified and self-reported lexical knowledge (partial correlation = .83, $p = .000$). The same observation held for the L2 advanced learners (partial correlation = .83, $p = .000$), as well as for the L2 intermediate group (partial correlation = .94, $p = .000$). Therefore, it was concluded that verifiable self-report of word familiarity alone, independent of all other variables, could explain a significant proportion of the variation in the vocabulary knowledge of language users who are at different levels of language proficiency.

This result invited for a further exploration of whether verifiable self-report was the single "best" predictor of vocabulary knowledge for the studied groups or whether the combination of several predictor variables could provide good models for practical assessment of their knowledge of words. Moreover, all of the variables used in the study were highly intercorrelated, which also pointed to the possibility of selecting from a pool of predictors a smaller set that would be as efficient as the entire set of factors. Therefore, in an attempt to discover the smallest set of variables that could account for as large amount of the variance in the lexical knowledge of the three groups of participants as the entire set, I used several methods of variable selection, which allowed for a fine evaluation of the used factors from several perspectives before identifying the "best" smallest set. Moreover, it is axiomatic in the

regression analysis literature that there isn't one "best" method for variable selection, nor is there a "best" set of predictors. Rather, the results of several regression analysis methods would point to particular subsets of predictors, whose selection as being the "best" should be grounded in theory and previous research evidence. Of the various predictor selection procedures, I used three methods that are commonly used in variable selection to obtain the "best" regression equation, i.e. stepwise selection, all possible regressions, and direct search on ***t*** regression. At the same time, I was cautious with the interpretation of the results that each of these procedures yielded considering the fact that for model construction purposes theory should play a leading role in selecting the most meaningful set of variables (Pedhazur, 1997).

Stepwise selection

With stepwise selection procedure, every time a variable is added to the model, all of the previously added variables are re-evaluated; hence, making it possible to remove predictors that were earlier brought into the equation but have subsequently lost their power (Pedhazur, 1997). In this analysis I used liberal entry criteria (probability of F-to-enter [PIN] = .10) and removal criteria (probability of F-to-remove [POUT] = .15) in order to allow as many variables as possible to enter the equation and be reassessed.

For the group of NS, all variables entered the equation at PIN = .10 but at POUT = .15, the variables associated with number of associations and word frequency effects were removed. This suggested that having verifiable self-report of vocabulary knowledge, vocabulary size, and native-like typicality of associations in the model can explain a significant proportion of the variation in the lexical knowledge of NS, $F(3, 26) = 122.56$, $p = .000$, Adj. $R^2 = .93$. In fact, a model consisting of just those three variables was shown to be as good as the full model of five variables, $F(5, 24) = 67.90$, $p = .000$, Adj. $R^2 = .93$. This only means that collecting information about the number of associations that NS can produce, as well as their knowledge

of words from various frequency bands did not contribute in any significant way to predicting the overall state of their vocabulary knowledge in the presence of the other three variables.

For the non-NS, both the L2 advanced and intermediate learners, the only two variables that entered the equation and were retained in the model were self-perception of vocabulary knowledge and vocabulary size. The analysis revealed that these two variables could be powerful predictors of the general state of vocabulary knowledge of the advanced, $F(2, 14) = 613.32$, $p = .000$, Adj. $R^2 = .99$, as well as the intermediate learners, $F(2, 14) = 248.77$, $p = .000$, Adj. $R^2 = .97$, in that they could explain virtually all the variance in their lexical knowledge. Apparently, based on these results, word frequency effects, as well as the effort and time spent on testing associative behavior of advanced and intermediate learners did not prove to have contributed to predicting the general state of vocabulary knowledge of L2 learners at those two levels of proficiency. Of course, this conclusion holds true only if information about learners' vocabulary size and self-perceived word knowledge were previously obtained.

*Direct search on **t** regression*

I further proceeded with the search for the "best" predictors of vocabulary knowledge by using a second method of variable selection, direct search on ***t*** regression. This is a procedure where, based on the full model, all variables are ordered by their absolute value of ***t*** (**| *t* |**). After that, separate regression equations are computed by sequentially adding one variable at a time, in the rank order of the variables. Fortunately, this procedure confirmed the models for the three groups obtained by using the stepwise selection procedure. That is, for the NS, the model which included self-perception of vocabulary knowledge, vocabulary size, and native-like typicality of associations was the least biased model ($C_p = 2.01$[1]), compared to the models which included more variables (see Appendix F). Moreover, the proportion of the explained variation by the

[1] C_p = Mallow's statistic (The desired values are values close or smaller than k $^{\text{number of predictors}}$ + 1, i.e. for this analysis ($C_p = 5+1 = 6$)

three-predictor model was higher (Adj. R^2 = .93) than that of the four-variable model (Adj. R^2 = .92) or the full model (Adj. R^2 = .92). The results confirmed the findings of the stepwise selection method by showing that, based on this data set, verifiable self-report, vocabulary size, and native-like typicality of associations were the "best" predictors of lexical knowledge for NS of English. Adding a measure of word frequency effects and number of associations to the model insignificantly decreased the effect size but noticeably increased the total bias (Adj. R^2 = .92, C_p = 6.00).

For the two groups of non-NS, the direct search on ***t*** procedure was also in agreement with the previously obtained results. Verifiable self-report and a measure of vocabulary size were the two variables that not only explained almost all the variance among the advanced and intermediate learners (Adj. R^2_{Adv} = .99, Adj. R^2_{Intrm} = .97), but also provided the least biased subset of predictors ($C_{p\ Adv}$ = 1.33; $C_{p\ Intrm}$ = .32). Adding other variables not only did not contribute to the proportion of variation explained but only led to an increase in the total bias of the models (see Appendix G and Appendix H).

All possible regressions

Finally, the search for the "best" possible set of predictors of lexical knowledge proceeded with calculating all possible regression equations, which in this study amounted to 32 equations (all possible equations = $2^{k\ (number\ of\ predictors)} = 2^5 = 32$). The selection criteria for each group of models, from all possible equations, were based on meaningful increments in R^2 or Adj. R^2 and the total bias in the regression estimation measured by Mallow's statistic (C_p). The goal of the analysis was also to find a model that would comprise of a set of predictors which would explain the highest proportion of the variance in participants' lexical knowledge, measured by R^2, Adj. R^2 (which removes the sample bias associated with R^2), and would have the smallest total bias, measured by C_p statistic. Table 6 summarizes the results from the all possible regressions procedures.

Table 6

Summary of the analysis of all possible regressions

Number of predictors	Regressor subset	NS			Advanced			Intermediate		
		R^2	*Adj.* R^2	C_p	R^2	*Adj.* R^2	C_p	R^2	*Adj.* R^2	C_p
none		0	0	0	0	0	0	0	0	0
1	(X1) Self-reported VK	.90	.90	1.15	.95	.95	33.61	.93	.92	9.75
1	(X2) Vocabulary size	.77	.76	50.08	.96	.98	22.29	.76	.75	77.66
1	(X3) Word frequency effects	.73	.72	65.15	.95	.94	37.88	.74	.72	87.31
1	(X4) Native-like typicality	.19	.17	258.87	.55	.52	466.94	.62	.59	136.99
1	(X5) Number of associations	.26	.24	234.36	.92	.91	72.02	.59	.57	147.68
2	X1 X2	.93	.92	-9.22	.99	.99	-10.67	.97	.97	-11.68
2	X1 X3	.93	.92	-8.97	.99	.98	-7.30	.97	.97	-11.12
2	X1 X4	.92	.92	-7.66	.96	.95	13.37	.95	.94	-1.22
2	X1 X5	.92	.91	-6.39	.98	.98	-0.42	.96	.95	-4.87
2	X2 X3	.78	.77	42.40	.96	.96	19.78	.76	.73	75.24
2	X2 X4	.88	.75	47.55	.96	.96	18.52	.76	.73	74.54
2	X2 X5	.77	.75	47.54	.96	.96	19.73	.76	.73	75.30
2	X3 X4	.73	.72	62.55	.95	.94	31.40	.77	.73	73.65
2	X3 X5	.73	.71	62.02	.95	.94	33.83	.74	.71	83.23
2	X4 X5	.26	.21	232.35	.94	.93	40.32	.63	.58	130.21

3	X1 X2 X3	.93	.92	-11.23	.99	.99	-12.69	.97	.97	**-13.96**
3	X1 X2 X4	.93	.93	-13.99	.99	.99	-12.75	.97	.97	-13.77
3	X1 X2 X5	.93	.92	-12.83	.99	.99	-13.02	.97	.97	-13.68
3	X1 X3 X4	.93	.93	-13.72	.99	.98	-9.65	.92	.97	-13.51
3	X1 X3 X5	.93	.92	-12.98	.99	.98	-9.93	.97	.97	-13.51
3	X1 X4 X5	.92	.91	-9.75	.98	.98	-5.88	.96	.95	-6.90
3	X2 X3 X4	.79	.76	39.32	.96	.95	16.46	.77	.72	70.59
3	X2 X3 X5	.79	.76	39.77	.96	.95	16.72	.76	.71	73.04
3	X2 X4 X5	.77	.74	44.52	.97	.96	11.23	.77	.72	69.40
3	X3 X4 X5	.73	.70	60.00	.96	.95	18.03	.78	.72	67.56
4	X1 X2 X4 X5	.93	.92	-16.00	.99	.99	-15.98	.97	.96	-15.96
4	X1 X3 X4 X5	.93	.92	-15.72	.99	.98	-13.99	.97	.96	-15.98
4	X1 X2 X3 X5	.93	.92	-14.99	.99	.99	-15.12	.97	.96	64.47
4	X1 X2 X3 X4	.93	.92	-2.03	.99	.99	-14.75	.97	.96	-15.85
4	X2 X3 X4 X5	.79	.75	37.32	.97	.96	9.04	.78	.71	-15.55
5	X1 X2 X3 X4 X5	.93	.92	-18.00	.99	.99	-18.00	.97	96	-286.00

One-predictor models. The results obtained from conducting all possible regressions revealed that the one-predictor models had unacceptably high total bias, especially when used for the evaluation of non-NS' lexical knowledge. However, if a choice has to be made which of the five variables to be used for testing purposes, while completely disregarding the issue of total bias, verifiable self-report explained the highest percentage of the variance in the vocabulary knowledge of each of the three groups (Adj. R^2_{NS} = .90, Adj. R^2_{Adv} = .95, Adj. R^2_{Intrm} = .92), followed by vocabulary size (Adj. R^2_{NS} = .76; Adj. R^2_{Adv} = .98; Adj. R^2_{Intrm} = .75). Interestingly, even though self-report and vocabulary size were found to be strongly correlated, for the advanced learners self-report improved the predictive power of the analysis with as little as 3%, while the increments for the intermediate learners and the NS' group were between 17% and 14%, respectively. This implied that the advanced learners tended to have slightly higher metacognitive awareness of how much they knew about a word when they were tested to recognize it in isolation than the intermediate learners and the NS did, probably, as a result of their generally heightened overall metacognition when operating in a foreign language. This finding supported, to some extent, Bialystok and Ryan's (1985) hypothesis that, by virtue of the processes involved in learning a second language, metalinguistic skill should be higher for adult speakers of one language who are in the process of learning another, as well as for learners who already know two languages. The advanced and the intermediate learners met both conditions outlined by the authors, while the NS were found to be mostly monolingual (see description of participants in Chapter 2). Nonetheless, this distinction should not be overemphasized because all participants in the study showed a very high degree of metacognitive awareness and in the context of the present research the differences among the three groups are more artificial than meaningful.

Not surprisingly, the most biased predictors of lexical knowledge in the one-predictor models appeared to be native-like typicality of associations and number of associations.

Particularly, for the group of NS, having native-like typicality as a single factor of predicting their knowledge of words dramatically reduced the proportion of the explained variance to 17% only (Adj. R^2_{NS} = .17) and showed a very high degree of total bias ($C_{p\ NS}$ = 258.87). As for the non-NS, having native-like typicality of associations in a model not only did largely reduce the explained variation among the L2 learners (Adj. R^2_{Adv} = .52; Adj. R^2_{Intrm} = .14), but it also did not show sensitivity to distinguishing between language users at various levels of language proficiency. It also deserves to be mentioned that the magnitude of the total bias of the models containing native-like typicality was so high ($C_{p\ Adv}$ = 466.94, $C_{p\ NS}$ = 136.99) that I do not recommend native-like typicality of associations, though traditionally employed by L2 researchers, to be used as a single predictor variable of lexical knowledge neither for NS, nor for non-NS. The analyses further showed that the number of associations, rather than their native-like typicality, could be a better predictor of lexical knowledge because it increased the amount of the explained variation and decreased the bias, though not to a satisfactory level ($C_p = k^{\text{number of predictors}}+1 = 6$). Finally, the results revealed that a measure of word frequency effects had strong predictive power for the three groups of language users, yet less powerful than self-reported knowledge of words and vocabulary size. This comes to suggest that when a measuring instrument includes vocabulary items from several frequency bands, verifiable self-report or a measure of vocabulary size can better explain the variance in the knowledge of lexical items of NS and non-NS than a measure of word frequency effects alone.

In sum, models containing one predictor variables, e.g. self-perception of vocabulary knowledge, vocabulary size, word frequency effects, native-like typicality of associations, and number of associations showed an overall good potential to explain a large amount of the variation in the lexical knowledge of English language users. However, a major disadvantage of such models is their unacceptably high overall bias, i.e. bias that is associated with the residuals, number of participants, number of predictors, p-values, etc. Therefore, based on the

results of all possible regression method, I would strongly discourage the use of any one-predictor models for the assessment of lexical knowledge of NS and non-NS alike.

Two-predictor models. The two-predictor models obtained by means of all possible regression method completely supported the conclusions drawn from the application of the stepwise method of variable selection and the direct search on **t** regression. As for the NS' group, all models that contained verifiable self-report in combination with the other variables showed very high effect size and low total bias. In those models, the combination of self-report with the rest of the variables accounted for between 91% and 92% of the variance associated with the NS' lexical knowledge and the magnitude of total bias of the models was minimal. However, combinations between vocabulary size, word frequency effects, native-like typicality of associations, and number of associations tended to reduce the effect size with 15% to 71% and this reduction was especially noticeable when models comprising of native-like typicality and number of associations were compared to combinations of the other variables. Apparently, association tests would not reveal much about the lexical knowledge of educated NS of English; yet, when in combination with a size measure, the two variables (native-like typicality of associations and number of responses) did contribute to the predictive power of the breadth measure.

The results for the non-NS also supported the findings of the stepwise and direct search on **t** methods. For the advanced and intermediate learners, the models which contained verifiable self-report and vocabulary size could "best" explain the total amount of variation in the lexical knowledge of the L2 learners at these two levels of language proficiency. Moreover, verifiable self-report in combination with the other three variables maintained the same high levels of effect size (Adj. $R^2 > .94$) and acceptable degree of total bias ($C_p < 6$), with the exception of the model that contained self-report and typicality of word associations for the advanced group ($C_p = 13.97$). This comes to confirm the conclusions drawn in Chapter 2 that for

L2 advanced learners, decisions about their lexical knowledge made on the basis of the degree of native-like typicality their associative behavior shares with that of NS not only did not contribute to our understanding of the nature of their lexical knowledge but further obscured the issue of the usefulness of association tests for the purposes of studying the lexical knowledge of, primarily, L2 advanced learners. The analysis did show, though, that the model containing native-like typicality of associations, in addition to verifiable self-report, worked well for the intermediate learners (Adj. $R^2 = .94$, $C_p = -1.22$), which was probably because the intermediate learners, in general, failed to maintain consistency of their WA domain. In any event, number of associations, rather than shared native-like typicality, in combination with verifiable self-report comprised a model that yielded results as good as the models containing self-report and vocabulary size or word frequency effects.

Three-predictor models. The three-predictor models did not show to add any significant improvement to the effect size of the two-predictor models, as discussed in the preceding paragraphs. It was found that for the three groups of participants the models which contained verifiable self-report as a variable had the strongest predictive power and the lowest degree of total bias. In addition, the three-predictor models seemed to have the greatest potential to reveal differences that can be attributed to the language proficiency of the participants. Here the distinctions between the three levels of language proficiency could most clearly be seen in the specific combinations of predictors. While the "best" model for the NS comprised of self-report, vocabulary size, and native-like typicality of associations (Adj. $R^2_{NS} = .93$, $C_{p\,NS} = -13.99$), the "best" model for the advanced learners had number of associations in place of native-like typicality of associations (Adj. $R^2_{Adv} = .99$, $C_{p\,Adv} = -13.02$), and the intermediate learners had word frequency effects as a third variable in the model (Adj. $R^2_{Intrm} = .97$, $C_{p\,NS} = -13.96$). This result led to the conclusion that in the presence of self-perception and size of vocabulary, the choice of an additional factor should primarily depend on the language proficiency of the

learners. Association tests, which are usually designed to yield results about the extent to which the associations generated by L2 learners displayed a degree of native-likeness, appeared to be the least efficient and the most biased instruments for assessment of lexical knowledge across the three groups. Models that did not include verifiable self-report as a variable, but had three-predictor combinations of size, word frequency, native-like typicality, and number of associations showed reduction of their overall effect size with approximately 20% for both the NS and the L2 intermediate learners. However, they preserved their strong predictive power for the advanced learners. This suggested that for the intermediate students showing evidence of familiarity with vocabulary items from various frequency bands would yield more dependable results regarding their lexical competence than using more sophisticated instruments for measuring the quality and quantity of their knowledge. As for the advanced learners, any combination of the proposed five variables maintained a very high degree of effect size (Adj. $R^2_{Adv} > .95$) across all three-predictor models, which actually reflected the more sophisticated nature of their lexical competence.

Four-predictor models. The full model. The four-predictor models and the full model did not add meaningful increments to the effect size of the models with a smaller number of predictors. All four-predictor models, which included self-report as a factor in addition to any combination of three other variables maintained a very high effect size (between 92% and 99%) across the three groups and low values of total bias ($- 2.03 > C_p < -16.00$). The only exception was one model for the intermediate group (X1 X2 X3 X5), which had high predictive power (Adj. $R^2_{Intrm} = .96$), however, an unacceptable value of total bias ($C_{p\ Intrm} = 64.47$). Interestingly, not having self-report in the four-predictor model, reduced the potential of explaining the greatest amount of variation for the NS by 16% (Adj. $R^2_{NS} = .75$) and 26% for the intermediate learners (Adj. $R^2_{Intrm} = .71$), with a trivial effect for the advanced learners (Adj. $R^2_{Adv} = .96$). It should be added, though, that the figures only show reduction in the effect size, which does not mean that

the resultant effect sizes are small. On the contrary, following Cohen's guidelines, an effect size that equals .25 (R^2 = .25) is considered to be large, which in turn suggests that the discussed models are still powerful. However, if a choice should be made for test design purposes or assessment purposes, I propose the inclusion of a self-report category in all models, which would yield a high effect size and low total bias.

Interestingly, the full model, which contained all of the proposed variables, had the lowest values of total bias ($C_{p\ NS}$ = -16.00; $C_{p\ Adv}$ = -16.00, $C_{p\ Intrm}$ = -286.00) and very high effect sizes (Adj. R^2_{NS} = .92; Adj. R^2_{Adv} = .99; Adj. R^2_{Intrm} = .96). It virtually explained all the variance in the lexical knowledge among NS and non-NS alike, which led to the conclusion that the full model has a strong potential to account for the variation in the vocabulary knowledge of language users' who are at different levels of proficiency. The results of all possible regressions method added firm support to the argument made in Chapter 2 that the proposed three-dimensional model fully captures the macrolevel of lexical competence of speakers of English at various levels of language proficiency.

Conclusion

The analysis presented in this chapter had the practical goal of examining the potential of each variable, as well as various combinations of variables, to predict the greatest amount of variance in the lexical knowledge of NS, L2 advanced, and intermediate learners. The idea was to propose a model, comprising of the "best" set of predictors, to serve the purpose of vocabulary assessment of adult NS and non-NS of English by identifying the smallest set of variables that would be as practically efficient as the full model. The expectation was that this approach would also allow for examining the various models comparatively and help find out whether the lexical knowledge of NS, advanced, and intermediate learners can be studied by a common set of predictors or whether differences in the language proficiency make certain models more efficient for one group but not for the other. Two major questions were addressed

in the analysis, i.e. whether, for the practical purpose of assessment, it is necessary to obtain information about all measures associated with the three dimensions of lexical competence or whether it is possible that some measures are more revealing of learners' vocabulary than others? And the second question addressed was whether it would be plausible to identify a smaller than the full set of factors that could predict the same amount of variation among the three groups of participants and, at the same time, would be easy to work with?

To answer these questions I used several different procedures of regression analysis, i.e. stepwise method, direct search on ***t*** regression, and all possible regressions, since there is no one "best" method of identifying the most efficient predictor(s) of a given phenomenon. At the same time each procedure approaches the analysis of the variables in a different way, so it was assumed that the meeting point of the three procedures would provide the answers to the posited research questions. The decision about the usefulness of the models was based on three statistics, two of them are traditionally associated with the assessment of the practical significance of a model (R^2 and Adj. R^2) and the third one is used for the assessment of the total bias in a model (C_p).

Each of the regression procedures yielded a number of predictor sets whose practical significance was as good as the one of the full model; however the total bias of some of these models, which is associated with the residuals, number of subjects, number of predictors, p-values, etc., was unacceptably high. As a matter of fact, the practical significance of all models showed, according to Cohen's guidelines, large effect sizes ranging from Adj. $R^2 = .24$ to Adj. $R^2 = .99$, which is virtually all the variation that a predictor variable or set of variables can explain in the lexical knowledge of the participants. Therefore, the selection of the proposed models was primarily based on the magnitude of their total bias, which, regrettably, is rarely taken into consideration in L2 lexical research.

What was immediately noticeable in the analysis was the unreliability of the one-predictor models. The three regression procedures identified these models as being the most biased models, though their effect sizes were very high. Therefore, I strongly discourage the use of any one-predictor models for the assessment of lexical knowledge of NS and non-NS alike. Interestingly, the smallest "best" set of predictors of the overall state of lexical knowledge of the non-NS, identified by the three regression procedures, was a two-predictor model comprising of verifiable self-report and vocabulary size. This model has several unquestionable advantages that deserve to be mentioned, i.e. 1) it consists of only two variables that are easy to work with, which makes the model time-efficient; 2) it works equally well for advanced, as well as for intermediate learners; 3) it has the potential to account for as high proportion of the variation among the L2 learners at these proficiency levels as the models containing more variables; 4) it shows low values of total bias. Some of the three- and four-predictor models also showed good characteristics of practical applicability and low total bias. However, neither those models, nor the full five-predictor model proved to have meaningful increments in the variation explained, nor did they substantially improve the magnitude of the acceptable total bias. This leads to the conclusion that a five-predictor model does not necessarily explain better the variation in the lexical knowledge of language users at different level of language proficiency, though it does provide a more sophisticated picture of the researched phenomenon. Apparently, when the SW selection is not based on a specific word frequency but includes words randomly selected form a wide range of frequencies, obtaining information about word frequency effects on participants' vocabulary knowledge or about their associative behavior is not time and resource efficient, nor is it improving the research results. Most probably, this kind of information is implicitly reflected in the self-report provided by the participants, especially considering their overall heightened metacognitive awareness.

Finally, it was hypothesized earlier in the chapter that the predictors identified by the regression analysis as "best" would also point to a hierarchical order of importance that the proposed dimensions might have in the overall state of language users' lexical competence. In this regard, the selected "best" predictor set of variables suggests that metacognitive awareness and vocabulary size seem to be the two dimensions that are somewhat higher up in the hierarchy of vocabulary knowledge than quality. However, I should add here that the design of the study as well as the data I worked with do not allow me to draw any definitive conclusions in this regard. At this point I can only hypothesize that for the purposes of assessment of lexical knowledge, a measure of verifiable self-reported familiarity with the SW and a measure of vocabulary size can account equally well for differences in the general state of vocabulary knowledge among NS and non-NS alike.

On a final note, in his book *Assessing Vocabulary*, J. Read (2000) outlined various reasons for testing "depth" of vocabulary knowledge by using WA tests "as an effort to go beyond conventional tests of word meaning, while still employing simple item type" (p. 248). He also pointed out that L2 vocabulary assessment does need other measures that are practical to use and valid indicators of how well words are known. The author further emphasized that these measures should emerge from research that is giving a more macrolevel angle to studying the general state of learners' vocabulary. The analysis presented in this chapter is in effect such a proposal, which was guided by theory and, at the same time, was rooted in previous lexical research. It also contributes to the L2 research field with a novel finding that offers another perspective on efficient models of vocabulary assessment. Meara (1996), for example, argued that vocabulary size is the best measure of a macrolevel type that works well for L2 learners with vocabulary sizes of up to 5,000 – 6,000 words. Beyond this level, he further suggested, a measure of vocabulary size becomes less important than a measure of how well the vocabulary is organized in the learner's mind. In light of the reported analysis, I would argue that vocabulary

size is not the “best” predictor of learners’ lexical knowledge. Rather, a measure of vocabulary size, based on a random selection of SW from various frequency bands, combined with a measure of metacognitive awareness, i.e. some form of verifiable self-report of SW familiarity that would trigger analyzable knowledge and cognitive control, form the “best” set of predictors that work equally well with small and large vocabulary sizes. While the analysis of participants’ WA was found to be of little practical usefulness in this line of investigation, I do believe that associative behavior has a great value from a psycholinguistic point of view, i.e. for examining how non-NS’ qualitative and quantitative patterns of associative meaning compare to those of NS. This idea will be further explored the chapters that follow.

CHAPTER 4

ORGANIZATION OF THE MENTAL LEXICON OF ADULT NATIVE SPEAKERS AND L2 LEARNERS OF ENGLISH AT AN ADVANCED AND INTERMEDIATE LEVEL OF LANGUAGE PROFICIENCY

Introduction

One of the questions frequently asked in L2 lexical research is whether there are fundamental differences in the way speakers of two or more languages organize their mental lexicon compared to NS' organization. In other words, for a long time researchers have been interested in finding out whether or not L2 learners organize their lexical knowledge much the same way as NS do, with the major differences being related to the size of their vocabularies rather than to the patterns of their organization. There is a growing body of L2 lexical research that tries to investigate some of these issues by using test instruments, namely word association (henceforth WA) tests, which have a long tradition in L1 psychological research. It is, actually, this particular line of research that laid the foundations of the theoretical platform behind the value of WA tests as research instruments in L2 studies. Most of the assumptions in L1 research derived from the notion that language associative behavior could be revealing of the cognitive processes of human thought. Cognitive psychologists from this tradition (e.g., Cramer, 1968) believed that associative responses to SW reflect the functioning of thought processes of an individual, that is, an associative response was viewed as a simple unit of thought. Therefore, it was widely believed that understanding these simpler units of thought could help understand the more complex processes of thinking (Cramer, 1968, p.6). Within the same tradition of WA research, Deese (1965) took this idea even a step further by arguing that by

studying individuals' associative responses, researchers could get insights not only into the processes of thinking but also into the way individuals construct their bigger sets of meaning. He also contended that it is the very nature of meaning as a relational concept that makes it possible to account for certain relationships between words and natural phenomena, as well as between words and other words.

It was this particular approach to the analysis of WA that attracted the attention of several L2 lexical researchers (e.g., Miron & Wolfe, 1964; Lambert, 1972; Meara, 1978; Kruse, Pankhurst, & Sharwood Smith, 1987), whose experimental work paralleled to a large extent the L1 association studies. These researchers broadened the scope of L1 research by adding questions concerning how L2 learners' associations compared to NS' associative patterns. Their main point of departure was the finding that adult NS tend to show stable patterns of associative organization of meaning along commonly shared paradigmatic-syntagmatic patterns. By contrast, L2 researchers consistently found that non-NS failed to develop such stability of their associations and in that they notably differed from NS' association patterning; hence from NS' shared meaning patterns. This key finding led to several distinct, yet related, lines of L2 lexical research that used association data to study different aspects of L2 learners' lexical knowledge. For example, some researchers (e.g., Ervin, 1968; Szalay, 1984; Yoshida, 1990) focused on examining the socio-cultural influences on lexical acquisition. Another group of researchers (e.g., Kruse, Pankhurst, & Sharwood Smith, 1987, Schmitt, 1998; Wolter 2002) centered on investigating the degree of shared lexical commonality between NS' and L2 learners. Yet, a third group of scholars employed WA tests to probe "depth" of L2 learners' lexical knowledge (e.g., Reed, 1993; Vermer, 2001; Wolter, 2001; Greidanus & Nienhuis, 2001). Each of these bodies of research tried to shed light on how the structure of the mental lexicon of non-NS differs from that of NS and what the implications of these differences might be for language teaching and assessment. However, little research has been done on the role of language proficiency in the associative patterning of L2 learners' lexical knowledge, i.e. whether the

association patterns L2 learners build primarily depend on the degree of knowledge they have of the SW, whether an increase in language proficiency has any impact on the associative connections, or whether these connections depend on factors unrelated to language proficiency. Similarly, more research needs to be done to explore in-depth the nature of the differences between the structure of the NS' and non-NS' mental lexicon, i.e. whether they are predominantly qualitative or quantitative in nature. I believe that the answers to these questions will bring L2 researchers closer to understanding what factors have impact on the way NS and non-NS organize their lexical knowledge and how language teaching can facilitate the development of better structured L2 mental lexicon.

This chapter will address some of the outlined issues above as follows. First, I will review the theoretical framework behind using WA tests as developed and used in L1 psychological and cognitive research. I will focus on some traditional distinctions L1 researchers made between qualitative and quantitative features of WA domains which, unfortunately, are frequently disregarded in L2 WA research. I will also discuss some of the major findings of L1 experimental work concerning several well-researched relationships between certain characteristics of the SW and the qualitative and quantitative features of NS' WA domain. Next, I will discuss the application of WA tests in L2 experimental work and the research lines that were born out of this tradition. I will briefly overview each group of L2 WA studies and some of their most significant findings in order to outline the multi-faceted application of WA tests in L2 context and their value as a research tool. Finally, I will present the results of a qualitative and quantitative WA data collected for this study. The discussion will aim at shedding light on issues related to the organization of the L2 mental lexicon that have not been sufficiently well researched in the field.

L1 word association research

Theoretical background

Association theory has a long tradition in L1 psychological and cognitive research. The foundation of linguistic association testing was laid on the perceived connection between language and thought which challenged many researchers to study verbal associations in an attempt to shed light on the nature of human thought. They firmly believed that associations were cast in language and language, on the other hand, was an embodiment of human thought (Deese, 1965, p. 4). Traditional primary laws of association, i.e. the laws that tried to describe the necessary conditions for the formation of an association, were derived on stimulus-response principles. In fact, they held that the way the mind worked was a function of external to the mind events. In addition to the primary laws of association, there was a set of secondary laws as well. The secondary laws aimed at describing the conditions that influenced and modified the resulting associative responses and many psychologists followed that line of exploration. These laws were taken over from the British school during the 18th and the 19th centuries and were thought to be the basic principles on which the mind functioned. In other words, the mind was believed to function on the association of ideas that operated on the principles of contiguity, similarity or contrast, and frequency of occurrence (e.g., Deese, 1965; Cramer, 1968; Kruse, Pankhurst, & Sharwood Smith, 1987).

Interestingly, while early psychological experiments assumed that responses were essentially conditioned by an individual's personal background and experience, it was also noticed that there was a high degree of commonality of the primary and secondary responses despite individual differences in experience and background. This finding gave rise to a large number of studies that began analyzing quantitatively the degree of commonality of associative responses among various groups of participants. The studies initially focused on the analysis of the primary response (i.e. the response that occurred with the greatest frequency to SW) as a measure that provided an estimate of the strength of the association response commonality

among subjects. The idea of *response commonality* was originally used in reference to the occurrence of the primary response to a SW, but later it began to be used to refer to the frequency of occurrence of any three most commonly given associations to a SW, often determined in terms of their absolute frequency of occurrence in a set of WA data (Cramer, 1968).

Deese (1965) is one of the researchers with, probably, the greatest contribution to the WA line of research in L1. In his work, he strongly emphasized the importance of studying the semantic organization of meaning shared by groups of people by arguing that if people did not share common meaning, they would not be able to communicate. Said differently, he argued that common meaning in communication is largely determined by the existence of commonality of associative structures shared by different people (1965, p. 45). In addition to the commonly shared meanings, he underscored that one of the unique features of human language relates to the highly organized relationships between words in language. Therefore, he held that since the structure of associative relations stemmed primarily from language use, the best way to study relations among meanings of words was to adopt a semantic approach to the analysis of WA data. In the main, such understanding of semantic relationships among words in the mental lexicon is very much in concert with recent semantic field theory. Semantic field theory advances the idea that the meanings of words must be understood in relation to other words that shape a given semantic domain (Kittay & Lahrer, 1992). Thus, to understand the meaning of the noun *scent*, for example, one has to understand first its contrastive relation to other nouns such as *odor*, *aroma*, *stink*; its affinitive relation to the term *smell,* as well as its syntagmatic relations to words such as *unpleasant*, *delightful*, *give off*, *detect*, etc. Overall, researchers from the semantic field tradition agree that the relations which order a field are of two types: *paradigmatic*, words that are substitutable for one another in a well-formed syntactic string (e.g., synonyms, antonyms, meronyms, hyponyms, etc.), and *syntagmatic*, words that collocate well in

a grammatical string and have semantic affinities (e.g., something *gives off a bad smell* but cannot *emanate* or *excrete* it) (ibid.).

However, many WA researchers (e.g., Ervin, 1961, Deese, 1965, Stolz & Tiffany, 1972) take a broader view on the nature of associative relations between words in the minds of language users by adopting a distinction between paradigmatic and syntagmatic relations based on lexical class rather than on semantic relations. According to this broader categorization of responses, paradigmatic responses are the ones that belong to the same lexical category as the SW and syntagmatic responses are those that belong to a different lexical class than the stimulus. This approach is broader than the semantic field view in that it supports the linguistic assumption that members of a given lexical category can replace one another in a wide variety of sentences, i.e. they can potentially occupy equivalent positions within utterances. By contrast, stimulus-response members of different lexical categories usually occupy different positions within phrases and sentences and, more often than not, tend to be contiguous. In this regard, as Deese (1965) pointed out, a positional-equivalence notion of a lexical class is "a psychologically useful one" because it reflects certain important features in concept attainment (p. 100). In the main, the broad distinction between paradigmatic and syntagmatic associations, though not unanimously agreed upon, is in concert with current understanding of the relationships between different levels of language description. At the same time, it further emphasizes the difficulty of drawing a clear dividing line between syntax and meaning, despite their relative independence. The reason being that "both are structured within the language and both have extralinguistic referents… The only possible difference between grammar and meaning is that meaning is more influenced in its structure by extralinguistic relations (relations in the natural world), but grammar is, to a lesser extent, so structured (by, for example, social relations), so the distinction is one of degree rather than an absolute one" (Deese, 1965, pp. 100-101).

Description of the WA domain

The organization of the WA domain has been traditionally described in L1 research by means of quantitative and qualitative measures. Quantitative measures, such as the strength of the primary response, response commonality, response heterogeneity, response idiosyncrasy, availability of responses, number of responses, etc., have mostly been used as indicators of the quantitative characteristics of the organization of the associative domain. The qualitative measures, such as, the form classification of the responses (paradigmatic and syntagmatic), the semantic classification (e.g., synonyms, antonyms, meronyms, etc.) of the associations, their dictionary meanings, etc. have been applied to describing the qualitative characteristics of WA responses. As far as the factors that influence the quality and the quantity of any WA domain are concerned, it is logical to assume that characteristics of the SW as well as characteristics of the participants would have impact on the quality and the quantity of their associations. However, Deering (1963 cited in Cramer, 1968) established that it is the SW variables rather than the participant variables that determine the critical features of a WA field. Following this conclusion, the investigation of WA domains has been largely narrowed down to analyzing the relationships between SW and associative responses they evoke, rather than to examining the impact of participants' characteristics on the associative connections, especially if the later were not of a particular interest to a researcher. Several characteristics of the SW have been very well examined in L1 WA studies, including the affectivity of the stimulus, the effects of the semantic levels of the SW on the association responses, the impact of stimulus familiarity and form class on the responses, etc. The findings concerning two of those characteristics, namely stimulus familiarity and form class of the SW, are of particular relevance to the WA analysis that will be presented later in this chapter. Regrettably, these effects have been largely overlooked in L2 WA studies and little attention has been paid to how they affect L2 learners' mental lexicon organization. That is why, in the following paragraphs, I will briefly overview the relationships,

which were experimentally determined in L1 research that are of particular relevance to the proposed analyses in this and the following chapters.

Effects of SW frequency on the WA domain

L1 research has a long tradition in exploring the effects of familiarity with the SW on WA responses. By and large, familiarity has most often been determined by the frequency of occurrence of a SW, based on some source of a frequency count. The general notion behind this assumption has always been that person's knowledge of a word increases relative to the rate at which a word occurs in a given language (Stolz & Tiffany, 1972). Consequently, many L1 researchers, as well as L2 researchers, followed the practice of selecting their SW either from *The Teacher's Word book of 30, 000 words* (Thorndike & Lorge, 1944) or other frequency count lists – a practice which can be challenged on several points. To begin with, language is a dynamic phenomenon and the frequency of occurrence of the words comprising its lexicon change synchronically as well as diachronically. Unfortunately, though extremely valuable resources and precious pieces of scholarly work, most of the existing word frequency counts do not reflect those changes and, consequently, their relevance to the current dynamics of English becomes arguable. Along the same lines, while the relationship between knowledge of words and the frequency of occurrence of these words is fairly well-established across large groups of people, there are several problems linked to this generalization. The most serious of them concerns the dubious value of an assumption which takes that any frequency count can be representative of the relative frequency of words in any individual's or group's experience (Stolz & Tiffany, 1972, p. 40). It is also worth noting that besides word frequency, 'emotionality' (also *stimulus affectivity*) as a factor has been found to have far greater effects on response patterns than initially suspected (Cramer, 1968). For example, it was found that emotional stimuli elicited much more idiosyncratic responses than neutral stimuli did (e.g., Deering, 1963). This, in turn, drew researchers' attention to the fact that adding a second dimension to a SW, such as affectivity, makes it very difficult to distinguish the effects of emotionality from the effects of word

frequency (as a measure of familiarity) on the WA patterns. Therefore, in an attempt to remedy some of those problems, several L1 researchers began using in their studies the 4-level scale of word familiarity, proposed by Dale (1965), who distinguished between the following 4 stages of knowing a word:

Stage 1: "I never saw it before".

Stage 2: "I've heard of it, but I don't know what it means."

Stage 3: "I recognize it in context – it has something to do with …."

Stage 4: "I know it." (Dale, 1965, p. 398).

The scale is often used as a pre-check of self-reported word knowledge, especially in studies aiming at investigating participants' judgements involving words that they claim to have partial or no knowledge of (e.g., Shore & Durso, 1990). The scale or its modified by Canadian researchers Paribakht and Wesche (1993) version is widely used in L2 lexical studies. In summary, it would be safe to say that familiarity with the SW has been undoubtedly recognized as a factor that influences associative response patterns. Therefore, conceptualization of SW familiarity moved from assumptions of familiarity based on the frequency of occurrence of the SW to the use of participants' ratings of familiarity, which distinguish several degrees of knowing a word. In recent studies, L1 researchers seem to agree on the inclusion of at least three such levels of word familiarity: a level of unknown words, a level of frontier words (vaguely known), and a level of known words (Shore & Durso, 1991), which has opened new interests in experimental work related to exploring the relationship between self-reported levels of familiarity and actual knowledge of lexical items.

Effects of SW familiarity on the WA domain

The effects of SW familiarity on the patterns of NS' associative responses is another well-researched area in L1 context. Of all studied effects of SW familiarity, number of associative responses, commonality/ heterogeneity of responses, and type of associative responses are of particular interest to the author. Overall, interest in each of these descriptors

yielded groups of studies that specifically focused on how they interact with familiarity. Number of responses, as a variable that describes the quantitative characteristics of a WA domain, produced the most controversial results. For example, after reviewing a number of studies that examined the relationship between number of responses and SW familiarity, Cramer (1968) concluded that this line of research did not yield consistent results across a large body of experimental work. In support of her conclusion, the author reported studies which found that high-frequency words elicit more responses than low-frequency words (e.g., Noble, 1953; Underwood, 1959), as well as studies that found no relationship to exist between familiarity and number of responses (e.g., Davids, 1956; Wispé, 1954). Cramer argued that availability of responses could not be sufficiently well explained by familiarity of the stimulus alone because adding another dimension, such as emotionality or abstractness of the SW, interacted with familiarity in a way that largely masked its effect. The author concluded that both familiarity and emotionality contribute to the increased number of responses for high-frequency words, since high-frequency stimuli predominantly include neutral words and low-frequency words tend to be very often emotional words (Cramer, 1968, p. 58).

Response strength, also referred to as *commonality of responses*, is another well-described effect of SW familiarity on the quantitative characteristics of the WA domain. As mentioned earlier, L1 WA researchers noticed that there was a high degree of commonality of the primary and secondary responses of NS despite individual differences in background and experience among subjects, which encouraged a lot of researchers to analyze quantitatively the degree of commonality of associative responses among various groups of individuals. The idea of *response commonality* was originally used in reference to the occurrence of the primary response to a SW. But more recently it has started to be used to refer to the frequency of occurrence of any three most commonly given associations to a prompt word. In other words, the response that occurred with the greatest frequency to any SW, the primary response, or the three most common associations served as a measure of the strength of the association

response commonality among participants. By and large, research findings suggest that when commonality is defined by the frequency of the first three most common responses, high-frequency stimuli elicit the highest commonality score, medium frequency SW elicit the next highest, and low-frequency words elicit the lowest commonality score. However, several WA researchers noticed that the combined effect of increased commonality and frequency of the primary response was also positively related to, yet, another variable - response variability (also *number of different responses*), i.e. the greater the familiarity with the SW, the more difficult it could be to maintain a common response set. For some time, this relationship was treated as an apparent paradox in the L1 literature because it was traditionally accepted that response strength and response variability were inversely related, i.e. the greater the commonality of responses, the smaller the number of different responses. Surprisingly, subsequent research found that this relationship largely depended on the model used to elicit associations. For example, a model that asked participants to produce more than one association per stimulus allowed for a response competition, which, in turn, increased the chances of strong responses to be better represented than weak ones in a larger set of associations. On the contrary, a model based on one response per SW would be more likely to yield a greater variety of responses and a lower degree of commonality. This finding raised researchers' awareness that the model of response elicitation could greatly influence the quantitative characteristics of the WA domain.

Finally, a small number of studies examined the type of associative responses as a function of participants' familiarity with the SW. Overall, it has been found that the effects of word frequency, as a determinant of familiarity, on the number of syntagmatic and paradigmatic responses may vary according to the lexical category of the SW. While familiarity didn't seem to have any effect on the types of responses given to nouns and verbs, for adjectives, the low-frequency words elicited more syntagmatic responses and the high-frequency stimuli yielded more paradigmatic ones (Deese, 1965). Another, even more interesting, relationship was found

between part of speech of the SW and type of WA. On the whole, it was noticed that certain grammatical categories elicit a higher proportion of paradigmatic responses than other categories do. For example, Deese' experimental work (1962, 1965) showed that nouns elicit the greatest number of paradigmatic responses, followed by high-frequency adjectives, and verbs. On the other hand, he found that syntagmatic responses were most often elicited by adverbs, followed by adjectives, verbs, and nouns (ibid.). By the same token, type of responses was consistently found to vary as a function of the age of the participants. When compared to the responses obtained from adults, typical responses obtained from children showed changes in rather regular ways with increasing age. When SW variables were held constant, for example, Ervin (1961) observed that young children gave mostly syntagmatic responses, whereas older children and adults gave more associations that were from same lexical category with the SW. Earlier research interpreted this shift from predominantly syntagmatic to predominantly paradigmatic associative connections as a developmental phenomenon related to maturation and sophistication of cognitive processes. However, subsequent research (e.g., McNiel, 1966, 1970) suggested that shift in association patterns with increasing age might be explained either by the acquisition of new features of the stimuli, which would result in participant's choosing a response that shared the maximum number of features with the stimulus, or by change in the strategies used in searching and matching stimulus-responses features. Stolz and Tiffany (1972) confirmed the first hypothesis, arguing that cognitive skills, as well as knowledge of words and the acquisition of more features and meanings develop dramatically with young children. Therefore, isolating the effects of one of the factors and ignoring the effects of the other would inevitably result in an incomplete account of the shift in WA response pattern. The researchers further hypothesized that young children respond with overwhelmingly syntagmatic associations not because their cognitive processes are immature but because their familiarity with the SW is relatively low or non-existent. Their experiment supported this hypothesis and revealed that while the responses to unfamiliar adjectives of adults patterned very much like the

responses given by young children to common adjectives, their responses to familiar adjectives were similar in patterns with the ones provided by older children. Therefore, the authors concluded that the primary cause of the response shift was greater familiarity with word features rather than cognitive development alone. They also suggested that the methodological implication of this finding should be seen in the application of WA tests as "a rather sensitive index of the state of one's lexical knowledge about a given word" (Stolz & Tiffany, 1972, p. 45).

In a nutshell, word frequency as a measure of familiarity with the SW has been found to have several distinctive effects on WA patterns. Some of the most important findings of L1 WA research can be summarized as follows: 1) an increase in the familiarity with the SW results in an increase in the number of associative responses, especially when increased response availability is examined in relation to the emotionality of the SW. 2) Increased SW familiarity is positively related to response commonality as defined by the frequency of the primary response or the first three most common responses. 3) Certain grammatical categories elicit a higher proportion of paradigmatic responses than other categories do. For example, nouns elicit the greatest number of paradigmatic responses, followed by high-frequency adjectives, and verbs, while syntagmatic responses are most often elicited by adverbs, followed by adjectives, verbs, and nouns. 4) SW familiarity combines its effects on type of associative responses (syntagmatic or paradigmatic) with the form class of the stimulus. This relationship is particularly salient with adjectives, where the higher the frequency of the stimulus, the greater the number of paradigmatic associations. It also deserves to be mentioned that the theoretical background, the research lines, as well as the findings of L1 WA studies served as a foundation for the development of L2 WA studies in lexical research. L2 WA studies adopted not only the methods of analysis of WA data but also confirmed most of the findings concerning L1. More importantly, they added another dimension to the study of verbal associative behavior by comparing L2 associative connections with the features of NS' associative networks. The following section will specifically discuss the contribution of L2 lexical studies to the WA area of research.

Word association tests in L2 research

Research on the associative behavior of L2 learners has been a subject of interest in FL and L2 studies for a long time. As a matter of fact, it closely paralleled L1 WA research both in time and approaches to the analysis of association data. However, the research questions in L2 WA studies are very different from the ones asked by L1 researchers. While a L1 WA study aims at examining certain aspects of the organization of lexical knowledge of NS of a language, L2 researchers use WA tests to estimate degrees of *native-likeness* of L2 learners' lexical organization by examining how their associative patterns compare with the ones of NS. One of the basic assumptions underpinning this approach is that NS have remarkably stable WA patterns which reflect the complex lexical and semantic links in their mental lexicon (Read, 1993). On the other hand, L2 learners tend to have much more diverse and unstable patterns of responses, though, there is evidence that as learners' proficiency increases, some aspects of their associative connections develop towards NS' norms (e.g., Meara, 1978; Vermeer et al., 2001). In this respect, as Kruse et al. (1987) point out, WA tests are often used as instruments that hold a potential to offer a measure of degrees of native-likeness; hence, they are often used by "L2 researchers who are interested in the developing lexicon as a measure for establishing baselines for various experimental investigations" (p. 141). Consequently, we can distinguish several relatively distinct lines of L2 WA research, which differ mainly in the perspective from which L2 association patterns are discussed. That is, WA are interpreted 1) from a socio-cultural perspective; 2) from the point of view of language proficiency and its effects on associative behavior; 3) as an indicator of "depth", or quality, of vocabulary knowledge; and 4) as an indicator of the organization of the L2 mental lexicon. I will briefly review each of these lines of research in order to emphasize the multi-faceted role that WA tests play in L2 research. Also, by reviewing the L2 contribution to WA experimental work, I will outline some of the most important findings derived from studying the WA behavior of L2 learners, the way these findings compare

with L1 research, and how they complement discussions about the organization of the mental lexicon of speakers of more than one language.

Socio-cultural tradition in L2 WA research

The socio-cultural line of research examines the notion of native-likeness from the point of view of the effects of social and cultural factors on the acquisition of L2 lexical concepts. In this approach, the associations are viewed as related to non-language variables, rather than to language ones, such as culture-specific meaning, cognitive structures, or shared cultural experiences (Szalay & Windle, 1968, p. 43). Some of the earliest WA studies among adult bilinguals hypothesized that experiences and memories of various kinds are linguistically stored independently in the different languages an individual speaks (e.g., Kolers, 1963; Ervin, 1968). However, more recent WA studies have reached a different conclusion from that of Kolers (1963) and Ervin (1968). For example, Yoshida (1990) has noticed that her participants, Japanese college students who have lived at least for 2 years in the US prior to entering a Japanese university, tend to produce language-dependent associative responses to SW from her culture category but not so much to culturally-neutral stimuli. In her analysis, if a concept were unique to the L2 culture, the culturally-bound responses given by the bilinguals but absent from the response domain of the Japanese control group served as evidence that these concepts tend to generate culturally-based learning. Other researchers from the cultural paradigm (e.g., Szalay, 1984; Szalay & Brent, 1967; Szalay & Windle, 1968) took the idea of cultural influence even further by arguing that the acquisition of cultural concepts is accompanied by cognitive restructuring of the conceptual system. Szalay (1984) argued that the analysis of the associative responses provided by different cultural groups allowed for getting insights into the cultural beliefs shared by the members of these groups. He further contended that “only by comparing the perceptual/ semantic representation developed by different culture groups do we discover how these representations vary from culture to culture, how they depend on our background, on our culture, and how people’s behavior and their relationship to each

other depend on these shared subjective representations." (Szalay, 1984, p. 74). Szalay et al. earlier research (1968, 1972), for example, uncovered differences in the response patterns of Korean participants traceable to the influence of the culture in which the L2 was acquired and not so much to the language of response itself. Thus, Korean participants who acquired English in an American setting and experienced the culture of the L2 in the process of acquisition produced responses which were very similar to the NS' associations. By contrast, there was a very high incidence of intergroup association patterns for the Korean learners who didn't have a cultural experience, regardless of whether the participants responded in their L1 or in English. This was interpreted to indicate that L2 learners' conceptual system was not very likely to reorganize without extensive immersion experience in the L2 culture, compared to the conceptual system of participants who had such an exposure. Grabois' research (1997 cited in Lantolf, 1999) supported these findings and, to a great extent, built a bridge between the effects of culture on L2 lexical acquisition and the effects of L2 proficiency by studying comparatively the concept formation among learners that acquired a L2 naturally and those who acquired it through classroom instruction. As Lantolf (1999) points out, one of the important conclusions of Grabois' research is that while the intermediate level students didn't provide evidence for any reorganization of their native conceptual system, the advanced learners who had an immersion experience in the L2 culture for approximately 10 years did show strong evidence of such reorganization (p. 41). Overall, the socio-cultural perspective of L2 lexical acquisition has developed as a research area in its own rights. Similarly, it has greatly contributed to our understanding of the complexities involved in the attainment of lexical concepts by L2 learners. In the main, the conclusions drawn by the researchers working in this paradigm strongly emphasize the subtle interaction and the inextricable relationship between language and culture, which readily projects in the acquisition of a new language.

Assessing language proficiency by means of WA tests

Another line of L2 WA research, paralleling the socio-cultural paradigm, attempted to examine the relationship between associative patterns and language proficiency. By and large, L2 researchers who employed WA test in their studies aimed at ascertaining whether WA tests can be used as an instrument for measuring language proficiency or whether they are not suitable for that purpose (Kruse et al., 1987). Some of the earliest research in this direction (e.g., Lambert, 1972; Davis & Wertheimer, 1967; Kolers, 1963; Riegel, Ramsey, & Riegel, 1967) revealed that the number of responses given to SW increased with an increase in language proficiency; hence, it confirmed a clear relation between the number of associative responses and proficiency level. However, the degree of commonality between responses was found to be largely unaffected by changes in the number of responses and remained relatively independent of the changes in the number of associations. In addition, some researchers observed that the lower proficiency groups tended to translate the stimuli (e.g., Davis & Wertheimer, 1967) or to give more often directly translated responses to concrete than to abstract stimuli (e.g., Kolers, 1963; Meara, 1978). However, one of the problems with conclusions relating translations to proficiency lies in the difficulty to directly attribute translation of responses to participants' level of proficiency. The reason being that this effect may well be a consequence of cross-cultural similarities or cultural heritage rather than just translation as a L2 strategy. Overall, studies that attempted to determine whether WA tests can be used as sensitive predictors of language proficiency are far from being in agreement about their potential. Some researchers (e.g., Kruse et al., 1987) have concluded that WA tests do not show much promise as a means of assessing proficiency, whereas others (e.g., Piper & Leicester, 1980) have argued that WA behavior is sensitive to language proficiency. The contradictory conclusions of these studies are largely due to certain methodological weaknesses such as methods of assessment of general proficiency, methods of SW selection, scoring procedures, etc. Nonetheless, Wolter (2002) has suggested on an optimistic note that refinement of scoring procedures in addition to careful consideration

of proficiency assessment and SW selection may yield testing procedures valid for distinguishing more proficient from less proficient L2 learners.

WA tests and depth of vocabulary knowledge

Most recently, several attempts have been made to examine quality, or depth, of word knowledge by employing association tests as a means of tapping L2 learners' meaning representations. By and large, researchers from this paradigm seem to share a common belief that WA tests are a valuable tool for assessment of the structure of lexical knowledge. However, they vastly differ in their interpretation of association patterns as indicators of quality of lexical knowledge. Some researchers (e.g., Kruse et al., 1987; Schmitt, 1998; Wolter, 2002) see the value of WA test as a useful instrument to measure L2 learners' ability to produce nativelike associations. Schmitt (1998), for example, argued that the attribute on which quality of L2 responses should be judged is the degree of "native-likeness" of L2 associations. In his analysis, the most common response pattern provided by his norming group was three (54%) and two (34%) common responses, which made him argue that this high degree of commonality among NS' responses (88%) is an indicator of native-likeness of association behavior against which non-NS' associations should be matched. Hence, a higher degree of native-likeness of L2 associations will be indicative of a higher quality of lexical knowledge. Another group of researchers seem to favor Meara's proposal (1978, 1984) of looking at patterns of association responses as characterizing the structural organization of lexical knowledge of language users. Overall, Meara (1996) finds the term *depth* of vocabulary knowledge very confusing in the context of assessment of L2 lexical knowledge. Instead, he proposes to call this dimension simply "quality" of vocabulary knowledge, meaning quality of the associative links that L2 learners develop as revealing of the organization of their mental lexicon. He also hints at the possibility that when L2 learners develop sufficiently large in size vocabularies, it is the structural organization of that knowledge that becomes of primary importance. In his view, it is essential to know whether L2 learners' lexical knowledge is organized in semantic connections

similar to those of NS or whether they simply tag L2 words into their L1 lexicon. Taking such an approach to analyzing association patterns gives the advantage of determining whether non-NS' WA patterns are broadly comparable or not with the ones of NS. It also allows for finding out whether there are structural overlaps between NS' and non-NS' association domains as proficiency increases or whether non-NS' lexicon develops in a fundamentally different from NS' lexicon way, independent of language proficiency

Several researchers (e.g., Meara, 1978; Piper & Leicester, 1980; Read, 1993; Wolter, 2001; Vermeer, 2001; Greidanus & Nienhuis, 2001) have adopted the aforementioned point of view and focused on examining patterns of paradigmatic, syntagmatic, and phonological connections that L2 learners and NS build in organizing their lexical knowledge. Their analyses are based on the assumption that the paradigmatic, syntagmatic, and phonological patterns of associations are revealing of different types of semantic relationships between words in the mental representations of language users (Clark, 1970 for a review). It should be mentioned that most of the traditional assumptions about the nature of the semantic organization of L2 learners' lexicon derive from the early research conducted by Meara (1978, 1984). Based on his experimental work, Meara arrived to the general conclusion that there are significant differences in the way NS and L2 learners structure their semantic relationships. He identified the following major differences between NS' and non-NS' lexical organization: 1) NS of English tend to produce a large number of common responses, which are semantically strongly linked. In contrast, L2 learners produce a greater number of varied and unpredictable responses that are for the most part semantically unrelated. 2) A large number of L2 learners' association responses seem to be a result of phonological or orthographic misinterpretations of the SW rather than meaning-based misinterpretations. 3) NS typically have very stable response patterns over time, whereas L2 learners seem to be very unstable in their response production, which makes it difficult to use WA as a sensitive index of developmental progress. Finally, 4) adult NS tend to organize their lexical knowledge primarily along paradigmatic lines, whereas L2

learners' lexical organization most often falls along syntagmatic and phonological lines. The last conclusion, at least on the surface, echoes findings of L1 WA research that suggest a developmental shift from predominantly syntagmatic responses for children to overwhelmingly paradigmatic patterns for adults that is indicative of cognitive maturation.

In sum, it can be said that the results of past studies point to the broad generalization that L2 learners' patterns of lexical item connection are fundamentally different from the ones of NS. Though, it should be noted here that little attention has been paid to identifying the nature of these differences, i.e. whether these differences are quantitative, qualitative, or both. More specifically, no attempt has been made to compare quantitatively, as well as qualitatively the two types of lexicons, especially with regard to the effects of increased language proficiency. To clarify some of these issues, I will analyze WA data I collected from NS and non-NS of English with regard to the quantitative and qualitative characteristics of their WA domains. The analyses will try to answer several important questions concerning the organization of the mental lexicon of L2 learners, which have not been satisfactorily answered in L2 lexical research yet. For example, where should the differences between NS' and L2 learners' mental lexicon organization be looked for? How does an increase in language proficiency affect the semantic organization of L2 lexical knowledge? How do the quantitative and the qualitative features of lexical knowledge interact? The distinction between qualitative and quantitative features of WA domain adopted in the analysis is the one used by L1 researchers, discussed earlier in the chapter. In an attempt to shed light on these issues, I will first compare the quantitative characteristics of the organization of the WA of three groups of participants: advanced language learners, intermediate language learners, and NS. The quantitative features will be analyzed by measures of (i) response strength; (ii) size of the associative-response domain, and (iii) measures reflecting the consistency of the associative-response domain. Next, I will examine the WA data qualitatively with regard to the (i) paradigmatic, (ii) syntagmatic, and (iii) morphological/ phonological WA patterns given to words with which participants report a varying

degree of familiarity. The analysis will focus on WA given to SW participants report to know and SW they vaguely know (i.e. frontier words). The analysis will aim at finding how familiarity, lexical category, and frequency of occurrence of words affect the quantitative and qualitative features of language users' mental lexicon, when proficiency is taken into consideration. Finally, by broadly comparing the qualitative and quantitative characteristics of the WA domains of the three groups of participants, I will try to specify the factors responsible for the differences in the WA patterning among different proficiency groups.

This study will test the following null hypotheses:

1. There are no quantitative difference between the mental lexicon of NS, advanced, and intermediate learners of English as measured by (i) the strength of the three most popular responses, (ii) response commonality, (iii) total number of responses, (iv) number of different responses, and (v) response idiosyncrasy.
2. There are no qualitative difference between the mental lexicon of NS, advanced, and intermediate learners of English as measured by (i) the paradigmatic, (ii) syntagmatic, and (iii) morphological/ phonological WA responses given to known and frontier SW.

Quantitative analysis of the WA domains

Method

A three-group (NS, L2 advanced, and intermediate learners) between-participant design was used. The dependent variables were five different quantitative measures of lexicon organization, i.e. the three most popular responses, response commonality, total number of responses, number of different responses, and response idiosyncrasy. Five hypotheses of the type H_o: $\mu_{NS} = \mu_{advanced} = \mu_{intermediate}$ were tested.

Participants

Sixty four adults, both NS and non-NS speakers of English from the same pool of subjects participated in this study, i.e. NS (*n* = 30), advanced (*n* = 17) and intermediate learners of English (*n* = 17).

Materials

The WA data were collected in written form. First, all participants were asked to identify on a 4-level scale how well they knew each of the 73 words, randomly selected from a dictionary containing 16,045 words. Provided the participants chose either option III (*I think the words means* _____) or IV (*I know that the word means* _____) and verified their familiarity by supplying a synonym, explanation, or translation of the SW, they were asked to write down up to three words with which they associate the SW. This procedure of WA data collection renders the WA test a *continuous associations test* because the stimulus was presented only once and served as a point of departure for the production of a relatively short chain of associations. I decided to ask the participants to provide up to three responses, instead of just one or as many as they could think of, in order to obtain a reasonable amount of associations that would form a sufficiently large association domain for each SW.

The SW were listed alphabetically in the test in the order in which they were selected from the dictionary. All SW were content words from different syntactic categories. For most of the words the intended syntactic category was specified. The following categories were represented in the sample: nouns (*n* = 42), verbs (*n* = 16), adjectives (*n* = 13), and adverbs (*n* = 2). By using SW from several grammatical classes, selected by means of a systematic sampling with a random start, I attempted to collect association data that would reflect the organization of the whole mental lexicon, which ultimately contains lexical items from all form classes.

To avoid bias in the association data, which very often derives from over-representation of high frequency words in test samples, special attention was paid to the frequency of

occurrence of the stimuli. The frequency of occurrence of each SW in *The Educator's Word Frequency Guide* (Zeno et al, 1995). The *Guide* itself contains current frequency count information, which is based on over 14 million words taken from a wide range of materials that students in the United States are likely to encounter throughout their education years. The frequency of each SW was identified by its Standard Frequency Index (SFI) and the analysis of the resultant sample of prompt words showed a relatively equal representation of stimuli from several frequency bands. There were 13 items (18 %) with SFI between 20.0 and 30.0, 23 items (32 %) with SFI ranging from 30.0 and 40.0, 19 items (26 %) with SFI between 40.0 and 50.0, and 18 items (24 %) with SFI between 50.0 and 70.0+. Because there was a fairly even distribution of several frequency bands in the sample, and for the sake of simplicity, I divided the SW into three frequency categories: low, mid, and high frequency words. The category of low frequency stimuli (n = 24) included words in the range between SFI = 20.0 to approximately SFI = 35.0; the words classified in the mid frequency band (n = 25) were in the range from SFI = 36.0 to SFI = 45.0; and the high frequency category (n = 24) consisted of words spreading over SFI = 46.0 to SFI = 70+ (see Appendix I).

Scoring procedures

In the tradition of L1 WA research, the quantitative characteristics of WA are essentially associated with measures indicating the degree of organization of the response domain (Cramer, 1968). To investigate those characteristics, I used several quantitative measures frequently employed by L1 WA researchers. These were:

1) Measures related to the *associative-response strength* reflected by:
    - *response commonality* - determined by scoring each response in terms of the absolute frequency of its occurrence among the responses of a group;

- *strength of the three most popular responses to a SW* - determined by the frequency of occurrence or nonoccurrence of any of the three most common responses to a stimulus.

2) Measures related to the *size of the associative-response domain* reflected by:
   - *total number of responses* - determined by the total number of responses elicited for each SW;
   - *number of different responses*, i.e. the number of different responses given to every SW obtained across the members of each group.
3) Measures related to the *consistency of the associative-response domain* as reflected by:
   - *response idiosyncrasy* - determined by the number of responses to a SW given by only one participant in a group.

Scores on each of these measures were obtained for every participant.

The first step in quantifying the WA was to compile a list of the associations for each group. For an association to be considered in the analysis, a participant should have demonstrated a certain degree of familiarity with the SW by providing synonym, a brief explanation, or a translation of the SW. All responses were lemmatized and the following items were listed as separate associations:

1) Base words and some of their inflections were combined in one item, e.g., inflection *–s* for nouns (e.g., *train* and *trains*); inflection *-s* for verbs (e.g., *forget* and *forgets*); inflections *–er* and *–est* for adjectives (e.g., *tall – taller - tallest*) and adverbs (e.g., *fast – faster - fastest*).
2) Irregularly inflected forms (e.g., *children, better, went*, etc.) were treated as separate from the base forms (e.g., *child*, *good*, *go*, etc.) associations.

3) Multiword responses (e.g., *good opportunities, marching line, give up, spread out,* etc.) were listed and scored as one item.
4) All derivations were treated as independent items (e.g., *cleaning, cleaned, breathless, immoral, resignation, disagreement, scrutinize, unfamiliar*, etc.).

Next, dependent on the frequency of occurrence of each association in the WA data of a group, the associations in each of the lists were classified as idiosyncratic (given by only one participant in a group) or common (given by, at least, two participants). The absolute frequency of occurrence of each common association was also tallied. For example, for the SW *inception*, the list of WA for the NS' group contained several common responses (e.g., *beginning* (3), *birth* (3), *start* (3), *first* (2), *initial* (2), *plan/s* (2)) and several idiosyncratic responses (e.g., *beginning of ideas*, *creation*, *first time*, *new*). Two other such lists were constructed with the associations supplied by the advanced group and the intermediate learners.

After the lists were compiled, each participant's association was scored according to the five specified criteria. For example, one of the native speaking participants responded to the SW *inception* with the following three associations: *initial*, *birth*, *new*. Therefore, based on the list of NS' associations, his/ her responses were scored as follows:

1) *Response commonality* - a score of 6 was assigned to the associations based on their absolute frequency in the NS' list (2 (*initial*)+3 (*birth*)+1 (*new*) = 6).
2) *Strength of the three most popular responses* - the participant gave only one of the three most popular responses (i.e. *birth*), so a score of 3 was assigned for the response.
3) *Total number of responses* - the participant gave the maximum number of associations (i.e. 3 responses), which carried a score of 3.
4) *Number of different responses* - because the participant was the first in his/ her group to supply these three associations, a maximum score of 3 was assigned.

5) *Response idiosyncrasy* - the participant gave only one idiosyncratic response (i.e. *new*) so, it received a sore of 1.

Results

Participant scores on each association measure were analyzed separately using a One-way Analysis of Variance (ANOVA) to compare the three proficiency groups (NS, advanced, and intermediate learners). I took the strategy of doing five independent ANOVAs rather than a single multivariate ANOVA because I had substantive interest in each of the dependent measures. The analyses yielded a significant group effect on all dependent variables (see Appendix J) indicating statistically significant differences among the group of NS, L2 advanced, and L2 intermediate learners for strength of the three most popular responses, $F(2, 61) = 6.425$, $p = .000$; response commonality, $F(2, 61) = 35.67$, $p = .000$; total number of responses, $F(2, 61) = 38.17$, $p = .000$; number of different responses, $F(2, 61) = 25.99$, $p = .000$; and response idiosyncrasy, $F(2, 61) = 26.35$, $p = .000$. Along with the statistical significance of the results, I was also concerned with the practical significance of the measured relationships, i.e. I was interested in the proportion of the total variability in each of the dependent variables that could be explained by the grouping variable (language proficiency). Measures of effect size (ω^2) were calculated to examine the strength of the relationship between group (the independent variable) and each of the dependent variables. The effect sizes indicated that group can explain only 14% of the variability in the three most popular responses ($\omega^2 = .14$), 52% of the variability in response commonality ($\omega^2 = .52$), 54% of the variability in the total number of responses ($\omega^2 = .54$), and 44% of the variation in the number of different responses ($\omega^2 = .44$) and response idiosyncrasy ($\omega^2 = .44$). According to Cohen's guidelines, with the exception of the effect size for the availability of the three most popular responses, which is considered moderate, the rest of the effect sizes are relatively large. This shows that the dependent variables provide practically

meaningful measures of the quantitative features of the organization of the mental lexicon of language users when language proficiency is used as a grouping variable.

The results of the analyses supported a conclusion that the quantitative characteristics of the NS' mental lexicon are different from the quantitative features of advanced and intermediate learners' lexicons. The key questions advanced earlier, however, could only be answered by exploring the nature of quantitative differences in the lexical organization among the three groups. To this end, post hoc pairwise comparisons were conducted to specifically compare the groups on the five quantitative measures. The post hoc comparisons, in which the significance level was adjusted to .05 based on Bonferroni rationale, revealed that the mean differences among the groups were statistically significant in all subsets of comparisons but one. The only non-significant mean difference was found between the NS and advanced learners in the strength of the three most popular responses (95% *CI* = -4.498, 1.691, *p* = .05).

The post hoc comparisons were revealing from where the quantitative differences among the three groups stemmed. As previously mentioned, to estimate the size and the heterogeneity of the WA domain of each group, I used two measures commonly applied to the quantitative description of WA responses in L1 research: total number of responses and number of different responses. On average, the advanced learners gave a significantly greater number of responses than NS (95% *CI* = 3.200, 7.357; *p* = .05) and intermediate learners did (95% *CI* = 51.935, 58,536; *p* = .05), as well as a greater number of different responses to the SW than the NS (95% *CI* = 18.137, 22.342; *p* = .05) and the intermediate learners did (95% *CI* = 39.213, 43,964; *p* = .05) (see Figure 3).

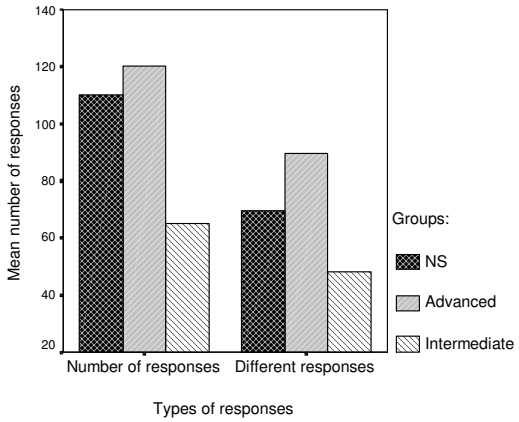


Figure 3. Mean distribution of participants' total number of responses and number of different responses generated to familiar words.

While it was not surprising to find that the intermediate learners had significantly smaller in size and less diverse WA repertoire than the NS and the advanced learners, it was surprising to find that the L2 advanced learners' associative domain can be significantly larger in size and more diverse than the NS' WA domain. Generally, it is logical to expect that the number of different responses and the total number of responses generated to a SW are directly linked to the overall size of vocabulary that an individual has. In examining the vocabulary size of the participants in this study (Chapter 2), I found that the intermediate learners had a significantly smaller vocabulary size (M = 6,033 words) than the other two groups, whereas the advanced learners had a non-significantly smaller in size vocabulary (M = 8,736 words) than the average vocabulary size of NS (M = 9,447 words). Obviously, the positive relationship between the number of words individuals know and the number of connections, as well as the diversity of connections they build among these words, holds true for mostly smaller vocabularies. The data indicate that L2 vocabularies between 5,000 and 6,000 words fall under this category and are characterized by a lower degree of connectivity and diversity in the connections among words than larger vocabularies. Larger vocabularies, containing 8,700 words or more, tend to be more connected and allow language users to associate every lexical item with a much greater number

of different things and ideas. Apparently, when it comes to vocabularies of over 8,700 words, factors other than proficiency influence the size and the heterogeneity of the WA connections, since the advanced learners performed better on both measures. In conclusion it can be said that vocabulary size of language users, both NS and non-NS, has a key effect on the connectivity of their mental lexicon. The larger the vocabulary, the more connected it should be expected to be, especially in terms of number and variety of connections.

The strength and the consistency of the associative responses of the three groups were examined by comparing groups' means of the absolute frequency with which a response occurred in each group's data, the strength of their three most popular responses, and the degree of idiosyncrasy of participants' associations. As a whole, the distribution patterns of each of these measures were noticeably different across the three groups (see Figure 4).

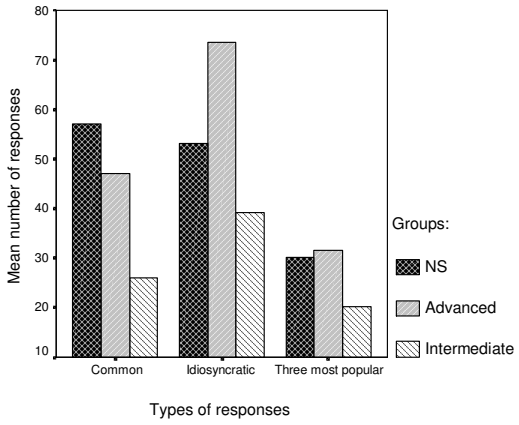


Figure 4. Mean distribution of NS', advanced, and intermediate learners' common, idiosyncratic, and the three most popular responses generated to familiar words.

The NS gave a significantly greater number of common responses, compared to the advanced group (95% *CI* = 8.472; 11.461; *p*=.05) or the intermediate group (95% *CI* = 29.414, 32.402; *p* = .05), and a smaller number of idiosyncratic responses than the advanced (95% *CI* =- 23.132, -17.359; *p* = .05) and intermediate learners (95% *CI* = 11.162, 16.936; *p* = .05). When I

calculated what percentage of the total number of responses their common and the idiosyncratic associations accounted for, I found that there was a slight difference in the proportion of common (52%) and idiosyncratic (48%) responses that they produced, which showed that for this group of NS giving common responses was as typical as giving idiosyncratic associations. As far as the frequency of occurrence of the three most popular responses is concerned, the data for this group revealed that the top three responses accounted for only 27% of all responses they generated. By and large, this response pattern for the NS' group agrees with the findings of other researchers (e.g., Rozenzweig, 1961; Yoshida, 1990) who suggested that American participants tend to show a high degree of commonality of responses, usually higher than other participant populations. For example, in Yoshida's study (1990), the American group top three responses given to the SW from her culture category accounted for 54% of all the responses, whereas in the Japanese control group they accounted for a much lower proportion (39%). The present findings do not quite support such a high degree of availability of the three most popular responses in the NS' data. Moreover, the three groups had very similar proportions of their top three responses: 27% for the NS, 26% for the advanced learners, and 31% for the intermediate learners. These results suggest that, when the degree of response commonality is measured by the availability (or the absence) of the three most popular responses, the effects of the SW on the response domain may obscure the effects of language proficiency on commonality of associative connections. As a matter of fact, the difference between the NS and the advanced learners on this variable proved to be the only statistically non-significant difference in the present analysis.

Interestingly, it was immediately noticeable in the mean distribution of the three types of responses (common, idiosyncratic, the three most popular responses) that the two groups of non-NS shared strikingly similar patterns of commonality and consistency of their WA domains in that, regardless of their differences in language proficiency, both groups gave more idiosyncratic than common responses to the SW. More specifically, the common responses of

the advanced and the intermediate learners accounted for 39% and 40%, respectively, of all the responses they generated, whereas the proportion of their idiosyncratic responses was much higher, but very similar in magnitude: 61% for the advanced learners and 60% for the intermediate learners. Apparently, the lower degree of commonality non-NS' WA responses, when compared to NS's commonality, is a distinctive pattern that characterizes the WA domain of L2 learners. More importantly, increase in language proficiency did not seem to have any effect on its overall pattern, which means that, at least for intermediate and advanced learners of English, the higher degree of idiosyncrasy of the responses to SW they know is an attribute of the L2 learner rather than a consequence of loose organization of their mental lexicon. I will further elaborate on this point in the discussion section.

Discussion

The examination of the quantitative characteristics of the mental lexicon of the language users participating in the study (NS, advanced, and intermediate learners of English) revealed that NS' and non-NS' mental lexicons are quantitatively different. The analysis of the size and the diversity of their WA domains showed that the intermediate learners had a WA repertoire much smaller in size and less diverse than the NS and advanced learners, whereas the advanced learners' associative domain was noticeably larger in size and more heterogeneous than the NS'. In light of the generally well-established positive relationship between vocabulary size and size of the WA domain, one possible conclusion to be drawn from these results is that L2 vocabularies of approximately 6,000 words, or intermediate learners' vocabularies, are characterized by relatively few and mostly loose connections among words. Larger vocabularies, containing 8,700 words or more, show greater connectivity among words and allow language users to associate every lexical item with a much greater number of different things and ideas. Hence, language users with larger vocabularies have a considerably richer connections, both in size and heterogeneity.

One of the most unexpected findings in the analysis concerns the size and the heterogeneity of the advanced learners' WA domain, i.e. the total number of associations and the number of different associations they generated were significantly greater than those of NS. This finding inevitably raises the question of the reason for the greater heterogeneity and the larger in size WA domain of the advanced learners in comparison to the NS'. Generally, this is not an easy question to answer because, on a larger scale, it pertains to the more global issue of multicompetence. More specifically, it touches upon the issue whether L2 processing can be cut from L1 or whether they represent a whole combined system at all levels of language processing. Even though the present study does not directly investigate this issue, based on the results discussed above, it is possible to hypothesize that the mental lexicon of proficient L2 speakers is an enriched one compared to the mental lexicon of a monolingual language user.

In this regard, several researchers (e.g., Meara, 1983; Cook, 1992; Kecskes & Papp, 2000) have pointed out that there is no reason to believe that a person who speaks two languages should linguistically behave as a monolingual speaker. Cook (1992) further argues that L2 learners have for a long time been assumed to be aiming at L1 competence. Consequently, they have been compared with the native monolinguals' competence without any regard to evidence that supports multicompetence as a distinctive state of mind. By the same token, Kecskes and Papp (2000) emphasize that cross-linguistic influences work in both directions, i.e. from L1 to L2 as well as from L2 to L1, and as language proficiency increases, the effects of multicompetence at the conceptual level become stronger and more salient. Findings about the interaction between the native language and the L2 lexicon also support the proposed hypothesis. For example, the suggestion that L1 and L2 share the same mental lexicon and that the L2 is somewhat residually activated when a bilingual is in a monolingual state also support the notion that multicompetence is a distinctive state of mind. With this in mind, it is possible to speculate that the advanced learners in this study generated a greater number of associations and a greater variety of associative responses than the NS did, because

a SW in English activated associative links both in English and in their native language (Bulgarian) and, as a result, they produced associations relevant both to L2, as well as to L1.

Whether or not this is the case, the data collected for this study do not allow me to draw any definitive conclusions regarding this issue, though it would make an interesting follow-up study to focus specifically on investigating the extent to which L1 is contributing to a WA task. One way that one could follow this up is to examine a group of bilingual speakers with high proficiency in their second language and determine whether this diverse association set is a general feature of highly proficient bilinguals. At this point, it would be fair to say that the size and the heterogeneity of the advanced learners' WA domain are directly related to the average size of their vocabulary. Similarly, these two characteristics should be looked at as a reflection of the complexity and sophistication of the associative connections proficient L2 users build among the words in their mental lexicon.

Another important way in which the quantitative characteristics of the NS' and non-NS' mental lexicon organization differed concerns the degree of commonality and consistency of their response domains. As found in other WA studies, the NS in this experiment gave a significantly greater number of common responses and a smaller number of idiosyncratic responses than the advanced and intermediate learners. Nonetheless, when I calculated what percentage of the total number of NS' responses those two types accounted for, I found that there was a minute 4% difference between the proportion of their common and the proportion of their idiosyncratic responses. Moreover, the occurrence of the three most popular responses in the NS" data accounted for only 27% of all the whole WA domain, compared to 54% in Yoshida's study (1990), for example. Furthermore, when I compared the strength of the three most popular responses across the three groups of participants, I found that they shared very similar degree of consistency. This naturally led to the conclusion that the NS in this study had a somewhat different WA behavior than NS groups in other studies in that the participants in this

study did not show the same high degree of response commonality. Instead, giving common WA responses was as typical as giving idiosyncratic associations for this group.

I attribute this overall difference in the NS' WA behavior to some of the specificities of the experiment and test design. The experiment design of this study differs in several important ways from the methodology used in other similar L2 WA studies. First, many WA L2 researchers select their SW either from lists containing WA normative data (e.g., Kent-Rosanoff list, 1910) or simply use high-frequency words; nonetheless, this does not prevent them from making generalizations concerning the whole of the L2 mental lexicon. There are at least two problems associated with such a SW selection: First, though valuable resources, the available norming lists that are commonly used by L2 researchers were compiled decades ago; thus, they cannot fully reflect the dynamic nature of associative verbal behavior. Second, as pointed by Meara (1982), most of the words in these norming lists contain high frequency words that tend to elicit the same responses both in the respondents' native and second language, which substantially clouds the overall picture of the L2 WA domain. Third, high-frequency words elicit a high proportion of primary, secondary, and tertiary responses and very few idiosyncratic responses, which may further complicate the interpretation of results of an analysis based on those types of responses. In this study, I tried to address all these problems in an attempt to avoid bias in the collected data linked to the methodology of SW selection. The SW were selected by means of a spaced sampling procedure from a dictionary that contains 16,045 words, which permits generalizations about the nature of the L2 WA domain concerning a very specific vocabulary size. Moreover, several frequency bands were well represented in the sample, which made it possible to collect WA data generated to stimuli with various frequency of occurrence.

A second way the test sample I used differs from the SW lists used in other L2 studies concerns the lexical category of the SW. Many L2 researchers select their SW from a single lexical category, i.e. only nouns, verbs, or adjectives, while the sample used in this study contained content words from all lexical categories. Although the effects of the SW lexical

category on the qualitative characteristics of the WA domain are very well researched in L1 context, to my knowledge, no study has examined the effects of lexical category on the quantitative characteristics of the organization of the L2 mental lexicon, i.e. whether different lexical categories elicit from NS and non-NS broadly comparable quantitative patterns of associations or whether different lexical categories elicit quantitatively distinctive patterns of responses. I feel that the lexical category of the SW does have impact on the size, heterogeneity, and consistency of the WA domain but this issue will be further statistically examined in the next chapter. Nonetheless, it is worth noting that having all lexical categories represented in the test sample might have influenced the quantitative features of the WA domains of both NS and non-NS. Moreover, by choosing this representative sample, I can better generalize to the whole lexicon than prior studies can.

Another methodological difference between the present study and other L2 experiments concerns SW familiarity. Most L2 researchers do not pre-check their participants' degree of familiarity with the SW either because the test-takers are expected to be familiar with high frequency stimuli or because the researchers are not specifically concerned with the effect of SW familiarity on the L2 WA organization. In this study, all participants were pre-tested on their familiarity with the SW and the analysis of the association data was based on responses given to words that participants verified a certain degree of familiarity. As a result, I believe that the discussed assessment is a better indicator of the total state of each group's associative connections surrounding known words.

Finally, there is a procedural difference in the WA data collection between this study and other L2 WA studies, which might well have influenced the degree of commonality of NS' associations. The participants of most of the other L2 WA studies provided their WA in a controlled environment either orally or in a written form, whereas the participants in this study completed the test as a take-home test, without time constraints. By not limiting the time for the test completion, I wanted to avoid the effects of fatigue and allow the participants to work at their

own pace, given that they had to respond to a fairly large number of items. However, there is a high probability that this may have seriously influenced the degree of response commonality. In this regard, Clark (1970) comments that WA can be characteristically different dependent on the "rules that the player" has to follow, i.e. when individuals are allowed to take their time, they generally react with richer images and personal associations, which gives way to more idiosyncratic responses. On the contrary, when individuals are urged to respond quickly within certain time constraints, their associations become less idiosyncratic and more predictable in the sense that the responses they give become very much alike (Clark, 1970, p. 272). I doubt whether this alone could account for the differences found between the three groups since all groups followed the same "rules" but it is reasonable to speculate that the lack of time constraints might have promoted a higher degree of idiosyncratic responding across groups than in other studies.

In summary, the discussed methodological differences between the present experiment and test design and other L2 WA experimental studies may to a large extent explain the fairly lower degree of response commonality of the NS' group. Nonetheless, it should be pointed out here that the degree of commonality and consistency of the NS' associative behavior should be described from at least two perspectives: on the one hand, as a reflection of NS' strength of WA connections and, on the other hand, as a reflection of the influences of certain characteristics of the SW on their responses. This conclusion calls for special attention to be paid to methodological issues in WA experiment designs, such SW selection, method of response elicitation, etc. before drawing conclusion based on data that do not allow for broad generalizations.

So far I have discussed the NS' associative behavior. Now I turn to non-NS' quantitative features of associative responses. The analysis revealed that their associations shared very similar patterns of commonality and consistency, which, however, were not influenced by their differing levels of language proficiency. Both non-NS groups gave more idiosyncratic than

common responses compared to NS. Apparently, the lower degree of commonality of responses and the higher degree of idiosyncrasy of non-NS' associations, also found in several other studies (e.g., Lambert, 1972; Meara, 1978; Kruse et al., 1987, Wolter, 2002), is a distinctive pattern that characterizes L2 learners' WA domain as a whole. More importantly, increase in language proficiency does not seem to have any significant effect on the overall quantitative response pattern of L2 learners. One possible interpretation of this finding is that, at least, for intermediate and advanced learners of English the high degree of idiosyncrasy and the relatively low degree of commonality of their associations can be thought of as typical of L2 learner's mental lexicon structure rather than as a flaw in the organization of their lexical knowledge. In other words, it is highly plausible that this WA behavior is a result of L2 learners bilingualism. While beginning learners consistently show a high degree of commonality of their response patterns because their vocabularies are small in size (Kruse et al., 1987), an increase in proficiency, as confirmed by the results of this study, leads to a considerable increase in the production of idiosyncratic associations rather than to an increase in the common responses. In fairness, it should be mentioned that as language proficiency increased, the commonality of advanced learners' responses also strengthened, though, it did not come close to the degree of NS' commonality. At the same time, the idiosyncrasy of their associations also immensely augmented and their idiosyncratic responses significantly outnumbered the ones of NS and intermediate learners. So, it will be safe to conclude that response commonality and idiosyncrasy increase with proficiency, but the overall pattern remains the same.

Interestingly, the conclusion concerning the lesser degree of L2 learners' commonality of responses is consistently confirmed, regardless how commonality is measured. In Chapter 2, I compared what I called there *typicality of associations* across the three groups of participants, where *typicality* was a measure of the degree of *native-likeness* of non-NS' associations, i.e. non-NS' typicality was measured with respect to the absolute occurrence of NS' associations in the norm list. The results of that analysis also showed that neither the advanced nor the

intermediate learners have reached the extent of typicality of associative behavior of NS. In the present analysis, commonality of responses was based on within-group associations. Nonetheless, the results once again confirmed that an increase in language proficiency does not improve the overall quantitative pattern of commonality and consistency of L2 learners' WA to the extent that it would approximate the strength of NS' associative commonality. In the context of this conclusion, then, it can be speculated that the L2 WA behavior is motivated by learners' native language, in which case it should be assumed that, in general, L1 speakers of Bulgarian maintain looser WA domains than L1 speakers of English. Thus, when Bulgarian learners of English as a second language associate in L2, they transfer this WA feature of their native language onto their L2 WA behavior, which gives the impression of loosely associated lexicon in comparison to NS' lexicon. Yet, another possible reason for the high idiosyncrasy of non-NS' responses can be looked for in the effects of instruction and the use different teaching materials, especially if the fact that both L2 groups acquired English in a non-English speaking setting is taken into account. My results are more consistent with the first hypothesis, though it would be interesting to explore the other possibility as well.

Qualitative analysis of word associations

Method

The purpose of this section is to investigate whether there are qualitative differences in the structure of the mental lexicon of NS and non-NS of English at two levels of language proficiency. To this end, I compared the WA of NS, advanced, and intermediate language learners with regard to the types of associative responses they made (paradigmatic, syntagmatic, and phonological).

A three-language proficiency group (NS, advanced, and intermediate learners) between subjects design was used. The dependent variables were type of associative responses (paradigmatic, syntagmatic, and phonological) provided for each SW participants verified they

were familiar with, as well as for SW for which they showed only vague familiarity (frontier words). Thus, unlike the larger study, the main dependent variables used item means as a random factor, rather than subject means, since the SW were randomly selected and I was more interested in each group's performance rather than in individual variation in associative behavior.

Participants

Fifty-one adults, native and non-native speakers from the pool of 64 participants, were involved in the study. To ensure equal number of participants responding to each SW in the test, 17 students were randomly selected from the group of NS, which originally consisted of 30 participants. Thus, there were 17 participants in each group (NS, advanced, and intermediate learners of English) who responded to 73 SW. The SW, not the participants, were used as a between-group factor since I was more interested in the proportion of responses generated to each SW than in individual variation among members of each group in their production of different types of associations.

Materials

The data were collected in the manner described earlier in this chapter.

WA classification procedures

Two lists of associations were compiled for each group: one with associations to words with which participant showed a satisfactory degree of familiarity, and a second one with responses to words that participants believed or thought they knew, but in actuality they did not have sufficient knowledge. It has been suggested in the L2 WA literature that different degrees of familiarity with SW may evoke different WA patterning of L2 responses. However, this relationship has not been sufficiently well researched in L2 context, as it has in L1, so the two

lists of responses were compiled in an attempt to investigate this issue. Given that the associations were classified in two lists based on degree of familiarity, it was of paramount importance to establish criteria for acceptability of synonyms, translations, or brief explanations that satisfactorily proved familiarity with the SW. For a synonym or a translation of the SW to be considered an acceptable verification of SW familiarity, they had to be semantically and syntactically relevant to the SW. For example, the SW *hard* was specified as an adverb in the test. So, synonyms such as *strenuously, laboriously, persistently, tightly,* etc. satisfied the verification requirement and all associations following those or similar synonyms were included in the list of familiar words (List 1). Synonyms such as *difficult, strong, tough, solid, firm*, etc. rendered the SW vaguely familiar because of participants' failure to respond correctly to the lexical category of the SW. Therefore, the associations that followed such synonyms were included in the list for frontier words (List2). The requirement that a response should adequately reflect the semantic and the syntactic property of the SW was applied to the translations, as well. When a brief definition/ or explanation was provided, the decision about its acceptability was based on criteria used in lexicography for defining words by part of speech. These criteria were discussed in detail in Chapter 2.

The description of the qualitative characteristics of WA responses in L1 research is traditionally based on the analysis of several types of responses: paradigmatic, syntagmatic, phonological (clang). Each participant's association was classified as one of those types dependent on its relationship to the SW. The distinction between the different types of responses adopted in this study is the one commonly used by L1 and L2 researchers (e.g., Ervin, 1961; Deese, 1965; Clark, 1970; Piper & Leicester, 1980; Söderman, 1993; Reed, 1993; Wolter, 2001). An association was classified as paradigmatic if it shared same lexical class with the SW (e.g., *advantageous – beneficial*) and as syntagmatic if it belonged to a different lexical class or could form a syntactic string with the SW (e.g., *advantageous – behoove; pillar - community*). Phonological (clang) associations were defined by Meara (1978), as responses

that rhyme with the SW (e.g., *inception* - *inspection*), have a similar consonant cluster as the SW (e.g., *crampons* – *tampons*), repeat a morpheme of the stimulus (e.g., *withdraw* – *with*, *draw*), or preserve the consonants unchanged (e.g., *beat* – *bit*). In other words, those are associations related to stimuli in phonological terms only (Singleton, 1997), without being semantically or syntactically linked to it. Unlike Söderman (1993) and Wolter (2001), who classified the morphological derivations as repetitions under the category "other" (e.g., *sleep* – *sleepy*; *beautiful* – beauty; *thief* – thieves; *foot* – *feet* (Söderman, 1993); and *concentration* – *concentrate* (Wolter, 2001), I treated morphological derivations as either syntagmatic or paradigmatic because they show knowledge of word derivation processes rather than some sort of an anomalous association behavior. In addition, derivations also bare a semantic relationship with the SW, which is of paradigmatic (e.g., *amoral* – *moral*) or syntagmatic (e.g., *advantageous* – *disadvantage*) nature that cannot and should not be disregarded. The key principle of classifying a response as clang was lack of semantic or syntactic relatedness to the SW that was realized in any of the ways specified by the criteria suggested by Meara (1978).

It is impressive how infrequently the difficulties of classifying certain responses are commented in L2 WA research, though it is crucial for any WA qualitative analysis to use strict criteria for classifying ambiguous WA responses. Not to mention that any simplified approach to WA coding inevitably results in a bias in the interpretation of the results. It has been only recently that L2 researchers (e.g., Piper & Leicester, 1980; Söderman, 1993; Wolter, 2001) have been more specific about the criteria they applied to their data classification by listing detailed descriptions of response coding, most of which, by the way, follow the tradition of specifications employed by L1 WA research. Therefore, to maintain consistency in the response classification procedure in this study, all categories into which the responses were classified were specified in advance following principles used in other similar studies. Unfortunately, there is no certain way to eliminate the grammatical ambiguity of some responses, especially in cases in which responses could be coded into either paradigmatic or syntagmatic category, but the

following criteria were applied in the classification of the association data in an attempt to limit inconsistencies associated with ambiguity:

1) A response was counted as paradigmatic when it was from the same word class as the SW.
2) Multiword responses (e.g., *good opportunities, marching line, bed cover,* etc.) were classified according to the relationship between the SW and the lexical class of the head of the phrase.
3) A response was classified as syntagmatic when, generally, it did not belong to the same form class of the SW.
4) In the cases when the SW was repeated (e.g., *second – second place*, *back – back up*), a response was counted as syntagmatic (Piper & Leicester, 1980).
5) When a response showed a clear sequential relationship to the SW, even though it was from the same form class with the stimulus, it was classified as syntagmatic (e.g., *pillar – salt; cassava - leaves*).
6) Following Deese (1965), in all cases in which a response could be classified as either paradigmatic or syntagmatic (e.g., *blanket (n.)– cover*), it was coded as (potentially) syntagmatic.
7) Associations were classified as clang responses when they were related to a SW only in phonological terms, showing no semantic or syntactic relationship to it (e.g., *virtuosity – virtue*), or when they revealed some other kind of misinterpretation of the SW (e.g., *choke (n.) – writing, board; blanket (n.) - test*).

I believe that this detailed specification of response classification criteria greatly reduced the number of subjective judgments in coding the associations as they were consistently applied across the whole set of data. At the same time, most of them were used by other L1 and L2 WA researchers, which would allowed for comparisons of the present analysis with findings of other similar studies.

Results

To test the influence of language proficiency on the type of association responses, a One-Way ANOVA was conducted on the proportions of responses from each association type (paradigmatic, syntagmatic, phonological) elicited by each known stimulus word, using language proficiency as a between-subjects factor. The results of the analysis of the three types of responses that participants generated to SW they knew yielded a non-significant group effect on the outcome variables (see Appendix K), thus indicating no statistically significant differences among the three groups (NS, L2 advanced, and intermediate learners) in their mean proportion of paradigmatic, $F(2, 216) = 2.63$, $p > .05$, and syntagmatic responses, $F(2, 216) = .814$ $p > .05$. Interestingly, none of the participants produced clang associations for the words they knew, which means that familiarity with the stimuli rather than, as previously thought, loose organization of the L2 mental lexicon motivates elicitation of more phonologically than semantically or syntactically based associations. The overall ANOVA supported the conclusion that L2 advanced and intermediate learners' qualitative characteristics of lexical knowledge organization are not different from the qualitative characteristics of NS' mental lexicon. Moreover, the overall pattern of responses was the same across the three groups, i.e. the participants of the three groups gave more paradigmatic than syntagmatic responses to the words they were familiar with, and no clang associations at all (see Figure 5).

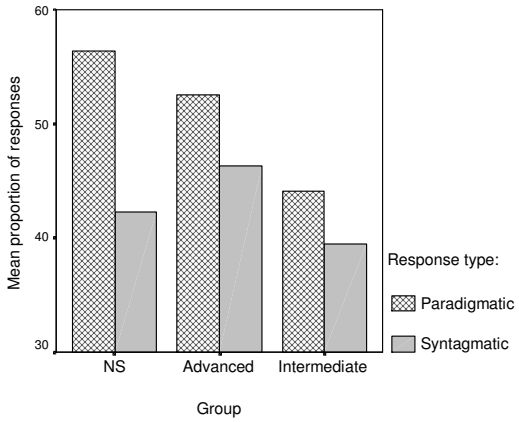


Figure 5. Mean proportion of paradigmatic and syntagmatic responses of NS, advanced, and intermediate learners given to SW participants know.

It was immediately noticeable on the graph that the response patterns were strikingly similar, i.e. adult NS' and non-NS' mental lexicons were connected predominantly along paradigmatic lines for the words they know. In other words, contrary to findings of other studies (e.g., Meara, 1978; Piper & Leicester, 1980; Wolter, 2002), which found that the major qualitative difference between NS' and L2 learners' association patterns was linked to L2 learners' overwhelmingly syntagmatic connection between words in their mental lexicon, my results supported a qualitatively similar overall pattern, which was only quantitatively different across the three groups. When I calculated the proportion of paradigmatic and syntagmatic responses within each group's total number of responses, I found that for the three groups the proportion of paradigmatic associations accounted for 51% of each group's total number of responses, and the syntagmatic accounted for 49%. This added further support to the argument I made earlier that the differences of word connections between NS and L2 learners, who have obtained an intermediate and advanced level of proficiency, were related to the size of their paradigmatic or syntagmatic domain rather than to the type of their association connections.

Interestingly, the analysis of participants' association patterns of the words they thought or believed they knew but, in reality, they didn't have sufficient familiarity with (List 2) revealed a

different picture. It was easily noticeable that this set of associations contained responses given as a result of misinterpretations of the grammatical class of the SW, as well as semantic misinterpretations of the SW. It was relatively easy to distinguish between those two types of misinterpretations by considering the synonyms, the explanations, or the translations that participants gave to the SW before generating their responses. For example, explanations such as *to take away courage*, or *to release stress*, or *to relax* given to the SW *unnerve* were taken to be semantic misinterpretations since the explanations revealed a misinterpretation of the SW based on overgeneralization of the meaning of the prefix and the free morpheme, though the grammatical glass was preserved. Whereas explanations such as *difficult*, *solid*, *not soft*, *not easy*, etc. to the stimulus *hard*, which was intended to be responded as an adverb in the test, were taken to be lexical class misinterpretations, since the brief explanations revealed participants' confusion about the syntactic category of the SW. Initially I thought that it would be a good idea to analyse the grammatical misinterpretations separately from the semantic misinterpretations and see whether there were differences among the groups on the interpretations alone. However, the statistical analysis showed that the three groups differed only in their semantic misinterpretations of the stimuli they vaguely knew, $F(2, 216) = 8.33$, $p = .000$, but not in misinterpretations of the SW lexical class, $F(2, 216) = 1.18$, $p > .05$ (see Figure 6).

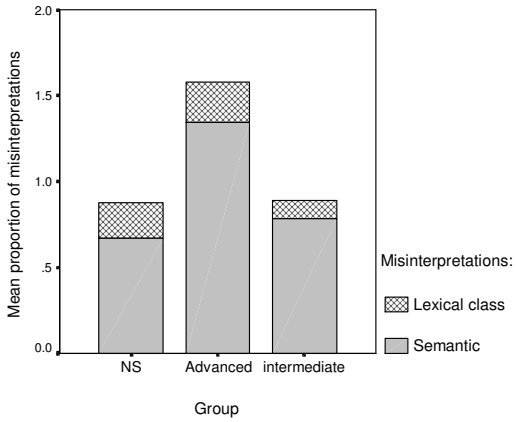


Figure 6. Mean proportion of misinterpretations of lexical class and meaning of SW participants do not know.

This result convinced me not to analyze the lexical class misinterpretations separately from the semantic ones but to examine them as one whole set of misinterpreted SW. It also showed that all participants had a strong sense of grammatical category for the words they vaguely knew and their misinterpretations were mostly semantic in nature rather than syntactic.

I further examined the patterns of paradigmatic, syntagmatic and clang associations generated to SW participants vaguely knew by using the same criteria described earlier. The analysis showed that the three groups differed significantly only in the mean proportion of their syntagmatic responses, $F(2, 216) = 5.60$, $p < .00$, but not in the mean proportion of their paradigmatic, $F(2, 216) = 1.79$, $p > 05$, and clang associations, $F(2, 216) = 1.99$, $p > .05$. Means and standard deviations are presented in Appendix L. Furthermore, the multiple pair-wise comparisons conducted at .05 level of significance clearly showed where exactly the differences stemmed from. As shown in Figure 7, the advanced learners produced a greater proportion of paradigmatic as well as syntagmatic responses than the other two groups did. However, the only significant effect was for the mean proportion of their syntagmatic associations compared to NS' syntagmatic responses (95% CI = 4.881, 31.544; $p < .05$).

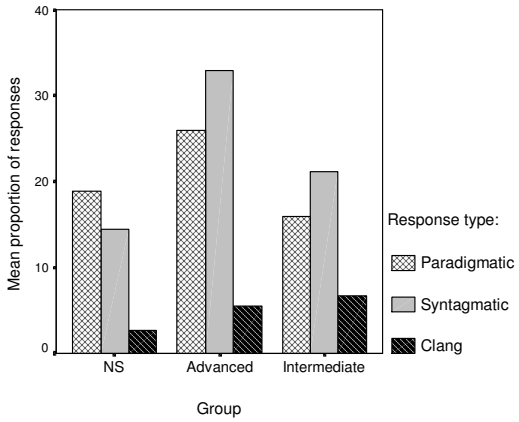


Figure 7. Association patterns of NS, advanced, and intermediate learners elicited to vaguely familiar SW.

As a whole, the most interesting finding of this analysis was related to the overall change in the response pattern for the advanced and intermediate learners. When they responded to words they vaguely knew, L2 learners as a group supplied much more syntagmatic than paradigmatic, thus confirming the hypothesis that familiarity with prompt words affected the response patterns. It was also noticeable that while there were no clang associations generated to the stimuli participants knew, vaguely familiar stimuli elicited semantically unrelated but phonologically linked to the form of the SW associations. Nonetheless, as the graph shows, the vaguely familiar SW apparently attracted few clang associations among all participants, though they did show a decrease as proficiency increased.

In summary, the three groups revealed a strong sense of grammatical category of stimuli they thought they knew but they actually did not, which showed that misinterpretations of words by NS or L2 learners at intermediate and advanced level of language proficiency are mostly meaning based. The degree of familiarity with words proved to have an effect on the response patterns of NS and L2 learners. While for the words participants knew, the response pattern across the three groups showed paradigmatic - syntagmatic dominance, it noticeably changed with a decrease of level of familiarity. Similarly, the lower the proficiency, the more detectable

the difference in the response pattern was. In other words, while the NS still maintained a paradigmatic - syntagmatic dominance, there was a shift in the response pattern of advanced and intermediate learners to somewhat syntagmatic - paradigmatic preponderance in their responses to frontier words. The implications of this shift will be discussed in the following section.

Discussion

The analysis of the associative responses of the three groups of language users aimed at investigating whether NS, advanced, and intermediate learners of English organize their mental lexicon in qualitatively different patterns. In addition, I included a second dimension to this examination by looking at paradigmatic, syntagmatic, and clang associations with respect to the effects of participants' degree of familiarity with the SW.

Interestingly, the results of the associations given to words that participants knew revealed that there were no qualitative differences in their overall association patterns. That is, all participants gave predominantly paradigmatic associations and less syntagmatic responses to stimuli they knew. Apparently, as proficiency develops and the size of vocabulary increases, the potential to generate semantically or syntactically related responses develops towards NS' capacity. Quite surprisingly, the overall response pattern to SW participants knew across the three groups showed a clear tendency to a paradigmatic–syntagmatic dominance. Thus, this finding adds strong support to Söderman's arguments (1993) that adult L2 learners pattern their responses very much like adult NS, even at lower levels of language proficiency, and that SW familiarity can have a greater effect on response patterning than usually acknowledged.

By and large, there are surprisingly few studies in the field concerning the overall patterns of L2 associative responses, which, nonetheless, yielded surprisingly controversial results. The most popular argument about associative patterning of L2 learners is based on findings in L1 concerning a syntagmatic-paradigmatic shift noticed with children as they grow

and develop their lexicons (e.g., Ervin, 1961). Several L2 researchers have been attracted to the idea of response shift and have tried to draw parallels between L1 children's and L2 learners' developing lexicons. Their major point of interest has been whether there is a tendency for L2 learners to produce more syntagmatic and clang associations than paradigmatic responses, as L1 children tend to do, or whether a L1 adult response pattern (i.e. paradigmatic-syntagmatic) will dominate from very early in the learning process (e.g., Meara, 1978, 1982; Wolter, 2001). Meara (1978) reported that the participants in his study, a group of L2 learners of French preparing for the 0-level examination in French, produced high proportion of paradigmatic responses, few and quite unsystematic syntagmatic associations, and a surprisingly large number of clang and un-French associations. Thus, the researcher concluded that this was indicative of a lack of a proper semantic organization for the L2 mental vocabulary, which could largely explain why L2 learners experienced difficulties in processing both written and spoken language. It should be noted that Meara (1978) used only one group in his study, assuming relatively equal level of proficiency among his participants, which could have been any level. Nonetheless, he made broad generalizations concerning L2 learners in general. Similarly, Wolter (2001) also studied only one group of non-NS, L2 Japanese learners of English, in comparison to NS of English, but added the dimension of SW familiarity to his analysis. In the main, his findings echoed Meara's (1978) conclusion in that his analysis revealed considerable difference in the patterning between NS and L2 learners for the words they know. That is, he found that syntagmatic connections play a more important role in the organization of non-NS' lexicon, though, he was in haste to add that the L2 syntagmatic dominance did not necessarily mean that the organization of the L2 mental lexicons would be inferior to the L1 lexical organization.

Regrettably, very few studies involve more than one group of L2 learners, which may explain why generalizations concerning the organization of the L2 lexicon do not hint at the possibility that factors other than the broad distinction native - non-native speaker can have

specific effects on the structure of the mental lexicon. Even more surprising is that almost no research attention has been devoted to examining how factors unrelated to L2 learners, but related to characteristics of the SW used in WA tests can influence the association patterning of L2 participants. To my knowledge, the few notable exceptions in this regard are studies conducted by Piper and Leicester (1980), Söderman (1993), as well as the present study, which involved more than one group of non-NS in investigated the effects of language proficiency in combination with variables pertaining to SW familiarity on the WA patterns of L2 learners. The results of this analysis, relating to the effects of SW familiarity on the WA patterns, are more in concert with Söderman's (1993) findings because the researcher made a finer distinctions in her grouping variable (language proficiency) than Piper and Leicester (1980) did in their study. Her four groups of participants, Finnish-Swedish learners of English, included at least three broadly distinguishable levels of proficiency, i.e. a beginner's, intermediate, and advanced level, whereas Piper et al. compared two largely different groups of L2 learners, advanced and beginning Japanese learners of English, with NS. In the main, this study confirms the very general conclusion from both previous studies that there are no differences in the patterns of lexical organization between NS and L2 learners at higher levels of proficiency of English in that they build their lexical connections very much like NS do. Apparently, as proficiency increases, learners vocabulary size increases too which, in turn, allows them to generate a greater number of responses of both paradigmatic and syntagmatic character. The most important thing is that the paradigmatic-syntagmatic pattern is well preserved along these proficiency levels. Similarly, my findings do not support the hypothesis adopted from L1 research of a developmental shift from more "child-like" syntagmatic associations to more "adult-like" paradigmatic responses in L2 associations. On the contrary, my inquiries into the subject suggest that this need not be the case and L2 learners with an intermediate size of vocabulary (approximately 5,000 words and above) have stable paradigmatic-syntagmatic patterns of connection among the words they know. Moreover, none of the participants produced clang associations for the words in this

category, which allowed to hypothesize that SW familiarity rather than loose organization of the L2 mental lexicon motivates elicitation of more phonologically than semantically or syntactically based associations.

This hypothesis was confirmed by the analysis of participants' association patterns of words they thought or believed they knew but, in reality, they did not. Those words were called vaguely familiar words because participants were able to place them in the correct syntactic category but they either had partial knowledge of their meaning or misinterpreted them based on form similarities. Durso and Shore (1991) referred to such words as "frontier words", a label they adopted from Trembly (1966, cited in Durso & Shore, 1991), meaning words that participants judge as familiar but fail to acceptably define. Examining the associations that the participants of this study generated to vaguely familiar words allowed for finding out how a lower level of familiarity would affect the overall patterning of associations among the three groups, if at all.

Somewhat as a surprise came the result that, on the whole, the category of clang associations attracted very few responses among the three groups of participants and they did not show a decrease with an increase in proficiency, as expected. Usually, this is the category of responses that is of greatest interest to L2 WA researchers because it reveals the greatest variation among native and non-native speaking participants. However, in this study it turned out to be the least interesting, from research point of view, category because the analysis showed no statistically significant difference among the three groups on this type of responses. I attribute this result to the very stringent criteria that I used in the specification of what responses could be classified as clang associations. That is, I strictly adhered to the inclusion of phonological misinterpretations of the SW only, which probably can explain why so few responses fell under this category. Unfortunately, very few researchers (e.g., Meara, 1978, Söderman 1993; Wolter 2001) have been specific about their response classification criteria, which in turn makes it difficult to interpret the present analysis in light of findings of previous

studies. Furthermore, I feel that some decisions of including certain types of responses under the heading “clang - other” associations in previous research are not well justified. For example, in addition to the pure phonological misinterpretations, Wolter (2001) included morphological derivations (e.g., *concentrate* – *concentration*) as well as what he called “nonsensical” responses (e.g., *anticipation* – stand) (p. 52) under his category of clang - other responses. Söderman (1993) identified morphological derivations, anomalous responses (associations that show no obvious relationship to the SW), and purely Swedish responses (translations from L1) under a separate category that she called “other”. I feel that these criteria are neither theoretically nor practically justifiable because, on the one hand, the morphological derivations are an inseparable part of the organization of any mental lexicon. On the other hand, judgments based on researcher’s intuitions of “nonsensical” can be very subjective, especially when no specific semantic criteria are used. Let’s say, a stimulus *eternity* would always elicit from me a response *white*; it may elicit from other people associations such as *love*, *heaven*, *conversation* etc. The question is: on what grounds some of those associations would be judged as related to the SW and others as “nonsensical”? Or, how would a researcher objectively justify that *white* as an association sounds more Bulgarian-like and *conversation* is more American-like, for example. Therefore, in an attempt to avoid any such subjective judgments of associative responses the category that I had for “clang” responses covered very specific cases. Unfortunately, for reasons I mentioned earlier, the inclusion criteria do not allow for comparison of my participants’ performance on the phonological associations with results from other studies. Nonetheless, the analysis supports a conclusion that with an increase in language proficiency, learners tend to produce fewer clang associations, which, anyway, comprise a very small proportion of their response repertoire. Clang associations accounted for only 7% (n = 13) of NS’ responses to vaguely familiar words, 9% (n = 28) of the advanced learners’ associations, and 16% (n = 31) of the intermediate learners’ WA. Furthermore, I can say that the L2 learners in this study, markedly do not show preference for connecting the lexical items they vaguely

know along phonological lines, which might to a great extend be influenced by their adult associative habits in their native language.

Probably one of the most interesting findings of this line of analysis concerns the overall change in the response pattern of the advanced and intermediate learners when they responded to words they were vaguely familiar with. That is, although for the words L2 learners knew, their association pattern showed paradigmatic-syntagmatic dominance, this pattern noticeably changed with a decrease in their level of familiarity with the SW. In other words, while the NS still maintained paradigmatic - syntagmatic dominance, there was a shift in the response pattern of advanced and intermediate learners to syntagmatic - paradigmatic preponderance in their responses. Apparently, less familiar words tend to activate more syntactic than semantic connections with L2 learners than they do with NS. Similar results have received varying interpretations in the L2 literature. Wolter (2001), for example speculated that a syntagmatic preference of associating less familiar words (in the VKS category 3) was indicative of a childlike associative behavior that, in turn, was suggestive of a loosely structured mental lexicon which is at an early stage of its development (p. 60). However, the author found out that his participants maintained the same syntactic dominance not only for words they thought they knew but for prompt words they verified they knew too. Consequently, after reviewing some L1 studies on child-like associations and the few L2 studies on WA patterning, the author argued that there was not sufficient empirical evidence to conclude that a syntagmatically dominated lexicon would be less functional than a paradigmatically dominated one (p.63). Thus, he summarized in his argument speculations about this issue put forward by several other researchers earlier (e.g., Ervin, 1961; Stolz & Tiffany, 1972; Piper & Leicester, 1980; Söderman, 1993). Unfortunately, they did not elaborate on the implications of a syntagmatically-dominated lexicon for adult language users.

With regard to this issue, I would like to distance myself from the “child-like/ nonnative-like” metaphor and all the speculations that stem from it because, on the one hand, there is not

sufficient L2 research to support such a speculation. On the other hand, recent L1 research suggests that SW familiarity, and respectively size of vocabulary, have a great deal to do with the predominant production of syntagmatic responses, and that it is not simply a developmental issue. Along these lines, an adult L2 learner is different in many ways from a child, not only physically and cognitively, but linguistically as well. As pointed out by Singleton (1997), an adult L2 learner is already aware of the way words relate to the extralinguistic reality of his/ her experience, as well as the ways in which they may relate to each other in his/ her native language. Therefore, the production of relatively more syntagmatic than paradigmatic responses to SW L2 learners have partial knowledge of can be a result of several reasons other than unstable, underdeveloped, child-like lexicon. Though, I fully agree that a L2 lexicon is developing, but so is the L1 lexicon. At this point, I can only suggest some possible reasons for a shift in response patterning because this study does not directly address the issue of shift. I also believe that each of the ideas discussed below deserves to be empirically investigated in its own rights.

One plausible explanation of L2 learners' shift pattern when they respond to vaguely familiar words may relate to differences in formal instruction between NS and L2 learners. Ervin (1961) suggested that it was very possible that older American children's production of more paradigmatic than syntagmatic and clang responses reflected their educational experience of practicing a lot of substitution of antonyms and synonyms in sentences at school. She further speculated that this may also account for the greater response commonality of associations of adult Americans, compared to Europeans, found by several researchers (e.g., Rozenzweig, 1961, p. 370). Taking this a step further I suggest that, considering adult NS' general awareness of lexical category, they tend to give more paradigmatic responses even to words they vaguely know as a result of both their strong sense of grammatical class as well as educational experience. The L2 learners participating in this study have acquired English through formal instruction that, for the most part, emphasizes development of language skills rather than

language alone. Therefore, a less familiar SW may more readily activate in the mind of a L2 learner some sort of a syntactic string in which the word may have been experienced, or assumed to have been experienced, rather than semantic features that would promote a selection of a paradigmatic association.

Another feasible explanation is suggested by Clark (1970), who argues "although syntagmatic responses first *appear* to be different from paradigmatic responses, they are produced by rules that belong to the same class of rules stated for paradigmatic responses – the simplicity of production rule" (p. 284). In Clark's account this rule states: "Perform the least change on the lowest feature, with the restriction that the result must correspond to an English word" (ibid.). Following this line of reasoning, it is highly possible that for SW L2 learners do not have sufficient knowledge, "the least change" operation simply entails a left-to-right completion of a syntactic string that includes the SW or a realization of the selectional features of the SW after recognition of its lexical category. For example, the SW *forgo* was misinterpreted by one advanced L2 learner to mean *forgive* and he/ she responded to it with *unjust*, *relief*. These associations show a clear syntagmatic preference, whereas the primary responses of participants in the same group who knew the word were overwhelmingly paradigmatic (e.g., *give up*, *skip*, *refuse*, *forget*, etc.). In contrast, NS who had a similar semantic confusion as the advanced learner generated mostly paradigmatic associations (e.g., *change mind*, *go farther*, *resent*, etc.), which may be interpreted to mean that for a NS "simplicity of production" entails preserving the lexical class of the SW in associative responses, thus, responding to a vaguely familiar words very much like responding to a familiar one.

Yet, another explanation derives from findings in L1 research concerning the conditions that provoke adults to generate more syntagmatic responses for less familiar words than an expected paradigmatic predominance. Stolz and Tiffany's study (1972) is probably the most commonly cited in the L2 literature research that provided strong support to a hypothesis put forward earlier by Ervin's (1961). Ervin (1961) suggested that an increase in the paradigmatic

responses of older children does not necessarily demonstrate a systematic decrease in the remaining responses. Stolz et al. (1972) took this a step further and argued that syntagmatic responses do not dramatically decline for adults. Rather, their rates of occurrence tend to stabilize earlier in the learning process (compared to paradigmatic or clang connections, for example), thus, indicating information learned early in an individual's experience with a given word (p. 44). On the other hand, the researchers also reported that adult responses to unfamiliar adjectives pattern very much like children's responses to familiar words, i.e. they show a tendency of syntagmatic prevalence. Therefore, the results made them conclude that the primary cause of a response shift in children, as well as in adults, is the acquisition of additional lexical material and overall development of a person's lexicon rather than maturation of more sophisticated mental processes (p. 45). This conclusion points to several questions concerning the L2 lexicon, which definitely deserve research attention. For example, how does SW familiarity influence non-NS' WA patterns responses compared to NS'? Is the difference in the response pattern between familiar and unfamiliar words for these broadly defined groups so dramatic as to be attributed to loosely organized mental lexicon? How is overall size of vocabulary related to certain association connections? The presented data give some answers to these questions and, as shown in Table 7, the response pattern changed for the three groups but not so dramatically as to indicate a syntagmatic-paradigmatic response shift. Rather, it seems to indicate what Wolter (2001) calls "a shift from semantically meaningless (i.e. clang and unclassifiable responses) to semantically meaningful responses" (p. 63), which is more related to lexical development of vocabulary items and much less to stability of connections among them. As the table shows, the intermediate learners generated the greatest proportion of clang associations, which were the only type of associative responses acknowledged as non-semantically related, while the advanced learners' and the NS' proportion of responses was very much in the same range. Therefore, I can fairly confidently say that in the process of transition from form-based misinterpretations to semantically-related associations, which is

Table 7

Proportions of response types for familiar and vaguely familiar words

Stimulus type	NS	Advanced	Intermediate
Familiar words:			
• paradigmatic	51%	51%	51%
• syntagmatic	49%	49%	49%
• clang	-	-	-
Vaguely familiar words:			
• paradigmatic	52%	40%	43%
• syntagmatic	41%	51%	41%
• clang	7%	9%	16%

largely dependent on the degree of word familiarity, L2 learners' clang associations become paradigmatically or syntagmatically stable as a function of certain characteristics of the words that change their status from vaguely familiar to familiar. This view would support a conclusion that the L2 mental lexicon is not subject to developmental processes that manifest themselves in a complete paradigmatic dominance. Rather, a language user's lexicon strives for a balance between paradigmatic and syntagmatic connections as it expands and perhaps some lexical categories, certain frequency bands, or both, would affect the type of associative connections favored by an individual. This hypothesis will be the focus of the next chapter, which will examine the effects of certain characteristics of SW, in particular lexical class and frequency of occurrence, on the associative connections among words in the mental lexicon of NS and non-NS and how they interact with language proficiency and familiarity, because there are good reasons to believe that the way language users associate lexical items in their mental stores cannot be independent of features of the lexical input.

Conclusion

Word association tests have served various research purposes, dependent on the theoretical perspective taken in the interpretation of WA data. One of their practical uses as a research tool relates to contemporary theories of associative structure, which propose that WA represent the way in which semantic information is structured in memory (Nelson, 1977). As pointed out by Nelson (1977), “from this view point the study of word association structure is another approach to the organization of semantic memory, a subject worthy of studying on its own terms, without regard to its connection to linguistic or cognitive function” (p. 102). In the context of L2 research, WA tests have been very often employed to examine how L2 learners organize their lexical and semantic information, i.e. whether their lexical knowledge is structured much the same way as NS’ knowledge or whether there are fundamental differences in the way they build connections among the lexical items in their mental stores. Psycholinguistic research interested in the conceptual representation in bilingual memory has taken the use of WA tests even a step further by assuming that the word association task reflects not only conceptual but lexical processing as well (van Hell & de Groot, 1998). The more interesting question that they have been trying to find an answer, however, is whether conceptual representations are purely language specific or purely shared between languages. While the present study does not allow for speculations concerning bilingual conceptual representations, it was designed to investigate the qualitative and quantitative features of the lexical store of L2 learners. That is, the aim of the presented analysis was to gain insight into the organization of L2 lexical knowledge by comparing the features of their WA domain with those of NS’. In addition, it focused on the role of language proficiency in the associative patterning by looking at the effects of increased language proficiency and word familiarity on the L2 associative connections. It also aimed at finding out whether differences in the NS and non-NS lexicons are predominantly qualitative or quantitative in nature in an attempt to elaborate on some interpretations of L2 WA results advanced by previously conducted L2 WA studies.

In brief, the results of present study support a conclusion that familiarity with words plays a central role in the way NS and non-NS organize qualitatively their mental lexicons. When participants were familiar with stimuli, there were no qualitative differences in their overall association patterns, in which words were connected mostly along paradigmatic than syntagmatic lines. However, a lower degree of familiarity broadly distinguished NS from non-NS by pointing to differences in their associative connections that were, otherwise, unaffected by the L2 learners' level of proficiency. In this regard I argued that the detected differences can hardly be considered "a shift" in response type that indicates a loosely connected L2 lexicon that resembles the lexical organization of L1 English speaking children. Rather, it indicates a transition from semantically meaningless (clang associations) to semantically meaningful responses, which is related to lexical development of vocabulary items that comes with increased familiarity and not so much to stability of connections. Actually, the NS and the L2 learners participating in this study showed little preference for connecting the items they vaguely knew along phonological lines, which was most probably influenced by their adult associative habits and strong sense of grammatical class of the stimuli.

However, the examination of the quantitative characteristics of the mental lexicon of the three groups revealed that NS' and non-NS' mental lexicons are quantitatively different, dependent not so much on word familiarity than on the proficiency of the L2 learners. The analysis showed that the intermediate learners had a much smaller and less diverse WA repertoire than the NS and advanced learners, whereas the advanced learners' associative domain was noticeably larger in size and more heterogeneous than NS' domain, which made me conclude that language users with larger vocabularies have a considerably richer connections, both in size and heterogeneity. Another important way in which the quantitative characteristics of the NS' and non-NS' mental lexicons differed, this time regardless of language proficiency, concerns the degree of commonality and consistency of their response domains, i.e. the NS gave a greater number of common responses and a smaller number of idiosyncratic

responses than the advanced and intermediate learners. In this regard I suggested that the lower degree of commonality and the higher degree of idiosyncrasy of non-NS' associations, found in other L2 WA studies as well, should be seen as a distinctive pattern that characterizes L2 learners' WA domain rather than a flaw in the organization of their lexical knowledge. It was also hypothesized that this typical of L2 learners associative behavior may well be a result of their bilingualism. In other words, while beginning learners' associations have been found to maintain a high degree of commonality because of their small vocabularies, an increase in proficiency, as confirmed by this study, leads to an increase in the production of idiosyncratic associations rather than to an increase in the common responses. Therefore, based on the results of this study, it can be concluded that response commonality and idiosyncrasy increase with proficiency, however, the overall pattern remains unchanged.

In sum, the analysis of the WA data generated by Bulgarian advanced and intermediate learners of English revealed quantitative but not qualitative differences in the way they structure their lexical knowledge, compared to NS of English. The quantitative differences were most noticeable in the intermediate learners' WA domain, which suggests that they are a result of quantitatively small vocabularies of approximately 6,000 words. As vocabulary size increases to 8,000 words and above, the quantitative differences vanish, with the exception of commonality of responses, which does not come close to NS' commonality.

CHAPTER 5

EFFECTS OF LEXICAL CATEGORY AND FREQUENCY OF OCCURRENCE OF STIMULUS WORDS ON THE QUALITATIVE AND QUANTITATIVE CHARACTERISTICS OF NATIVE AND NON-NATIVE SPEAKERS' WORD ASSOCIATION DOMAINS

By and large, the effects of the SW characteristics on the WA domain of L2 learners have received very little attention in L2 research. Previously conducted L2 WA studies have given little consideration to the likelihood that the characteristics of words themselves used as SW in WA tests impose limitations on the conclusions that can be drawn from WA data. Considering the different functional role that words from different lexical categories have, it is reasonable to hypothesize that the distribution of associative connections that language users build among words might also vary with the lexical class of the SW. It is also reasonable to hypothesize that frequency of occurrence of words might influence the quality and quantity of associative responses. Regrettably, while such effects have been well examined in the L1 literature, the impact of those characteristics has been largely overlooked in L2 WA studies. Describing differences between L1 and L2 WA domains without considering that some of those differences might be attributable to the word characteristics themselves, ignores a fundamental issue in the lexical acquisition process.

Thus, the focus of this analysis will be on whether or not characteristics of the SW, in particular lexical category and frequency of occurrence, affected the qualitative and quantitative features of the WA domain of NS and non-NS of English, at advanced and intermediate level of proficiency. It is important to know more about these effects because, on the one hand, better

understanding of how stimuli affect WA responses of NS and non-NS would raise researchers' awareness of item selection bias in L2 research. On the other hand, the effects of lexical category and word frequency have not been researched extensively in L2 studies, with the notable exception of Söderman (1993), who looked at the effects of high and low frequency words on the qualitative associative patterns of NS and advanced learners of English. Moreover, there is a good reason to believe that the choice of SW has a major impact on the type of WA responses and can explain to a large extent some "unexpected" results of WA analyses. Adult language users might not react the same way to high versus low frequency words, nor would they maintain the same pattern of responses to all lexical categories. In Chapter 4, I have examined issues related to the qualitative and quantitative patterns of associative responses of NS and L2 learners, focusing on their mental organization of lexical knowledge without regard to possible influences stemming from particular SW characteristics. The purpose of the present analysis would be to shed light on issues concerning the impact of lexical category and frequency of occurrence of SW on the features of the WA domain of NS and non-NS with regard to participants' level of proficiency and degree of familiarity with the stimuli.

Method

Participants

Fifty-one adult participants NS (n = 17), advanced learners (n = 17), and intermediate learners (n = 17) from the previous study participated in the current study.

Design

A 3 lexical category (nouns, verbs, adjectives) X 3 frequency band (low, mid, high) X 3 language proficiency group (NS, advanced learners, and intermediate learners) between-subjects design was used. The types of associations (paradigmatic, syntagmatic, phonological)

generated to 36 stimuli were treated as a random factor since variation in WA types across participants in each language group was the main dependent variable of interest. Quantitative features of associations such as the associative-response strength, the size of the WA domain, and the WA response idiosyncrasy were secondary dependent variables of interest.

Stimulus Selection Procedures

The analysis was based on associative responses generated to 36 SW, which were selected from the total of 73 stimuli used in the study. To balance for lexical category and frequency of occurrence of the stimuli, 12 SW were selected from each of the three lexical categories (nouns, verbs, adjectives), 4 from each frequency band (low, mid, high) (see Appendix M). There were only 12 adjectives in the sample, so they were all included for testing. The nouns and the verbs were randomly selected from the bigger sample, based on their frequency of occurrence. The procedure of dividing the SW into three frequency bands, the generation of responses, and acceptability of proof of SW familiarity were described in the previous chapter. I conducted separate analyses of the associations generated to SW participants knew, and associations produced to SW they vaguely knew, aiming at finding out whether familiarity and proficiency interact word frequency, and lexical category and gave effects on the qualitative and the quantitative features of the WA domain of the participants.

WA classification procedures

The classification of responses was based on the criteria specified in Chapter 4. Briefly, the qualitative characteristics of WA responses were based on the analysis of three types of responses: paradigmatic, syntagmatic, phonological (clang). An associative response was classified as paradigmatic if it shared same lexical class with the SW (e.g., *advantageous – profitable*) and as syntagmatic if it belonged to a different lexical class or could form a syntactic string with the SW (e.g., *weaken – old; pillar - salt*). The phonological (clang) category included responses related to a stimulus only in phonological terms without showing any semantic or syntactic relatedness to it (e.g., *beaten – bit*; *crampons - tampons*). As described in Chapter 4,

the quantitative features of the WA domain were examined by measures related to (a) the associative-response strength (reflected by response commonality and strength of the three most popular responses), (b) the size of the WA domain (reflected by total number of responses and number of different responses), and (c) the consistency of the WA domain (reflected by response idiosyncrasy).

Results

The relationship between language proficiency of the participants, lexical category, and frequency of occurrence of the SW was evaluated by conducting four 3 x 3 x 3 MANOVAs. The independent variables were language proficiency (NS, advanced, and intermediate learners), lexical category (nouns, verbs, and adjectives), and SW frequency (low, mid, high). The dependent variables were quality of connectivity of associations measured by the proportion of paradigmatic, syntagmatic, and phonological responses, as well as quantitative features of the WA domain measured by total number of responses, number of different responses, number of idiosyncratic responses, commonality of associations, and frequency of the three most popular responses. Separate analyses were conducted for stimuli participants knew and ones they vaguely knew. I will first present the results concerning the qualitative characteristics and then I will proceed with the results pertaining to the quantitative features of the WA responses.

Qualitative features of the WA domains

The qualitative characteristics of the WA domain for stimuli participants knew were examined by means of a 3 (groups) x 3 (lexical categories) x 3 (frequency bands) MANOVA with proportion of paradigmatic, syntagmatic, and clang associations as dependent variables. Means and standard deviations for the effects of word frequency and lexical category on participant's performance on the dependent variables are presented in Appendix N. The 3 lexical category X 3 frequency band X 3 language proficiency group MANOVA indicated a non-significant triple interactions (see Appendix O), significant interaction between group and lexical

category for participants' paradigmatic associations, $F(4, 81) = 3.56$, $p < .05$, partial $\eta^2 = .15$, but a non-significant interaction for their syntagmatic responses, $F(4, 81) = 1.10$, $p > .05$, partial $\eta^2 = .05$. There were also non-significant interactions between group and word frequency on participants' paradigmatic performance, $F(4, 81) = 1.87$, $p > .05$, partial $\eta^2 = .09$, and syntagmatic associations, $F(4, 81) = .76$, $p > .05$, partial $\eta^2 = .04$, as well as between lexical category and word frequency for paradigmatic, $F(4, 81) = .85$, $p > .05$, partial $\eta^2 = .04$, and syntagmatic associations, $F(4, 81) = 1.87$, $p > .05$, partial $\eta^2 = .08$. This showed that differences in the proportion of participants' syntagmatic responses did not depend on the combined effect of lexical category and frequency of occurrence of the SW in any significant way. In addition, frequency of occurrence of the SW did not interact with proficiency in the generation of paradigmatic responses. However, lexical category showed a significant effect on groups' proportion of paradigmatic responses. In other words, the relationship of paradigmatic responses to the lexical category of the stimuli showed noticeable dependency on participants' level of proficiency.

Because the interaction between proficiency and lexical category on participants' paradigmatic performance was significant, I chose to ignore the lexical category and group main effects and, instead, examine lexical category simple effects, i.e. the differences in the proportion of paradigmatic responses elicited by stimuli from the class of nouns, verbs, and adjectives for each group of participants. To control for Type I error across the three simple effects, alpha was set for each at .017. There were no significant differences in the generation of paradigmatic associations among the three groups when participants responded to nouns, $F(2, 33) = .65$, $p > .017$, and verbs, $F(2, 33) = 2.34$, $p > .017$. However, their performance significantly varied when they generated associations to adjectives, $F(2, 33) = 6.33$, $p > .05$. The pairwise comparisons, with alpha set at .006 to control for Type I error, revealed that when responding the adjectives, NS produced a much greater proportion of paradigmatic responses ($M = 53.7$, $SD = 26.1$) than advanced ($M = 25.1$, $SD = 13.3$) or intermediate learners ($M = 29.0$,

SD = 29.3) did. One implication of this finding is that choosing to use only adjectives as SW in a WA test would most probably result in findings about the structure of the mental lexicon of L2 learners that would be different from findings based on WA data collected for noun or verb stimuli. The results of the present analysis may have also some serious implications for SW selection as found in prior L1 research, which will be touched upon in the discussion section of this chapter.

Interestingly, SW frequency did not have a significant main effect on the proportion of elicited paradigmatic, $F(2, 81) = 1.87$, $p > .05$, partial $\eta^2 = .04$, and syntagmatic responses, $F(2, 81) = 2.94$, $p > .05$, partial $\eta^2 = .07$, across the three proficiency levels. The graph on Figure 8 shows a very surprising distribution of the proportion of the two response types.

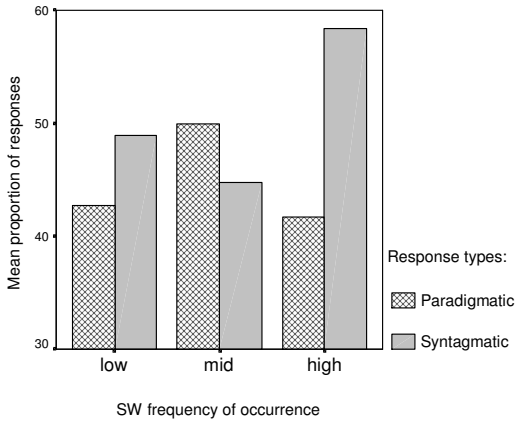


Figure 8. Effects of SW frequency on adult NS and non-NS' paradigmatic and syntagmatic responses to familiar stimuli.

Contrary to the expectation of paradigmatic - syntagmatic dominance across all frequency bands, the graph clearly shows that familiar words from the high and low frequency bands evoked mostly syntagmatic responses, whereas for the mid-frequency stimuli the participants maintained a fine balance between their paradigmatic and syntagmatic responses, with a preference for paradigmatic associative connections. I will later address this interesting

distribution in the context of using only high frequency SW in test samples and in light of discussions concerning findings about effects of word frequency in L1 and L2 research.

The analysis also pointed to a significant main effect of lexical category on the participants' proportion of syntagmatic associations, $F(2, 81) = 23.82$, $p = .000$, partial $\eta^2 = .37$, which indicated that there were differences in participants' proportion of syntagmatic responses resulting from the lexical category of the SW. Follow-up analyses of this main effect were further conducted in an attempt to identify the source of the differences in participants' syntagmatic performance. The follow-up tests consisted of all pairwise comparisons among the three types of lexical categories (nouns, verbs, and adjectives). The Bonferroni procedure was used to control for Type I error across the pairwise comparisons. The results indicated that for adults NS and non-NS, nouns tended to elicit a significantly smaller proportion of syntagmatic responses ($M = 28.2$) than verbs (95% CI = -51.144, -23.136, $p = .000$) and adjectives did (95% CI = -44.317, -16.309, $p = .000$). In addition, even though there were no significant differences between the proportion of syntagmatic responses generated to verbs and adjectives, verbs tended to attract a greater proportion of such responses ($M = 65.3$) than adjectives did ($M = 58.5$). Thus, the results further supported the conclusion that different lexical categories have specific effects on the type of associative connections adult language users develop, regardless of their language proficiency. Apparently, different lexical categories are selective to a particular type of associative connections, which was largely predicted by Clark's (1970) linguistic account of WA and Chomsky's (1965) proposal of selectional features of different lexical categories. That is, assuming that nouns do not have strong selectional features, according to Chomsky, we should expect that they would elicit much fewer syntagmatic responses compared to verbs and adjectives. In the main, the results of this analysis supported the aforementioned hypothesis.

The 3 lexical category X 3 frequency band X 3 language proficiency group MANOVA for the vaguely familiar SW showed that all interactions between the independent variables were not significant (see Appendix Q). Thus, differences among all participants' proportion of

paradigmatic, syntagmatic, and clang associations given to words they vaguely knew depended neither on a specific lexical category nor on a specific frequency of occurrence of the SW. Means and SD are presented in Appendix P.

There were only two significant main effects, a significant main effect of language proficiency on the proportion of syntagmatic responses generated, $F(2, 81) = 4.31$, $p = .000$, $\eta^2 = .10$, which confirmed the findings from the analysis of the larger data set, and a significant main effect for lexical category on participants' syntagmatic associations, $F(2, 81) = 11.39$, $p = .000$, $\eta^2 = .22$. Figure 9 shows the distribution of all associative responses of which only the syntagmatic responses were statistically different across the three lexical categories.

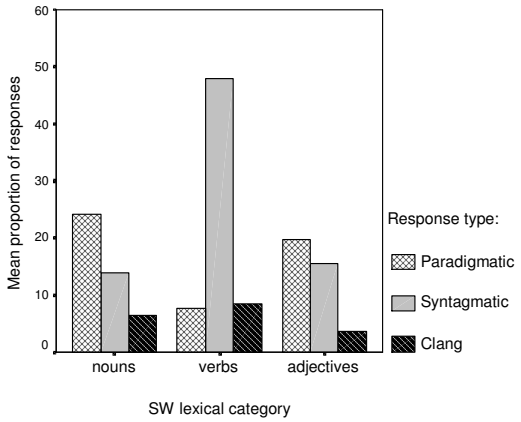


Figure 9. Effects of SW lexical category on the associative responses of adult NS and non-NS to vaguely familiar stimuli.

The pairwise comparisons of the syntagmatic associations, set at .05 alpha level of significance to control for Type I error, revealed that verbs attracted a significantly greater proportion of syntagmatic responses ($M = 47.9$) than nouns ($M = 13.9$) or adjectives did (15.6). Thus, it would be safe to say that for the vaguely familiar words the different lexical categories preserved their selectional characteristics to specific types of associative responses which they showed to have for familiar stimuli, with verbs being predominantly syntagmatic, and nouns being mostly paradigmatic. Another immediately noticeable effect of the interaction between a lower degree

of familiarity and lexical category was the predominance of paradigmatic associations over the syntagmatic for the category of adjectives. Let me remind here that the misinterpretations of the SW, which rendered them vaguely familiar, were mostly meaning-based (not lexical class confusions). Also, in the context of the present analysis the SW frequency of occurrence did not have any impact on participants' performance because there were no significant interactions between any of the three independent variables. Thus, while NS maintained paradigmatic dominance when responding to both familiar and vaguely familiar adjectives, non-NS showed a tendency to elicit more syntagmatic responses (58% on average) than paradigmatic (27.5% on average) for familiar adjectives, which trend reversed when they responded to vaguely familiar adjectives. The lower degree of familiarity evoked responses that more closely resembled NS' associative behavior in that they generated more paradigmatic (20% on average) than syntagmatic responses (15.5% on average).

As for the impact of frequency of occurrence of the SW on the proportion of response types, the three groups responded in a very similar way to the 36 stimuli (see Appendix Q). As shown in Figure 10, the high and the low frequency vaguely familiar stimuli still attracted more syntagmatic than paradigmatic responses, as they did for the familiar stimuli. However the mid frequency SW this time had a syntagmatic bent, non-significant though, which was not the case with familiar stimuli. As expected, clang associations decreased as frequency of occurrence increased, which was especially noticeable with the high frequency SW, though their proportion was very small anyway. On the whole, for the three groups of participants, the mid frequency stimuli tended to elicit a balanced ratio of paradigmatic (26% on average) and syntagmatic responses (29% on average), whereas the low and high frequency SW evoked approximately twice as more syntagmatic than paradigmatic responses.

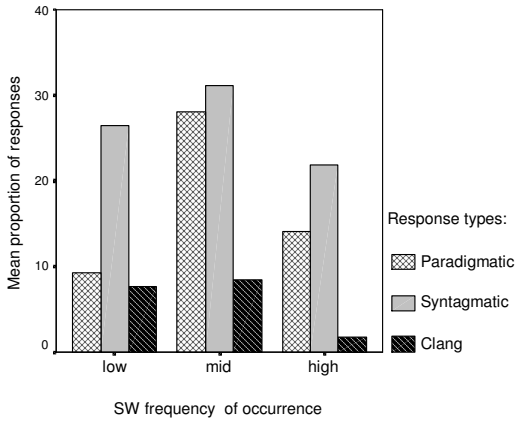


Figure 10. Effects of SW frequency on adult NS and non-NS' associative responses to vaguely familiar stimuli.

Summary of key findings with regards to effects of lexical category and word frequency on the qualitative features of participants' WA domains.

The analysis of the qualitative features of participants' WA domains revealed that for familiar stimuli, the level of proficiency strongly interacted with the lexical category of the SW and had a significant effect on participants' paradigmatic associations. Further investigation into the matter showed that groups differed in their production of same class associations when they responded to adjectives, but not when they responding to nouns or verbs. In particular, NS produced twice the proportion of paradigmatic responses to adjectives than non-NS did. This is a very interesting finding because the variation in the response pattern to familiar stimuli can be attributed neither to a particular "developmental" stage through which learners' vocabularies are going, especially favored in L2 WA research when syntagmatic dominance is detected, nor to frequency of occurrence of stimuli, because frequency did not interact with language proficiency. Therefore, I would argue that for adjectives non-NS preferably arrange their associative connections along syntagmatic lines and this organization is independent of an increase in level of proficiency or frequency of occurrence of this specific lexical class. Consequently, choosing to use only adjectives as SW in WA tests would inevitably result in

findings about differences in the structure of the mental lexicon of L2 learners and NS, which, as the results of this study show, apply only to this lexical class. For all participants, verbs were predominantly syntagmatic and elicited on average 65% syntagmatic associations, while nouns were the exact opposite and evoked on average 72% paradigmatic responses. Given this distribution of responses to the various lexical classes and the fact that non-NS speakers tend to generate syntagmatic responses to adjectives, it is not surprising at all why non-NS have been found to maintain a predominantly syntagmatic pattern of responses with lexically balanced test samples (e.g., Piper & Leicester, 1980; Wolter, 2001). However, it is surprising that the syntagmatic dominance has been attributed to lack of stability of L2 associative connections or explained away by positing lower stages in the development of their lexical organization. Similarly, it is not hard to see the implicit, sometimes even explicit, comparison between L1 "child-like" patterning and "non-nativelike" lexical organization showing through such explanations, which as an analogy confines interpretations of L2 WA data to findings relevant to the acquisition of English as a native language. In the main, my analysis completely supports findings from L1 (e.g., Deese, 1962; Fillenbaum & Jones, 1965) arguing that NS of English maintain an overwhelming paradigmatic dominance for nouns (68 %), paradigmatic preference for adjectives (53 %) and an overwhelming syntagmatic dominance for verbs (67 %) for words they know. Once again I will distance myself from the much favored "childlike/ non-nativelike" metaphor in L2 research and argue that non-NS, especially after they have accumulated vocabulary over 4,000 words, simply maintain a slightly different pattern of associative connections than NS do. Non-NS, or at least Bulgarian L2 learners of English, maintain the same type of connections as NS do for nouns and verbs. However, they preferably develop syntagmatic connections between adjectives and other lexical categories, which does not mean that their adjectival links are underdeveloped or unstable.

Interestingly, Piper and Leicester (1980) obtained very similar results in their study of advanced and beginning Japanese L2 learners in comparison to NS of English. Their test

sample was based on a high frequency SW list which contained 24 stimuli, 8 from each of the three lexical categories, which makes it a lexically balanced sample. The researchers found that their advanced learners generated on average 25% paradigmatic responses to adjectives (25% in this study), 34% to verbs (29% in my study), and 66% to nouns (69% in this study). Quite surprisingly, these results were explained in their overall discussion in light of the developmental shift toward paradigmatic responding in word association test found in L1 but in reference to their beginning learners only, who elicited an even lower proportion of paradigmatic responses to adjectives (16%) than the advanced learners did. Unfortunately, this is the only L2 study that reports results about L2 learners' proportions of associations with regard to the lexical category of the SW and participants' proficiency level, but the results of my study largely agree with the results of Piper and Leicester's study regarding the associations of advanced Japanese learners.

This allows for hypothesizing that the structure of the mental lexicon of NS and non-NS of English differs qualitatively in the way they organize their connections for adjectives but not for nouns and verbs. I would suggest two possible reasons for this peculiar behavior and, in my view, each of them deserves to be further investigated. First, it seems reasonable to suggest that the dissimilarity in the way NS and non-NS attend to adjectives can be attributed to the non-NS being influenced by the response mode of their mother tongue. Piper and Leicester (1980) hinted at this possibility but then hastily dismissed it. In Japanese, true adjectives of native Japanese origin are often considered to be a special type of verbs and there is a considerable overlap between the two lexical classes. For example, true Japanese adjectives have tenses and moods and the negative form of verbs is considered to be adjectival in origin (Bleiler, 1963). Therefore, the hypothesis that Japanese L2 learners might maintain verb-like associations (i.e. mostly syntagmatic) for the adjectives in the new language should not be dismissed without being experimentally rejected. Similarly, it would be interesting to further investigate this hypothesis by comparing response patterns of NS of Bulgarian with ones of NS of English for

pairs of translated stimuli. If, for example, the response mode of Bulgarian L2 learners of English resembles more closely the qualitative response pattern of monolingual Bulgarian speakers, this can be taken as evidence that the native language knowledge of the syntactic properties of words influences the associative connections among words in the L2. This possibility should not be excluded at all because in Bulgarian, for example, there is a strong semantic relationship between adjectives and nouns which is realized in a syntactic dependence of adjectives on the nouns they pre-modify. That is, nouns can be pre-modified only by adjectives, which agree in person, number, gender, and definiteness with the head of the NP. Moreover, unlike English, the class of nouns in Bulgarian does not share the syntactic property of adjectives to pre-modify other nouns, which draws clear lines between the syntactic properties and the syntactic selectional features of nouns, verbs, and adjectives. Therefore, it is quite possible that L2 learners attend to adjectival associative links in English as they would do in Bulgarian. Consequently, future research should examine how characteristics of the L1 itself may continue to influence the qualitative characteristics of the associations of the L2 lexicon in its advanced stages of development.

This hypothesis is largely supported by research concerned with the relationships between L1 and L2 lexicons, which has consistently found that a presentation of a stimulus in one of a bilingual's languages primes a response in the other language as well (e.g., Kirsner, Smith, Lockhart & Jain, 1984; Schwanenflugel & Rey, 1984). Likewise, when a person has a good command of two languages, lexical items are activated in both languages with those in the language that is not required being suppressed (Green, 1998). In addition to this, researchers are becoming increasingly aware that certain task-specific variables are likely to trigger specific language processing strategies on the part of L2 learners and that the nature of the task may hold important answers to processing issues (e.g., Schwanenflugel & Gavisk, 2002; Green, 1998). Considering these arguments, it would be reasonable to suggest that the lexical category of words plays a much greater role in the L2 mental organization of lexical knowledge than

previously acknowledged. Therefore, drawing conclusions about the qualitative features of the L2 mental lexicon based on research findings concerning the L1 and disregarding features of the lexical input is a largely unfounded approach that should be treated with caution.

L2 learners' predominance of syntagmatic responses for adjectives might well be linguistically motivated by the L2 itself. As mentioned in Chapter 4, Clark (1970) argued that paradigmatic and syntagmatic responses are essentially motivated by one and the same principle - the principle of simplicity of production, which states "perform the least change on the lowest feature, with the restriction that the result must correspond to an English word" (p. 280). Consequently, it is highly possible that for L2 learners "the least change" operation for adjectives more often entails a completion of a syntactic string in which they realize the selectional features of the adjectives rather than apply paradigmatic rules, i.e. the minimal contrast rule (generating a paradigmatic response that forms a minimal contrast pair with the stimulus, e.g. *long* – *short*), the marking rule (producing the unmarked form to a marked stimulus, e.g., *better* – *good*), the feature deletion and addition rule (giving responses in which a single feature is deleted or added, e.g. teach – learn, odour – smell), and finally, the category preservation rule (e.g., *speculate* – *consider*). For example, for the SW *amoral*, the most common response that the non-NS produced was *behavior* and it accounted for 14% of the total number of responses elicited by the stimulus, whereas the most commonly produced by the NS response was *moral*, which accounted for 18% of their total number of responses. The example clearly shows that the NS in this study preferably applied the minimal contrast rule, in Clark's terns, and gave a response that formed a minimal contrast pair with the stimulus, while the non-NS seemed to prefer the idiom completion rule, i.e. produce a word with which the stimulus forms a collocation or an idiom. Why non-NS do this and NS do not is a valid question that, regrettably, cannot be answered by this study.

The vaguely familiar words did not seem to be influenced by participants' level of proficiency, yet, lexical category again showed more distinctive effects than frequency of

occurrence. By and large, the different lexical categories preserved their selectional powers to specific types of associative responses with verbs being primarily syntagmatic, and adjectives and nouns being mostly paradigmatic. One of the immediately noticeable effects of the interaction between the lower degree of familiarity and lexical category for all participants was the predominance of paradigmatic associations over the syntagmatic for the category of adjectives. Thus, while NS maintained paradigmatic dominance when responding to both familiar and vaguely familiar adjectives, non-NS showed a clear tendency to elicit more syntagmatic responses for familiar adjectives but this trend reversed when they responded to vaguely familiar adjectives. So, the results showed that when non-NS had a lower level of familiarity, their response pattern closely resembled NS' associative behavior, which suggests that familiarity rather than proficiency promotes type-specific associations. Such an associative behavior also implies that NS and non-NS at higher level of proficiency (advanced and intermediate) establish similar patterns of connection of words in their mental lexicon. Yet, a high degree of familiarity encourages access to different type of associations for some lexical categories, in this case – adjectives, for NS and non-NS.

As for the impact of frequency of occurrence of the SW on the proportion of response types, this SW feature did not reveal any specific effects that would distinguish NS from non-NS in their response pattern. The three groups responded in a very similar way to both familiar and vaguely familiar words from various frequency bands. They maintained a clear syntagmatic preference across the three frequency bands for the SW they vaguely knew. For the ones they knew, stimuli with high and the low frequency of occurrence still attracted more syntagmatic than paradigmatic responses, while the mid frequency SW had a paradigmatic bent. Looking comparatively at the response pattern for the two different degrees of familiarity, the most striking result was that, regardless of degree of familiarity, the high and the low frequency words were overwhelmingly syntagmatic, while the mid frequency stimuli elicited relatively balanced proportions of syntagmatic and paradigmatic associations (see Table 8).

Table 8

Proportions of response types for familiar and vaguely familiar words

Stimulus type	Low frequency	Mid frequency	High frequency
Familiar words:			
• paradigmatic	42%	49%	41%
• syntagmatic	49%	45%	58%
• clang	-	-	-
Vaguely familiar words:			
• paradigmatic	10%	26%	14%
• syntagmatic	25%	29%	23%
• clang	9%	8%	2%

It seems, then, that the relative difficulty of the lexical task would influence more the type of responses than the level of proficiency of language users or their degree of familiarity with lexical items would. Unfortunately, there is only one L2 WA study on the effects of frequency of occurrence in the field and my results only partially support its findings. Söderman (1993) compared the WA of two groups of language users, NS and advanced learners, on frequent and infrequent stimuli. She found that level of proficiency did not play much of a role when her subjects responded to either word frequency. The results of my study fully support her conclusion, adding to the subject population intermediate L2 learners as well. She also noticed that some of her subjects, both NS and non-NS, tended to respond syntagmatically to both frequencies, which was contrary to her expectation that high frequency words would elicit primarily paradigmatic responses, while low frequency stimuli would promote mainly syntagmatic associations. In fact, the results of this study do not support a hypothesis such as the one proposed by Söderman (1993) because the participants in this study, NS and non-NS as a group, responded overwhelmingly along syntagmatic lines for the high frequency SW. My

results are more in line with Stolz and Tiffany (1972) who also found that their group of L1 users responded syntagmatically to high frequency words. The authors hypothesized that the syntagmatic rate stabilizes earlier in the word learning process than other types of responses. Therefore, adults responding syntagmatically to high frequency words "could be an indication that they represent information learned relatively early in a person's experience with a word" (p.44). I would take this hypothesis a step further by suggesting that not only do language users become well acquainted with high frequency words early in their language experience, but they also experience them in a much greater variety of contexts than mid or low frequency words. Also, high frequency words tend to have more different meanings than low frequency words, which makes them suitable for use in diverse situations. For example, one of the high frequency nouns used as a SW in the test was *studio* and there are five meanings listed under this entry in *Longman Dictionary of English Language and Culture* (1992), compared with one of the low frequency nouns in the list *cassava*, which has only one listed meaning in the same dictionary. Naturally, the more varied semantic content a word has the less restricted its use is to a particular context, which in turn encourages its use in a syntagmatic string. Low frequency words, on the other hand, have one or a couple of meanings at most, very few synonyms that can be given as paradigmatic responses; so, quite naturally, they also evoke more context-situated responses of syntagmatic nature. The mid frequency words showed a fine balance between the two types of responses, which most probably is a consequence of their moderate frequency of use and occurrence or their later integration into the lexicon. Most probably, these two conditions encourage both a frequent synonym match (i.e. a paradigmatic response) and a context-based association (i.e. a syntagmatic response). I am aware that the hypotheses I have just proposed offer only a general explanation of the syntagmatic variation of responses. This only comes to emphasize once again the difficulty of the task to explain the associative behavior of adults in a simple way, especially when studying and comparing adult NS with non-NS.

On a final note, I would say that the assumption that the highest level of lexical knowledge is represented by a paradigmatic response for both NS and non-NS is largely challenged by the results of this study. Therefore, the notion of a "paradigmatic shift" as an indication of a higher degree of lexical knowledge or a better organization of mental lexicon connections should be reconsidered in light of findings which strongly suggest that different lexical categories promote different types of connections in the mental lexicon of adult NS and non-NS alike.

Quantitative features of the word association domains

The 3 lexical category X 3 frequency band X 3 language proficiency group MANOVA on the quantitative features of the WA produced to familiar SW by NS and non-NS indicated non-significant triple interaction between lexical category, frequency band, and language proficiency group (see Appendix R). However, the interaction between SW frequency of occurrence and SW lexical category for number of different responses, $F(2, 81) = 4.32$, $p < .05$, partial $\eta^2 = .18$, degree of commonality, $F(2, 81) = 3.04$, $p < .05$, partial $\eta^2 = .13$, and idiosyncrasy of associations, $F(2, 81) = 5078$, $p < .05$, partial $\eta^2 = .22$ was significant, which showed that the combined effect of certain levels of the two SW characteristics influenced the production of different, common, and idiosyncratic responses, independent of proficiency level of the participants. Therefore, I chose to examine the simple effect of SW lexical category at each level of SW frequency for the three quantitative measures. To control for Type I error across the three simple lexical class effects, the alpha level for each was set at .017. The analysis revealed that for the low frequency SW there were no significant differences among participants' responses to stimuli from the three lexical categories for number of different responses, $F(2, 33) = 1.43$, $p > .017$; commonality of responses, $F(2, 33) = .06$, $p > .017$, and number of idiosyncratic associations, $F(2, 33) = 1.71$, $p > .017$. Overall, the low frequency stimuli evoked twice as many different and idiosyncratic responses than common responses, which suggested that it was the low frequency of occurrence rather than the lexical category of the stimuli that

affected participants' performance on the three quantitative measures. However, for the mid frequency SW, the lexical class of the SW had a significant effect for number of different responses, $F(2, 33) = 9.80$, $p \leq .017$; number of commonly produced associations, $F(2, 33) = 6.52$, $p \leq .0.17$, and number of idiosyncratic responses, $F(2, 33) = 7.29$, $p \leq .017$. The follow-up tests, with alpha set at .006 (.017 / 3) to control for Type I error over the three pairwise comparisons, revealed that mid frequency adjectives elicited a significantly greater number of different ($M = 25.8$, $SD = 9.9$) and idiosyncratic responses ($M = 21.9$, $SD = 8.3$) than mid frequency verbs did ($M = 12.8$, $SD = 10.5$ and $M = 11.5$, $SD = 9.1$, respectively). At the same time, mid frequency adjectives elicited a significantly greater number of common responses ($M = 10.9$, $SD = 8.0$) than mid frequency verbs did ($M = 3.0$, $SD = 3.7$). Mid frequency nouns tended to elicit a non-significantly smaller number of different ($M = 16.2$, $SD = 9.8$), idiosyncratic ($M = 13.6$, $SD = 9.2$), and common responses ($M = 9.9$, $SD = 10.8$). High frequency stimuli showed a significant main effect for number of commonly generated associations, $F(2, 33) = 13.88$, $p \leq .017$, and number of idiosyncratic associations, $F(2, 33) = 6.01$, $p \leq .017$. There were no significant differences in the number of different associations generated to the three lexical categories, $F(2, 33) = 3.32$, $p \leq .017$. The pairwise comparisons clearly pointed to the cause of the significant effect for lexical category. High frequency nouns evoked a significantly greater number of common responses ($M = 25.4$, $SD = 7.5$) than high frequency verbs did ($M = 12.9$, $SD = 6.7$). Similarly, since the number of idiosyncratic responses seemed to be inversely related to the number of commonly generated associations, it was logical to expect that participants would generate a significantly smaller number of idiosyncratic associations to high frequency nouns ($M = 15.8$, $SD = 5.5$) than they would to high frequency verbs ($M = 24.7$, $SD = 7.5$) and adjectives ($M = 24.3$, $SD = 4.3$).

Next, since the interactions between lexical category and word frequency were not significant for total number of responses and the strength of the three most popular associations the participants generated to the 12 SW I examined the main effects of SW frequency and SW

lexical class on the two variables, which were significant (see Appendix R). The follow-up analyses to these main effects were conducted to identify the source of variation in participants' associative behavior. The follow-up tests consisted of all pairwise comparisons among the three types of lexical categories (nouns, verbs, and adjectives), as well as the three frequency bands (low, mid, high) for two quantitative measures, i.e. total number of responses and strength of the three most popular responses. The Bonferroni procedure was used to control for Type I error across the pairwise comparisons. As expected, frequency of occurrence of the SW was directly related to total number of associations and frequency of occurrence of the three most popular responses, i.e. the lower the SW frequency, the lower the total number of responses, the lower the frequency of occurrence of the three most common responses (see Figure 11).

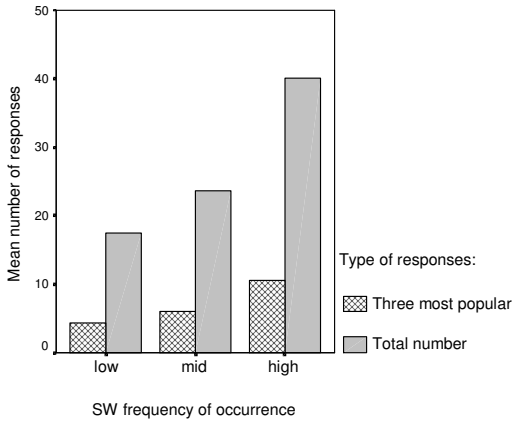


Figure 11. Effects of SW frequency on adult NS and non-NS' total number of responses and frequency of the three most popular associations generated to familiar stimuli.

Nonetheless, for both types of quantitative measures only the differences between the number of associations elicited by the low frequency SW in comparison to the number of responses generated to mid and high frequency stimuli were significant. In other words, high and mid frequency stimuli tended to elicit a much greater total number of responses compared to low frequency SW. In addition, the strength of the three most popular associations was much

greater for high and mid frequency words than for the low frequency stimuli, which as a whole evoked a small number of responses that were mostly idiosyncratic.

The lexical category of the stimuli also proved to have a significant main effect on the two quantitative measures (see Figure 12). Quite surprisingly, considering the nature of

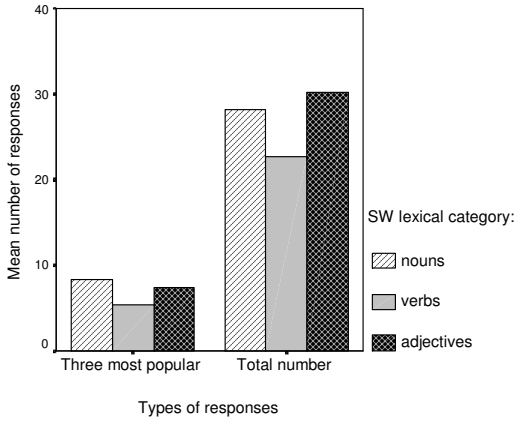


Figure 12. Effects of SW lexical category on adult NS and non-NS' total number of responses and frequency of the three most popular associations generated to familiar stimuli.

selectional features of verbs, this category evoked on average a significantly smaller number of associations (M = 22.6, SD = 1.8) than nouns (M = 28.2, SD = 1.8) and adjectives did (M = 30.1, SD = 1.8), as well as a lower frequency of occurrence of the three most popular responses (M = 5.3, SD = .8) than nouns (M = 8.3, SD = .8) and adjectives (M = 7.4, SD = .8). Overall, all participants seemed to produce more readily associations to nouns and adjectives than they did to verbs. Moreover, verbs tended to elicit quite infrequently three common responses, which suggested that using verbs as SW to investigate the strength of the associative domain of language users would yield significantly different results compared to test instruments that use adjectives, nouns, or both as stimuli.

The effect of proficiency was not of particular interest to this analysis because it was examined and discussed in detail in Chapter 4. I would only mention that the analysis based on

this balanced, with respect to lexical category and frequency of occurrence, set of 36 SW confirmed the results of the previous analysis, which was based on a less balanced sample of 73 stimuli.

As far as responses to vaguely familiar words were concerned, I examined only total number of responses generated to the 36 SW used in this analysis because the participants did not produce any common responses. Also, the number of different and idiosyncratic responses was the same as the total number of responses. Means and standard deviations are presented in Appendix S. The 3 language proficiency group X 3 frequency band X 3 lexical category MANOVA indicated non-significant interactions between proficiency and SW frequency, F (2, 81) = 1.15, p >.05, partial η^2 = .05; proficiency and SW lexical category, F (2, 81) = 1.09, p >.05, partial η^2 = .05, SW frequency and lexical category F (2, 81) = 2.30, p >.05, partial η^2 = .10, and the combined effect of proficiency, SW frequency, and SW lexical class, F(8, 81) = .36, p >.05, partial η^2 = .04. Since the primary purpose of this analysis was to examine the effects of the characteristics of the SW, the group main effect, F(2, 81) = 6.00, p <.05, partial η^2 = .13 was not of interest in this analysis, even though it showed to be significant. Instead, I focused on the significant main effect of SW frequency, F (2, 81) = 5.62, p <.05, partial η^2 = .12, and SW lexical category, F (2, 81) = 4.35, p < .05, partial η^2 = .10, on participants' number of responses to vaguely familiar SW. The pairwise comparisons among the three lexical categories and three frequency bands were controlled by using the Bonferroni procedure for control of Type I error. The results indicated that the participants more readily responded to vaguely familiar verbs than to adjectives (95% CI = -3.759, -.230, p = .021). They also generated a much greater number of responses to vaguely familiar SW from the mid frequency (M = 3.5) than to the high frequency band (M = 1.2). Overall, the high frequency vaguely familiar SW attracted the smallest number of responses, as did the low frequency vaguely familiar stimuli (M = 1.9). Apparently, mid frequency vaguely familiar words were the

source of greatest confusion among all participants just by virtue of their frequency of occurrence.

Summary of key findings with regards to impact of lexical category and word frequency on quantitative features of participants' WA domains.

Unlike the qualitative features of the WA domain, the quantitative characteristics of participants' WA were significantly influenced by the combined effect of the SW frequency of occurrence and lexical class. Although participants' total number of responses and strength of the three commonly given associations did not seem to be greatly affected by the two SW characteristics, their total number of responses, degree of commonality, and number of idiosyncratic associations were seriously influenced. Overall, the low frequency stimuli evoked twice as more different and idiosyncratic responses across the three lexical classes than common responses, which suggested that for these words it is the low frequency of occurrence rather than the lexical category of the stimuli that affects participants' performance on the three quantitative measures. By and large, when the stimuli were matched on both stimulus characteristics, as their frequency of occurrence increased the two variables started to work cooperatively for the mid and high frequency words. Interestingly, for the mid frequency words, the adjectives consistently attracted considerably more different, common, and idiosyncratic associations than verbs did. Nouns did not show to be significantly different from the adjectives, yet, they evoked a smaller number of responses across the three quantitative measures. The result suggests that adjectives seem to be the lexical class that most readily elicits both common and idiosyncratic responses. On the other hand, participants seemed to respond less willingly to verbs, which elicited on average a very small number of common associations (M = 3) and a relatively high number of idiosyncratic (M = 11.5) and different associations (M = 12.8), yet twice lower than adjectives. Not surprisingly, as frequency of occurrence increased, the number of responses generated to the three quantitative measures also increased and the analysis revealed that high frequency nouns were more often associated with common to

participants' things and ideas than verbs or adjectives were. Respectively, verbs and adjectives evoked more diverse and more idiosyncratic associative connections. Overall, as frequency of occurrence of words increased, the heterogeneity of the associative connections also systematically increased across the three lexical categories. However, the strength of the associative domain, as measured by each group's commonality of associations, showed a clear dependency on lexical category, with nouns clustering together more readily than adjectives, especially high frequency nouns, followed by adjectives and verbs.

Total number of responses and the availability of the three most common responses were also influenced by frequency of occurrence and lexical category of the stimuli but not by their combined effect. Quite expectedly, frequency of occurrence was directly related to both measures, i.e. the lower the SW frequency, the lower the total number of responses and the frequency of the three most common responses. While it was not surprising that low frequency stimuli evoked a small number of responses, I found it interesting that they elicited on average a relatively high proportion of primary (the three most common) responses. One would expect that since these words have a restricted semantic content that allows them to be experienced in a relatively restricted context, the strength of the three top responses would be much greater than for high or mid frequency words, for example. However, it was found that the proportion of the three commonly given associations was almost the same across the three frequency bands, i.e. 25% for low, 26% for mid, and 27% for high frequency stimuli. This suggests that there is no reason to expect that high or medium frequency stimuli would elicit the highest degree of commonality of primary responses, as traditionally accepted (insofar as proportion of responses , not absolute number, is used to measure the strength of the associative domain. When commonality is measured by the number of times a particular association is given to a particular stimulus, then, it seems that low frequency words promote lesser commonality than higher frequency ones. As a matter of fact, this finding is in concert with the findings of several L1 WA researchers, who noticed that the greater the frequency and familiarity with the SW, the more

difficult it can be to maintain a common response set. For some time, this relationship was treated as a paradox in the L1 literature because it was traditionally accepted that response strength and response variability were inversely related, until subsequent research found that the relationship primarily depended on the model used to elicit associations. For example, a model that asked participants to produce more than one association per stimulus allowed for a response competition that, consequently, increased the chances of strong responses to be better represented than weak ones. On the contrary, a model based on one response per SW would be more likely to yield a greater variety of responses and a lower degree of commonality, which raised researchers' awareness that the model of response elicitation could greatly influence the quantitative characteristics of the WA domain. The present study, which allowed for response competition, completely agrees with this conclusion and points to a very similar magnitude of strength of the primary responses across the three frequency bands, when it is accounted for by proportion of response commonality rather than by absolute number of commonly generated associative responses.

The lexical category of the stimuli also had distinguishable effects on the two quantitative measures. Interestingly, considering the selectional nature of verbs, this category evoked on average a much smaller number of associations, as well as a lower frequency of occurrence of the three most popular responses than nouns and adjectives. Overall, participants generated more readily associations to nouns and adjectives than they did to verbs. Moreover, the fact that verbs seldom promoted three common responses suggests that using verbs as SW to investigate the strength of the associative domain of language users would yield substantially different results compared to test instruments that employ adjectives, nouns, or both.

The analysis of the responses to vaguely familiar words did not turn out to be of much value mostly because a small number of the words were responded as vaguely familiar. It was found that the participants more readily responded to vaguely familiar verbs this time than to adjectives. They also generated a much greater number of responses to mid frequency than to

high frequency stimuli. Overall, the high frequency SW attracted the smallest number of responses, as did the low frequency stimuli. Apparently, mid frequency words were the source of greatest confusion among all participants, regardless of their proficiency level. In general, examining the qualitative features of the WA domain for the vaguely familiar words did not turn out to be revealing of any consistent associative behavior that would distinguish between participant's level of proficiency or associative habits. The analysis only showed that participants were well aware of their level of familiarity with low and high frequency words. Yet, the medium frequency elicited the greatest number of responses, which suggested that these words were more often recognized as familiar while they were actually not.

Conclusion

One of the greatest advantages of using WA tests in studying the organization of lexical knowledge is that they are fairly easy to use and, at the same time, have the potential to generate a wealth of data. However, as pointed elsewhere in the literature (e.g., Meara, 1982), little consideration has been given to SW selection though L2 researchers seem to be intuitively aware that certain characteristics of words have specific effects on the semantic organization of learners' lexical knowledge. Unfortunately, the lack of sufficient empirical research and methodology of using WA tests in L2 research have taken us little in the way of better understanding whether the bilingual mind responds differently to words from different lexical categories and frequency bands or whether there is a response pattern that governs the overall organization of lexical items. More importantly, when non-NS are compared to NS, it is of paramount significance to know whether generalizations concerning the differences between the two broadly defined groups pertain to fundamental differences in the way the monolingual and the bilingual mind builds semantic connections or whether certain characteristics of words result in patterns of associative links that are more specific to L2 learners and not so typical of NS.

By and large, the issue of principled considerations in SW selection, though raised in the literature (e.g. Meara, 1982, Wolter, 2002), has been largely overlooked by L2 researchers. Many researchers select their SW either by compiling idiosyncratic lists of prompt words with little or no discussion of why and how these words have been chosen or by using some standard lists of stimuli, the most frequently employed being Kent and Rosanoff list. This list was compiled in 1910 for a study of the WA produced by mentally ill subjects. The list consists of 100 items, mostly high frequency adjectives and few nouns that elicit relatively stable response patterns in normal NS of English. The extensive use of this list in both L1 and L2 research can most probably be explained by the very large number of normative associative responses collected from large groups of NS and not so much by any theoretical considerations of SW selection.

I touched upon some of the disadvantages of using standard lists in WA testing in Chapter 4, but in the context of the present discussion I will mention that the inclusion of high frequency words only in test samples to ensure familiarity is probably the greatest disadvantage of using standard lists. The reason is that high frequency words tend to elicit very similar responses in the native language as they do in the L2 (Meara, 1982) which in turn calls into question the value of any generalizations based on WA data elicited by means of such stimuli. Rozenzweig (1961), for example, used translations of Kent and Rosanoff's list to compare the association responses among speakers of English, French, German, and Italian. Not surprisingly, he found that the agreement among the four languages was impressive. Almost half of the comparisons in any pair of languages yielded same responses, which made him conclude that the cross-cultural community, at least the one of the Western European languages, largely shared associative connections and meanings, even though verbal forms were different (p. 357). In light of this conclusion it is fair to say, then, that using high frequency stimuli in L2 lexical research may easily obscure the effects of L2 learning and, instead, emphasize the effects of cross-linguistic influences. Sampling stimuli from only one frequency

band and one lexical category is another very common practice in L2 WA research. It should be noted that this practice severely restricts the conclusions of such experiments to generalizations concerning a limited part of L2 learners' vocabulary, which may or may not hold true for the whole lexicon. Therefore, knowing little about the effects of SW on the mental organization of lexical knowledge, in my view, takes us no further than simply identifying differences between NS and L2 learners without accounting for features pertaining to the input that promote them.

The analysis discussed in this chapter was conducted to shed some light on the issues outlined above. It focused on examining how lexical category and frequency of occurrence of the SW affect the qualitative and quantitative features of NS' and non-NS' WA domains. The analysis was based on associative responses generated to 36 SW, which were randomly selected from the bigger sample of stimuli and balanced for lexical category and frequency of occurrence. Previously conducted L2 WA research gave little consideration to the likelihood that word frequency or lexical category might influence associative behavior. This line of analysis proved the importance of careful consideration of SW selection and raised questions concerning the special effects of the lexical input on the lexical association process. The results confirmed the hypothesis that features of the lexical input have distinctive effects on the qualitative and quantitative characteristics of the connectivity among lexical items in the L1 and L2 mental lexicon. The analysis of the qualitative features of participants' WA domains revealed that for words participants knew, they built distinctive patterns of associations, i.e. verbs were responded predominantly syntagmatically, nouns -- paradigmatically, while the response type to adjectives was group specific. Participants significantly differed in their production of paradigmatic associations when they responded to adjectives only, more specifically L2 learners preferably associated adjectives with other lexical categories, while NS preferably maintained same class meaning connections. This is a very interesting finding because it not only challenges again the much favored by some L2 researchers explanation of a "developmental" stage through which L2 learners' vocabularies are going anytime a syntagmatic dominance is

detected, but it also points to the lexical class that triggers specific associative connections for the L2 learners. Consequently, choosing to use only adjectives as SW in WA tests, or even choosing to use a balanced for lexical category SW sample, would inevitably result in findings about qualitative differences in the structure of the mental lexicon of L2 learners and NS, which, as this study shows, pertain only to one particular lexical class, not to the whole lexicon. Moreover, deciding to use only high frequency SW to ensure familiarity would further add to the syntagmatic bias of L2 WA analysis results because, as shown by the results of this study, high frequency words tend to promote overwhelmingly syntagmatic meaning connections (as well as low frequency words, but for different reasons), while mid frequency stimuli elicit relatively balanced proportions of syntagmatic and paradigmatic associations. Therefore, I would recommend using mid frequency stimuli with lexically balanced WA test samples, instead of high frequency prompt words, because frequency of occurrence will neutralize the effects of lexical category and more realistically reflect whether qualitative differences among subjects actually exist. On a final note, the assumption that the highest level of lexical knowledge is represented by a paradigmatic response for both NS and non-NS was also largely challenged by the results of this study. Consequently, the notion of a "paradigmatic shift" as an indication of a higher degree of lexical knowledge or a better organization of mental lexicon connections should be reconsidered in light of findings which convincingly suggest that different lexical categories and frequency bands promote different types of connections in the mental lexicon of adult NS and non-NS alike.

The quantitative characteristics of participants' WA domains, in particular degree of commonality and number of different and idiosyncratic associations, were also seriously influenced by the combined effect of the SW frequency of occurrence and lexical class. Overall, when the stimuli were matched on both stimulus characteristics, as their frequency of occurrence increased the two variables started to work cooperatively for the mid and high frequency words, and not so much for the low frequency words, following a very specific pattern.

That is, the mid frequency adjectives and nouns consistently attracted considerably more different, common, and idiosyncratic associations than verbs did, yet, high frequency nouns promoted much greater commonality than the other two lexical classes. The results also suggested that adjectives were the lexical class that favored the greatest variation in the quantitative features of participants' meaning connections since they readily elicited both common and idiosyncratic responses. Verbs and nouns encouraged a more consistent associative behavior in that each lexical class favored opposing quantitative WA features, i.e. verbs provoked heterogeneity by attracting a relatively high number of idiosyncratic and different associations, while nouns promoted homogeneity by eliciting the highest degree of commonality of responses. In the main, as frequency of occurrence increased, the heterogeneity of the associative connections also systematically increased across the three lexical categories. However, the strength of the associative domain, measured by each group's commonality of associations, showed a clear dependency on lexical category, with nouns clustering around commonly shared things and ideas more readily than adjectives verbs. Therefore, the finding that verbs seldom promoted three common responses, or a high degree of response commonality, suggested that using verbs as SW to investigate the strength of the associative domain of L2 learners would yield substantially different results compared to test instruments that employ adjectives, nouns, or both.

In sum, characteristics of the SW, in particular lexical category and frequency of occurrence, have significant effects on the qualitative and quantitative features of the WA domain of NS and non-NS alike. The present analysis revealed that differences between L1 and L2 WA domains, respectively between the mental organization of their vocabulary connections, are largely attributable to the word characteristics themselves. Consequently, the results of this line of analysis have important implications for SW sampling as well as for using WA tests to compare the qualitative and quantitative features of NS' and L2 learners' organization of the mental lexicon.

CHAPTER 6

SUMMARY OF RESEARCH FINDINGS AND THEIR CONTRIBUTION TO THE FIELD

In my thesis I proposed a model of studying and assessment of NS' and L2 learners' lexical knowledge, which was also empirically tested by an experiment specifically designed to try out the model. The model bears on proposals from previous L2 lexical research; yet, it significantly differs from other frameworks in several ways. To begin with, the model outlines two major distinctions that need to be made in the study of lexical knowledge. The first major distinction is between receptive and productive vocabulary, which is frequently acknowledged in the L2 literature, and the second one pertains to the distinction between microlevel and macrolevel of word knowledge, which applies to both the productive and the receptive domains. The second distinction has never been explicitly proposed in the field, though Read (2000), for example, makes a mention about distinguishing between assessment of knowledge of individual words and overall knowledge of vocabulary. The microlevel of the proposed model refers to knowledge of the *grammar* of individual words (*grammar* is used in its broadest sense), i.e. knowledge of a word as a phonetic/ orthographic, phonological, morphological, syntactic, and semantic entry. The macrolevel is the level that describes the overall state of an individual's lexicon within three proposed dimensions: quantity, quality, and metacognitive awareness. While the role of quantity and quality of word knowledge has been extensively discussed and studied in L2 research, little mention has been made about the place of metacognitive awareness in a model that accounts for the overall state of vocabulary knowledge. In my framework, this dimension performs a bridging function between the micro- and the macrolevel

of knowledge in that it allows language users to monitor what aspects of the grammar of words they know in a language task that provokes them to claim knowledge of vocabulary items. Also, when we are dealing with adult language users, we are also dealing with the issue of metacognitive awareness, which in the context of L2 research relates to the question whether or not L2 learners' metacognitive awareness, applied to their lexical competence, is proficiency-driven.

It goes without saying that the receptive and the productive domains, as well as the micro- and the macrolevel are not at all independent. Instead, they interact, exchange information, and inform each other; therefore, the distinctions made in the model serve the single purpose of distinguishing between what assessment of knowledge of individual words entails and how it differs from assessment of the overall lexicon. And while the microlevel has received ample research attention in L2 studies, the macrolevel of lexical competence is still a largely under-researched area with many questions to be addressed and answered. I also believe that an approach to exploring the overall state of language users' vocabulary makes examining lexical knowledge more attainable a goal; hence, the approach that I took in this research is within the global trait tradition of modeling the lexicon. Thus, I focused on the macro-level of studying lexical knowledge and designed an experiment to empirically test the significance of the proposed macrolevel dimensions of lexical assessment when language proficiency is taken into consideration.

Three groups of subjects participated in the experiment: NS of English, Bulgarian L2 advanced, and intermediate learners of English. The research questions concerned the three dimensions of vocabulary knowledge posited to account for its macrostructure, i.e. quantity, quality, and metacognitive awareness. The goal of the experiment was to find out whether a three-dimensional model of studying lexical competence can account for the variation in the vocabulary knowledge among language users at different levels of language proficiency. To my knowledge, no previous study has investigated that matter with regard to more than one

proficiency level; though, the general assumption has always been that language proficiency is positively related to L2 learners' overall state of vocabulary.

Special attention was paid to the SW, which were sampled from a dictionary by means of a spaced sampling procedure. As simple as this procedure seems to be, the discussion on the sampling process conducted for the purposes of the experiment (see Chapter 2) raised a lot of questions about the importance of a host of decisions when sampling SW from a dictionary. I tried to justify, both theoretically and methodologically, all decisions concerning the inclusion and exclusion criteria that I used in the sampling in an attempt to obtain a test sample that would reflect participants' lexicon as a whole, in addition to being a reliable test instrument for future L2 lexical research. By addressing sampling issues in my research I also wanted to underscore the relationship between sampling and final interpretation of experimental results, which as a rule rather than as an exception is overlooked in L2 lexical research. The sample that was obtained from Oxford Student's Dictionary of Current English (Hornby, 1978) became representative of the 16,045-word dictionary content, which was estimated by an independent dictionary count. The procedure itself was designed to make the sample of SW a valid subset of potential items in the L2 speakers' vocabularies. Interestingly, similar well-justified samples have been obtained by L1 researchers mainly for studying the vocabulary size of NS; yet, to my knowledge, such a theoretically, empirically, and methodologically substantiated SW selection has not been done in L2 research.

The results of this line of analysis were very rewarding and can be summarized as follows:

A. *Breath of lexical knowledge* was measured by two variables, vocabulary size and knowledge of words from different frequency of occurrence, frequently employed by L2 researchers to account for this dimension.

1. Based on the 16,045-baseline lexicon, educated adults NS of English were found to know on average approximately 9,500 words (range 7,392 – 11,501 words). As a

proportion, this number showed that the native speaking participants in the study knew on average 59% of the estimated total size of the dictionary from which the SW were selected. Given the stringent criteria used to identify words as known, this proportion compared very well with L1 studies that also used abridged dictionaries to specifically measure the quantity of vocabulary knowledge of educated adult NS of English (e.g., D'Anna et al., 1991; Zechmeister et al., 1995). In addition, when the means derived from this experiment were compared with the average sizes across other studies, the results revealed that the estimates were compatible, though non-significantly smaller. Thus, the comparison supported a conclusion of a meaningful estimate of NS' vocabulary size against which non-NS' vocabulary could be dependably matched.

2. Not surprisingly, the vocabulary size of the non-NS largely depended on their level of language proficiency. It was found that the average vocabulary size of the advanced learners (8,737 words in the range 5,990 – 11,484 words) did not significantly differ from that of NS, which indicated that advanced learners could recognize, explain/ give a synonym or translate approximately as many words as NS could. However, the intermediate students' vocabulary size differed greatly from that of NS and advanced learners alike, since the average lexicon of an intermediate student was found to contain 6,033 words (range 4,127 – 7,939 words). In light of suggestions made by some L2 researchers concerning the size of small lexicons (e.g., Meara, 1996), the estimates derived from this study suggest that on average, L2 intermediate learners' vocabulary size falls within Meara's idea of a "small lexicon", i.e. a lexicon containing between 5,000 – 6,000 words. Yet, it needs to be pointed out that the approximations arrived at in this research are strictly limited to recognition of words in isolation, which renders the lexical task receptive in nature. However, as far as productive vocabulary or performance on communicative tasks are concerned, the estimated quantity of lexical repertoire may not be the same. Nonetheless, based on Laufer (1997), the intermediate learners could be

predicted to have reached the lexical threshold that would allow them to successfully transfer L1 reading strategies onto L2 reading tasks, for example. However, in terms of text coverage, especially regarding work with texts for academic purposes, the proposed figure of 3,000 word families being sufficient for comprehension of 95% text coverage (e.g., Nation, 1990; Laufer, 1992) may need serious revision.

3. The same differences among the three groups were found when their knowledge of words from different frequency bands was compared, which further confirmed that variation in the quantity of vocabulary knowledge is reflected by the overall proficiency level of L2 learners.

B. *Quality of lexical knowledge*, or *depth,* is the most controversial dimension in L2 lexical research. Yet, researchers seem to agree that using WA tests to probe L2 learners' *depth* of vocabulary knowledge can be a valuable instrument for collecting data, revealing of the way language users organize their mental lexicon. I initially used two variables commonly associated with the analysis of the quantitative features of language users' WA domain, i.e. native-like typicality of responses and number of associations, and subsequently added a third one -- within-group consistency of the WA domain. The rationale behind using three variables, where L2 researchers usually use just one (e.g., Schmitt, 1999; Wolter, 2001, 2002; Kruse et al., 1987, etc.), was to find out which variables would account for both the effects of language proficiency on the quality of participants' lexical organization, as well as the stability of connections that they build among words they know. The analysis resulted in the following findings:

1. The comparison among the three groups of participants on their shared native-like commonality of associative responses revealed that neither the advanced, nor the intermediate learners come even close to the relatively high degree of associative commonality that the NS maintained. Similar findings are traditionally interpreted in the L2 WA literature as an indication that L2 learners have a lesser degree of connectivity

among words and, consequently, a lower quality of their meaning connections and lexical organization. However, a closer look at the association data, as well as at the socio-cultural WA research, made me challenge the usefulness of this approach to analyzing WA data on which conclusions about the quality of L2 learners' meaning organization are made for, at least, two reasons. Firstly, positing that the quality of lexical organization depends on the degree of shared native-like commonality of associative meaning is in effect positing that semantic organization is solely linguistically motivated. However, such an assumption is widely challenged by WA studies from the socio-cultural line of research (e.g., Yoshida, 1990, Szalay, 1984; Szalay & Brent, 1967; Szalay & Windle, 1968), which have convincingly shown that meaning organization, in addition to being linguistically-driven, is also strongly influenced by many extra-linguistic factors. Secondly, not only hasn't this line of research arrived at any novel findings for more than two decades, but it also does not have the potential to distinguish between different levels of proficiency. The findings of my analysis pointed that even at an advanced level of language proficiency, L2 learners do not and probably cannot reach the extent of typicality characteristic of NS' associative behavior, which does not necessarily mean that their degree of lexical connectivity is of a lesser quality than that of NS.

2. This conclusion was further supported by the comparison of the number of associations that each group of participants generated to the SW used in the study. Generally, the number of associative responses that subjects produce in a continuous WA test is taken to indicate the size of participants' associative domain, i.e. to how many different things and ideas a word is associated in the mind of an individual. The analysis revealed that on average the advanced learners generated as many associations as the NS did, whereas the intermediate learners were able to produce a significantly smaller number of responses. The results also suggested a strong relationship between the number of

different things and ideas that an individual may associate any word with and the size of his/ her lexicon, i.e. a positive relationship between quantity and quality of lexical knowledge, when measured by the size of the WA domain.

3. Finally, a closer look at the association data prompted me to examine how the degree of within-group consistency of the WA domain compared across the three groups. It was easily noticeable in the data that just like most NS tended to maintain common word association networks, the L2 learners also had stable patterns of commonality of associative responses that did not, however, resemble the ones of NS. Therefore, I compared the degree of each group's typicality of associative responses as a measure of how well a word is integrated in the lexicon of the NS, advanced, and intermediate learners. The underpinning assumption was that degree of commonality of associative behavior, rather than degree of nativelike commonality, could be a more sensitive index of language proficiency. Moreover, in my view, this approach to the analysis of WA data could account not only for the difference in L2 learners' level of proficiency, but also for the linguistic and the extra-linguistic factors that influence associative relationships.

Not surprisingly, it was found that advanced learners' consistency of WA responses was as strong as the NS' commonality of responses, whereas intermediate learners' strength of the WA domain was noticeably less consistent compared to NS' and advanced learners' semantic stability of associations. This finding was in full agreement with the comparison of the size of the associative domain of the three groups of participants, which made me argue that number of responses and within-group consistency of the associative domain provide a much more meaningful estimate of *depth* of L2 learners' lexical knowledge than native-likeness of associations.

C. *Metacognitive awareness* was the third dimension of lexical competence that was proposed in the model. Using a rating scale and verified responses allowed for comparing participants' self-perception of the words they think they know with the amount of words they

actually responded to correctly. Consequently, this made it possible to draw conclusions about their degree of metacognitive awareness regarding the overall state of their vocabulary. Based on Bialystok and Ryan's (1985) metacognitive framework suggesting that analyzed knowledge and cognitive control are two major components of the framework, it was assumed that the point of intersection between these two metalinguistic skills can be accounted for by comparing participants' self-reported knowledge with their actual knowledge of the SW. In addition, it was important to find out whether language proficiency would have specific effects on participants' ability to exercise adequate control procedures and perform well on the lexical task, as far as the degree of their analyzed knowledge permitted.

Interestingly, the results revealed that in spite of the different levels of language proficiency and, respectively, various quality and quantity of lexical knowledge, all participants were equally skilled in making relatively difficult metalinguistic decisions when responding to the SW. This finding was in concert with the hypothesis advanced by Bialystok and Ryan (1985) that if metalinguistic skill is taken to be the control required to analyze a structured representation of a language, then adult L2 learners, who have already had an experience in acquiring one language, would find it relatively easy to solve metalinguistic problems in the new language. Apparently, it should be expected that L2 adult learners at an intermediate, or higher, level of language proficiency can exercise as heightened a degree of analyzed knowledge and cognitive control in the new language as NS can in their native language.

In a nutshell, the examination of the macrolevel of lexical knowledge, across its three dimensions is sensitive to L2 learners' level of language proficiency. Similarly, the regression analysis of the full set of variables used in the study showed that the full model captured virtually all the variation in the lexical competence among the participants (92% for the NS, 99% for the advanced, and 96% for the intermediate learners). However, the analysis also raised the question whether obtaining information about all measures associated with the three dimensions of lexical competence is practically meaningful or whether there is a possibility that

some measures are more revealing of the overall state of learners' vocabulary than others. Besides, a similar practice-oriented analysis of the potential of each variable and various combinations of variables to predict the greatest amount of variance in the lexical knowledge of NS and L2 advanced and intermediate learners has not been reported in the L2 literature. So, the idea of proposing a model, comprising of the "best" set of predictors that would serve the purposes of vocabulary assessment of adult NS and non-NS of English provoked the application of several regression analysis procedures to the variables used in the experiment.

The goal of the regression analyses conducted in the study was to examine the various models of variables comparatively and find out whether the lexical knowledge of NS, advanced, and intermediate learners can be studied by a common set of predictors or whether differences in the language proficiency make certain models more efficient for one group but not for the other. Several different procedures of regression analysis were used, since it is axiomatic in the research literature that there is no one "best" method of identifying the most efficient predictor(s) of a given phenomenon. The conclusions about the usefulness of the models were based on three statistics, two of which are traditionally associated with the assessment of the practical significance of any model (R^2 and Adj. R^2) and the third one is used for the assessment of the total bias in a model (C_p).

Each of the regression procedures yielded a number of predictor sets whose practical significance was as good as the one of the full model; however the total bias of some of these models was unacceptably high. Therefore, the identification of the "best" models was based on the magnitude of their total bias, which, unfortunately, is rarely taken into consideration in L2 lexical research. In summary, the analyses revealed that:

1. The one-predictor models were the most biased models, though, their effect sizes were very high. This finding made me argue strongly against the use of any one-predictor models for the assessment of lexical knowledge of NS and non-NS alike.

2. The smallest “best” set of predictors of the overall state of vocabulary knowledge of NS and non-NS, identified by the three regression procedures, was a two-predictor model comprising of verifiable self-report and vocabulary size.
3. Some of the three- and four-predictor models also showed good characteristics of practical applicability and low total bias. However, neither those models, nor the full five-predictor model proved to have meaningful increments in the variation of participants’ general knowledge of vocabulary explained, nor did they substantially lower the total bias.

These findings suggested that a five-predictor model does not necessarily explain significantly more variation in the lexical knowledge of language users at different level of language proficiency, though it does provide a more complete picture of the researched phenomenon. Apparently, when the SW selection is not based on specific word frequency bands but includes words randomly selected form a wide range of frequencies, obtaining information about the word frequency effects on participants’ general lexical knowledge or about their associative behavior is not time and resource efficient, nor does it add anything to our understanding of language users’ vocabulary status. It was also hypothesized that the predictors identified by the regression analysis as “best” indirectly pointed to the status the proposed dimensions might have in the overall state of language users’ lexical knowledge. In this regard, the selected “best” predictor set of variables suggested that metacognitive awareness and vocabulary size are the two dimensions that are more important in our understanding of L2 vocabulary than quality. Most probably, when they develop sufficiently well, quality features emerge as well. However, I should hasten to add that the design of the study as well as the data I worked with do not allow me to draw any definitive conclusions in this regard. At this point it is only justifiable to hypothesize that for the purposes of assessment of lexical knowledge, a measure of verifiable self-reported familiarity with the SW and a measure of the

vocabulary size can account equally well for the difference in the general state of vocabulary knowledge among NS and non-NS alike.

On a final note, it deserves to be pointed out that this line of analyses resulted in proposals of predictor models, which were guided by theory and, at the same time, were rooted in previous lexical research. It also contributes to the L2 research field with a novel finding that offers another perspective on efficient models of vocabulary assessment, suggesting that it is not "breadth and depth" of vocabulary knowledge but, rather, breadth and metacognitive awareness that capture the overall state of vocabulary.

The findings reported in the first part of the thesis suggested that analyzing WA data by using native-likeness of associative commonality as a measure of the quality of non-NS' lexical knowledge was an unproductive approach to investigating WA behavior. However, in general, associative behavior holds a great potential for learning more about how non-NS' qualitative and quantitative patterns of associative meaning connections compare to those of NS. Not to mention that the fundamental questions of what the L2 learners' mental lexicon looks like, how it develops, and in what respects it is similar or different from that of NS have not been satisfactorily researched and answered in the field yet. On the other hand, the assumption that NS' and non-NS' organization of the mental lexicon is fundamentally different has been for a long time in the field with little research attention being devoted to what the nature of the differences is and how it is affected by learner-related as well as word-related factors. Thus, the psycholinguistic analysis of the associative responses generated by the three groups aimed at unveiling whether or not NS, advanced, and intermediate learners of English organize their mental lexicon in qualitatively and quantitatively different patterns. I also investigated how word characteristics, such as lexical category and frequency of occurrence, influence participants' meaning connection patterns. In addition, I included a second dimension to this examination by looking at the WA domains with respect to the effects of participants' reported degree of familiarity with the SW – a dimension that is remarkably absent from L2 lexical research.

Instead, to ensure familiarity, most researchers use only high frequency words, which, in my view, severely restricts the interpretations to only this frequency band in the lexicon. Yet, it is a common practice among L2 researchers to generalize over the whole lexicon, which is unjustifiable. In the research discussed here, the SW were randomly selected from a variety of frequency bands and familiarity was established by using the VKS, which conclusively allowed for generalizing over the whole lexicon. To my knowledge, such in-depth analyses of L2 WA data relating to the qualitative and quantitative features of the associative domains of NS and non-NS and acknowledging several levels of language proficiency and a varying degree of SW familiarity has not been reported in the L2 literature.

Several interesting findings came out of this line of analyzing WA data. In summary, the examination of the qualitative characteristics of participants' associations led to the following conclusions:

1. The analysis of the associations given to words that participants knew revealed that there were no qualitative differences in their overall association patterns. That is, all participants gave predominantly paradigmatic associations and less syntagmatic responses to stimuli they knew. Apparently, as proficiency increases, learners' vocabulary size increases too which, in turn, allows them to generate a greater number of responses of both paradigmatic and syntagmatic character. The most important thing is that the paradigmatic-syntagmatic pattern was well preserved along the three proficiency levels. Similarly, my findings do not support the hypothesis that there is a developmental shift from more "child-like" syntagmatic associations to more "adult-like" paradigmatic responses in L2 associations. On the contrary, my findings suggest that this need not be the case. L2 learners with an intermediate size of vocabulary (approximately 5,000 words and above) already have stable paradigmatic-syntagmatic patterns of connection among the words they know. Moreover, none of the participants produced clang associations for the words in this category, which allowed for

hypothesizing that SW familiarity rather than loose organization of the L2 mental lexicon motivates elicitation of more phonologically than semantically or syntactically based associations. It is important to note that this is quite different from the developmental patterns shown by children in their L1. Even at a vocabulary size of 5,000 (which is attained by the typical child at age 6), children tend to show syntagmatic preponderance of associations (Schwanenflugel, personal communication). Thus, any analogy drawn between the L2 lexicon and the development of the L1 lexicon is certainly not one-to-one.

2. This hypothesis was confirmed further by the analysis of participants' association patterns of words they thought or believed they knew but, in reality, they did not. Those words were called vaguely familiar words because participants were able to place them in the correct syntactic category but they either had partial knowledge of their meaning or misinterpreted them based on form similarities. Although for the words L2 learners knew, their association pattern showed paradigmatic dominance, this pattern noticeably changed with a decrease in their level of familiarity with the SW. In other words, while the NS still maintained paradigmatic - syntagmatic dominance, there was a shift in the response pattern of advanced and intermediate learners to syntagmatic - paradigmatic preponderance of their associative patterns. Evidently, less familiar words tend to activate more syntactic than paradigmatic connections with L2 learners than they do with NS. Similar results have received varying interpretations in the L2 literature, regrettably, most often in the "child-like/ non-nativelike" spirit of interpreting syntagmatic dominance of associative connections. In my interpretation of the results I decidedly distanced myself from the "child-like/ nonnative-like" metaphor and all the speculations that stem from it because, on the one hand, there is not sufficient L2 research to support such a speculation. On the other hand, recent L1 research suggests that SW familiarity, and respectively size of vocabulary, have a great deal to do with the predominant production

of syntagmatic responses by adults, which means that the phenomenon is not simply a developmental issue. Moreover, an adult L2 learner is different in many ways from a child, not only physically and cognitively, but linguistically as well. Therefore, it was argued that the production of more syntagmatic than paradigmatic responses to words L2 learners had partial knowledge of could be explained by several reasons (discussed in Chapter 4) other than unstable, underdeveloped, or child-like lexicon, though, I fully agree that a L2 lexicon is developing, but so is the L1 lexicon.

3. The category of clang associations attracted very few responses and it did not show a decrease with an increase in participants' proficiency, as expected. Usually, this is the category of responses that is of greatest interest to L2 WA researchers because it shows the greatest variation among native and non-native speaking participants. However, in this study it turned out to be the least interesting category because the analysis revealed no statistically significant difference among the three groups on this type of responses.

The examination of the quantitative characteristics participants' WA revealed that for familiar words, NS' and non-NS' mental lexicons are quantitatively different dependent not so much on word familiarity than on the proficiency of the L2 learners.

1. The analysis showed that the intermediate learners had quantitatively a much smaller and less diverse WA repertoire than the NS and advanced learners, whereas the advanced learners' associative domain was noticeably larger in size and more heterogeneous than NS' domain.

1. The NS gave a greater number of common responses and a smaller number of idiosyncratic responses than the advanced and intermediate learners. In this regard I argued that the lower degree of commonality and the higher degree of idiosyncrasy of non-NS' associations, found in other L2 WA studies should be seen as a distinctive pattern that characterizes L2 learners' WA domain rather than a flaw in the quantitative organization of their lexical knowledge. It was also hypothesized that this typical of L2

learners associative behavior may well be a result of their bilingualism. In other words, while beginning learners' associations have been found to maintain a high degree of commonality because of their small vocabularies, an increase in proficiency, as confirmed by this study, leads to an increase in the production of idiosyncratic associations rather than to an increase in the common responses. Therefore, based on the results of the study, it was concluded that response commonality and idiosyncrasy increase with proficiency, however, the overall pattern remains unchanged.

3. The analysis of the vaguely familiar words did not turn out to be of much value because the participants did not produce the full range of responses to this category.

In sum, the analysis of the WA data generated by Bulgarian advanced and intermediate learners of English revealed quantitative but not qualitative differences in the way they structure their lexical knowledge, compared to NS of English. The quantitative differences were most noticeable in the intermediate learners' WA domain, which suggested that they were a result of smaller vocabularies. As vocabulary size increased to 8,000 words and more, the quantitative differences vanished, with the exception of commonality of responses, which does not come close to NS' commonality even at an advanced level of language proficiency. It was also argued that the L2 mental lexicon is not subject to developmental processes that manifest themselves in a complete paradigmatic dominance. Rather, a language user's lexicon strives for a balance between paradigmatic and syntagmatic connections as it expands and perhaps some lexical categories, certain frequency bands, or both would affect certain type of associative connections favored by an individual.

This hypothesized and absolutely unresearched topic in L2 studies was the subject of the final analysis discussed in the thesis. It focused on whether or not characteristics of the SW, in particular lexical category and frequency of occurrence, affect the qualitative and quantitative features of the WA domain of NS and non-NS. I decided to examine these effects because, on the one hand, I believe that knowing more about how word characteristics affect meaning

connections would raise researchers' awareness of the importance of a careful SW selection in empirical research. On the other hand, to my knowledge, these effects have not been extensively researched in L2 studies, with the exception of Söderman (1993), who only looked at the effects of high and low frequency SW on the qualitative association patterns of NS and advanced learners of English. I believe that when non-NS are compared to NS, it is of paramount importance to know whether generalizations concerning the differences between the two broadly defined groups pertain to fundamental differences in the way the monolingual and the bilingual mind build semantic connections, or whether certain characteristics of words result in patterns of associative links that are more specific to L2 learners than to NS. Therefore, the purpose of this analysis was to shed light on issues concerning the impact of lexical category and frequency of occurrence of SW on the features of the WA domain of NS and non-NS considering participants' level of proficiency and degree of familiarity with words. The examination was based on associative responses generated to 36 SW, randomly selected from the bigger sample, which were balanced for lexical category and frequency of occurrence. As far as the qualitative features of the WA domains are concerned, the most general conclusion was that lexical category had a far greater effect on participants' associations than frequency of occurrence.

1. For familiar stimuli, the level of proficiency strongly interacted with the lexical category of the SW and had a significant effect on participants' paradigmatic associations. Further investigation revealed that groups differed only in their production of same class associations when they responded to adjectives, but not when they responded to nouns or verbs. In particular, NS produced twice as greater proportion of paradigmatic responses to adjectives than non-NS did. Thus, based on this very interesting finding, it was argued that for adjectives non-NS preferably arrange their associative connections along syntagmatic lines and this organization is independent of an increase in level of proficiency or frequency of occurrence of this lexical class. Consequently, choosing to

use only adjectives as SW in WA tests would inevitably result in findings about differences in the structure of the mental lexicon of L2 learners and NS, which, as the results of this study show, apply only to this lexical class.

1. The participants did not differ in any significant way in their response pattern to nouns and verbs, which patterns, nonetheless, did not fit nicely into the paradigmatic–syntagmatic dominance posited for adult language users. For all participants, verbs were predominantly syntagmatic and elicited on average 65% syntagmatic associations, while nouns were the exact opposite and evoked on average 72% paradigmatic responses. This finding added extra support to the earlier made argument against the persistent in L2 research "childlike/ non-nativelike" metaphor by showing that non-NS, especially after they have accumulated vocabulary over 4,000 words, simply maintain a slightly different pattern of associative connections than NS do. Non-NS, or at least Bulgarian L2 learners of English, maintain the same type of connections as NS do for nouns and verbs. However, they preferably develop syntagmatic connections between adjectives and other lexical categories, which does not mean that their adjectival links are underdeveloped or unstable. The results of this study were in full agreement with the results obtained from advanced Japanese learners by Piper and Leicester (1980), which allowed for arguing that NS' and non-NS' of English structure of mental lexicon differs qualitatively in the way these two broadly defined groups organize their connections for adjectives, but not for nouns and verbs.
3. The vaguely familiar words did not seem to be influenced by participants' level of proficiency, yet, lexical category again showed more distinctive effects than word frequency. By and large, the different lexical categories preserved their selectional powers to specific types of associative responses with verbs being primarily syntagmatic, and adjectives and nouns being mostly paradigmatic. One immediately noticeable effect of the interaction between the lower degree of familiarity and lexical

category for the L2 learners was the predominance of paradigmatic over syntagmatic associations for the category of adjectives. Thus, while NS maintained paradigmatic dominance when responding to both familiar and vaguely familiar adjectives, non-NS showed a clear tendency to generate more syntagmatic responses to familiar adjectives but this trend reversed when they responded to vaguely familiar adjectives. So the results showed that when non-NS had a lower level of familiarity, their response pattern closely resembled NS' associative behavior, which suggests that familiarity rather than proficiency promotes type-specific associations. Such an associative behavior also implies that NS and non-NS at higher level of proficiency (advanced and intermediate) establish similar patterns of connection of words in their mental lexicon. Yet, a high degree of familiarity encourages access to different type of associations for some lexical categories, in this case adjectives, for NS and non-NS.

4. Frequency of occurrence of the SW did not reveal any specific effects on the proportion of response types that would distinguish NS from non-NS in their response pattern. The three groups responded similarly to familiar and vaguely familiar words from various frequency bands. They maintained a clear syntagmatic preference across the three frequency bands for the words they vaguely knew. For the ones they knew, stimuli with high and the low frequency of occurrence still attracted more syntagmatic than paradigmatic responses, while the mid frequency SW had a paradigmatic bent. Looking comparatively at the response pattern for the two degrees of familiarity, the most striking result was that, regardless of degree of familiarity, the high and the low frequency words were overwhelmingly syntagmatic, while the mid frequency stimuli elicited relatively balanced proportions of syntagmatic and paradigmatic associations.

In the main, this line of analysis revealed that the type of associative connections is very sensitive to lexical class of words, which comes to underscore the importance of cautious SW selection guided by specific research questions. In addition, the assumption that the highest

level of lexical knowledge is represented by a paradigmatic response for both NS and non-NS was largely challenged by the results of this line of analysis. In this spirit, the production of predominantly syntagmatic responses to verbs by all participants does not give the slightest reason to believe that they have a lesser degree of knowledge of verbs compared to nouns, for example. Therefore, the notion of a "paradigmatic shift" as an indication of a better quality of lexical knowledge or better organization of mental lexicon connections should be taken with a grain of salt, especially when applied to adult NS and non-NS.

The quantitative characteristics of participants' WA domains showed to be significantly influenced by the combined effect of the SW frequency of occurrence and lexical class.

1. The two SW characteristics had a strong impact on participants' number of different responses, degree of commonality, and number of idiosyncratic associations. Interestingly, for the mid frequency words, the adjectives consistently attracted considerably more different, common, and idiosyncratic associations than verbs did. Nouns did not show to be significantly different from adjectives; yet, they evoked a smaller number of responses across the three quantitative measures. The result suggested that adjectives seem to be the most diverse lexical class which readily elicits both common and idiosyncratic responses. On the other hand, participants seemed to respond less willingly to verbs, which elicited on average a very small number of common associations and a relatively high number of idiosyncratic and different associations. As frequency of occurrence increased, the total number of responses generated to the three quantitative measures also increased and high frequency nouns were more often associated with common to participants' things and ideas than verbs or adjectives were. Respectively, verbs and adjectives evoked more diverse and more idiosyncratic associations. Therefore, it can be summarized that as word frequency increased, the heterogeneity of the associative connections also systematically increased across the three lexical categories. However, the strength of the associative

domain, measured by each group's commonality of associations, showed a clear dependency on lexical category, with high frequency nouns clustering together readily, followed by adjectives and verbs.

1. Frequency of occurrence was directly related to total number of responses and the availability of the three most common responses, i.e. the lower the SW frequency, the lower the total number of responses as well as the frequency of the three most common responses. While it was not surprising that low frequency stimuli evoked a small number of responses, it was interesting that the proportion of the three commonly given associations was almost the same across the three frequency bands (25 % for low, 26 % for mid, and 27% for high frequency stimuli). This suggests that there is no reason to expect that high or medium frequency stimuli would elicit the highest degree of commonality of primary responses, as traditionally accepted, if proportion of responses is used to measure the strength of the associative domain.
3. Lexical category of the stimuli also had distinguishable effects on total number of responses and availability of the three most common associations. Verbs evoked on average a much smaller number of associations, as well as a lower frequency of occurrence of the three most popular responses than nouns and adjectives. Overall, participants generated associations to nouns and adjectives more readily than they did to verbs. Moreover, the fact that verbs seldom promoted three common responses suggests that using verbs as SW to investigate the strength of the associative domain of language users would yield substantially different results compared to test instruments that employ adjectives, nouns, or both.
4. The analysis of the responses to vaguely familiar words did not turn out to be of much value mostly because a small number of the SW were recognized as vaguely familiar.

In sum, this line of analysis proved the importance of careful consideration of SW selection and raised the question about the insights we gain into the L2 mental lexicon

organization, when comparing NS with non-NS, without regard to the special effects of the lexical input on the lexical acquisition process. The analysis presented in this thesis confirmed that features of the lexical input have distinctive effects on the qualitative and quantitative characteristics of the connectivity among lexical items in the L1 and L2 mental lexicon, which should not be disregarded in instruction, testing, and research.

Conclusion

One of the practical uses of WA tests as a research tool relates to contemporary theories of associative structure proposing that associations represent the way in which semantic information is structured in memory (Nelson, 1977). In the context of L2 research, WA tests have been very often employed to examine how L2 learners organize their lexical and semantic information and whether or not their lexical knowledge is structured much the same way as NS' knowledge. Psycholinguistic research interested in the conceptual representation in bilingual memory has taken the use of WA tests even a step further by arguing that the word association task reflects not only conceptual but lexical processing as well (van Hell & de Groot, 1998). The more interesting question, though, is whether conceptual representations are purely language specific or purely shared between languages. While the present study does not allow for speculations concerning bilingual conceptual representations, it was designed to investigate the qualitative and quantitative features of L2 learners' lexical store by considering their language proficiency and word familiarity. It also addressed several fundamental questions in L2 lexical research, which have not been sufficiently well researched in the field, such as whether differences in the NS and non-NS lexicons are predominantly qualitative or quantitative in nature and how certain features of words influence meaning connections. The wealth of novel findings allowed me to elaborate on some interpretations of L2 WA results advanced by previously conducted L2 WA studies, as well as to challenge some stagnant assumptions in the field, which needed to be scientifically challenged.

This study also raised several questions that need to be further researched. One of these concerns the relationship between the native language and the L2 in building meaning connections. It was hypothesized that the native language might influence both the qualitative and the quantitative features of participants' WA domains. For example, L2 studies have consistently found that non-NS cannot maintain the high degree of nativelike associative commonality shown by NS even at higher levels of language proficiency. Though the reasons are not clear yet, the possibility that the higher degree of heterogeneity and idiosyncrasy of non-NS' semantic connections might be L1-influenced has not been well-explored. In this regard, it would be worthwhile to discern whether this specific to non-NS associative behavior might be a consequence of L1 associative behavior or whether this is a result of the residually activated native language when performing in an L2 associative task. Very little research has been done in this direction, though there is evidence to believe that the native language influences the responses of the L2 but not vice versa (e.g., Riegel et al.,1967).

Another possible reason for the high idiosyncrasy of non-NS' responses that also deserves to be looked into is the effects of instruction and use of different teaching materials in L2 formal instruction, as well as the impact of informal training on the meaning connections developed by non-NS. It is evident that the mental lexicon of any L2 language user is a flexible system that changes along several dimensions but it would be interesting to examine comparatively how formal instruction and less formal training affect the qualitative and quantitative features of semantic connections and how this is reflected by language users' overall proficiency and ability to perform on academic tasks. On the one hand, it might be the case that instructional materials and teaching practices favor the development of certain types of meaning connections in the mental lexicon. For example, Politzer (1978) found that dialogue drills, acting out of dialogues, and translation drills greatly encourage the establishment of syntagmatic associations, whereas substitution drills promote primarily paradigmatic associations. He also found that the only activity that seemed completely counterproductive was

translation, while the only activity that approached a significant correlation with total number of responses was conversational practice (p. 209). On the other hand, it might be the case that informal training also favors specific connections, which might be of a different type. However, the more interesting issue that needs to be researched is how qualitative and quantitative differences between NS' and non-NS' lexicon relate to performance in academic tasks, which will promote evaluation of the notion of "differences between NS' and non-NS' mental lexicon organization", as well as re-evaluation of the paradigmatic-syntagmatic relationship in the context of performance.

Finally, some semantic generativists (e.g., Hale & Keyser, 1992; Juffs, 1996) argue that argument structure should be examined in the framework of lexical relational structure because lexical relational structure is the level of semantics that is relevant to syntax**.** Therefore, it would be useful to research L2 learners' knowledge of semantics - syntax correspondence by studying assignment of thematic roles to words as a way of studying depth of word knowledge and its relationship to syntactic structures. Such line of research will explore the assumption that a lexical parameter is crucially responsible for argument structure alternations (with respect to both number of arguments permitted to be assigned and variation in their mapping to grammatical function) rather than a purely syntactic parameter (Juffs, 1996). Similarly, it has been acknowledged in the field that there are some hidden dependencies in the second language learners' competence and some of these dependencies are lexical in nature, but syntactically analyzable (White, 1994). Unfortunately, seldom have systematic links between the lexicon and syntax been considered by L2 researchers, especially in light of a theory of lexical representation in syntactic structures, where derivational morphology becomes a very influential factor (Juffs, 1996). Therefore, exploring this relationship may shed light on how knowledge of lexical properties of words is realized syntactically in the L2.

Overall, the research presented in this thesis was motivated by an interest to shed some light on several elusive issues in L2 lexical research, in particular model-based investigation and

assessment of NS' and non-NS' lexical competence, the structure of their mental lexicon, and word features that affect the qualitative and quantitative characteristics of language users' meaning connections. The research was intended to bridge two traditions of examining lexical competence, i.e. applied linguistic and psycholinguistic, by exploring the lexicon from practical and structural point of view. And, as it always happens in scientific research, the study threw some valuable light on several overlooked and under-researched issues in the L2 lexical field, challenged some stagnant assumptions, and probably raised more questions than it answered.

REFERENCES

Aitchison, J. (1987). Words in the mind. An introduction to the mental lexicon. New York, NY: Basil Blackwell.

Aitchison, J. (1992). Linguistics. London: Hodder & Strongthon.

Anderman, G. (1996). The word is my oyster – the language learner and the translator. In G. Anderman and M. Rogers (Eds.), Words, words, words. The translator and the language learner (pp. 41–55). Philadelphia, PA: Multilingual Matters.

Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), Comprehension and teaching: research reviews (pp. 77-117). Newark: DE. International reading Association.

Anglin, J. M. (1993). Vocabulary development: A morphological analysis. Monographs of the Society for Research in Child Development, 58 (10, Serial No. 238).

Baayen, H., & Renouf, A. (1996). Chronicling the Times: productive lexical innovation in an English newspaper. Language, 72 (1), 69-96.

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

Bauer, E. (1983). English word-formation. Cambridge University Press.

Beglar, D., & Hunt, Alan (1999). Revising and validating the 2000 Word Level and University Word Level Vocabulary Tests. Language Testing, 6 (2), 131-162.

Bialystok, E., & Ryan, E. (1985). A metacognitive framework for the development of first and second language skills. In D. L. Forrest-Pressley, G. E. MacKinnon, and T. G. Waller

(Eds.), Metacognition, cognition, and human performance (pp. 207-246). Orlando: FL. Academic Press.

Bialystok, E., & Sharwood Smith, M. (1985). Interlanguage is not a state of mind: an evaluation of the construct for second language acquisition. Applied Linguistics, 6, 101-117.

Bleiler, E. (1963). Essential Japanese grammar. New York: Dover Publications.

Bloomfield, L. (1933). Language. New York: Henry Holt.

Bogaards, P. (2001). Lexical units and the learning of foreign language vocabulary. Studies in Second Language Acquisition, 23, 321–343.

Brown, A. L. (1980). Metacognitive development and reading. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), Theoretical issues in reading comprehension: Perspectives from cognitive psychology, linguistics, artificial intelligence, and education (pp. 453-481). Hillsdale, NJ: Lawrence Erlbaum.

Brownstein, S., Weiner, M., & Green, S. W. (1996). A pocket guide to vocabulary. Barron's Educational Series. Sofia: Apollo.

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

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

Chomsky, N. (1989). Some notes on economy of derivation and representation. MIT Working Papers in Linguistics 10; 43-74.

Clark, H. (1970). Word associations and linguistic theory. In J. Lyons (Ed), New horizons in linguistics (pp. 271-286). Middlesex, England: Penguin.

Coady, J. (1993). Research on ESL/ EFL vocabulary acquisition. In T. Huchin, M. Haynes, and J. Coady (Eds.), Second language reading and vocabulary learning. Norwood, NJ: Ablex.

Cook, V. (1988). Chomsky's Universal Grammar: an introduction. Oxford: Blackwell.

Cook, V. (1993). Linguistics and second language acquisition. New York, NY: St. Martin's Press.

Cook, V. (1995). Multicompetence and effects of age. In D. Singleton and Z. Lengyel (Eds.), The age factor in second language acquisition: a critical look at the Critical Period Hypothesis. Clevedon: Multilingual Matters.

Cramer, P. (1968). Word Association. New York, NY: Academic Press.

D'Anna, C. A., Zechmeister, E. B., & Hall, J. W. (1991). Toward a meaningful definition of vocabulary size. Journal of Reading Behavior, 23, 109-122.

Dale, E. (1965). Vocabulary measurement: techniques and major findings. Elementary English, 42, 395-401.

Davids, A. (1956). Personality dispositions, word frequency, and word associations. Journal of Personality, 24, 328-338.

Davis, B. J., & Wertheimer, M. (1967). Some determinants of associations to French and English words. Journal of Verbal Learning and Verbal Behavior, 6, 574-581.

De Bot, K. (1992). A bilingual production model: Levelt's speaking model adapted. Applied Linguistics, 13, (1), 1-24.

De Bot, K., Paribakht, T. S., & Wesche, M. B. (1997). Toward a lexical processing model for the study of second language vocabulary acquisition. Studies in Second Language Acquisition, 19, 309–929.

Deering, G. (1963). Affective stimuli and disturbance of thought processes. Journal of Consulting Psychology, 27, 338-343.

Deese, J. (1962). Form class and the determinants of association. Journal of Verbal Learning and Learning Behavior, 1, 79-84.

Deese, J. (1965). The structure of associations in language and thought. Baltimore, MD: The Johns Hopkins Press.

Diller, K. C. (1978). The language teaching controversy. Rowley, Mass.: Newbury House.

Dolch, E. W., & Leeds, D. (1953). Vocabulary tests and depth of meaning. Journal of Educational Research 4, 181-189.

Dubin, F., & Olshtain, E. (1986). Course design. Cambridge: Cambridge University Press.

Dupuy, H. J. (1974). The rationale, development and standardization of a basic word vocabulary test. Washington, DC: US Government printing Office.

Ehrlich, E., Flexner, S. B., Carruth, G., & Hawkins, J. M. (Eds.) (1980). Oxford American dictionary. New York: Oxford University Press.

Ellis, N. C. (2002). Frequency effects in language processing. Studies in Second Language Acquisition, 24, 143 – 188.

Epstein, S. D., Thránsson H., & Zwart, C. (1996). Introduction. In W. Abraham, S. D. Epstein, H. Thránsson, & C. Zwart (Eds.), Minimal ideas. Amsterdam/ Philadelphia, PA: John Benjamins.

Ervin, S. (1968). An analysis of the interaction of language topic, and listener. In J. Fishman (Ed.), Reading in the sociology of language (pp. 161-187). The Hague: Mouton.

Ervin, S. M. (1961). Changes with age in the verbal determinants of word association. American Journal of Psychology, 74, 361-372.

Færch, C., & Kasper, G. (1987). Introspection in second language research. Philadelphia, PA: Multilingual Matters.

Fillenbaum, S., & Jones, L. (1965). Grammatical contingencies in word association. Journal of Verbal Learning and Behavior, 4, 248-255.

Flavell, J. H. (1976). Metacognitive aspects of problem-solving. In B. Resnick (Ed.), The nature of intelligence. Hillsdale, NJ: Erlbaum.

Frankel, J. R., & Wallen, N. E. (2000). How to design and evaluate research in education: The McGraw-Hill Companies.

Gass, S. (1988). Second language vocabulary acquisition. Annual Review of Applied Linguistics, 9, 92-106.

Gass, S. (1990). Second and foreign language learning: same, different or none of the above? In B. VanPatten &J. F. Lee (Eds.), Second language acquisition/ foreign language learning (pp. 34-45). Clevedon-Philadelphia: PA, Multilingual Matters.

Gass, S. M., & Selinker, L. (2001). Second language acquisition: An introductory course. Mahwan, NJ: Lawrence Earlbaum.

Gitsaki, C. (1999). Second language lexical acquisition. A study of the development of collocational knowledge. International Scholars Publications.

Gombert, J. E. (1992). Metalinguistic development. Chicago: University of Chicago Press.

Goulden, R., Nation, P., & Read, J. (1990) How large can a receptive vocabulary be? Applied Linguistics, 11, 341-363.

Gove, P. B. (Ed.). (1961). Webster's Third New International Dictionary. Springfield. Mass.: G. & C. Merriam.

Grabois, H. (1997). Love and power. Word associations, lexical organization, and second language acquisition. Unpublished doctoral dissertation, Cornell University, Ithaca, New York.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. Bilingualism: Language and Cognition, 1, 67-81.

Greidanus, T., & Nienhuis, L. (2001). Testing the quality of word knowledge in a second language by means of word associations: Types of distractors and types of associations. The Modern Language Journal, 85 (4), 567-577.

Haegeman, L. (1994). Introduction to government and binding theory. Cambridge, MA: Blackwell.

Hartmann, G. W. (1941). A critique of the common method of estimating vocabulary size, together with some data on the absolute word knowledge of educated adults. Journal of Educational Psychology, 32, 351-358.

Haynes, M., & Baker, I. (1993). American and Chinese readers learning from lexical familiarization in English texts. In T. Huckin, M. Haynes, & J. Coady (Eds.), Second language reading and vocabulary acquisition (pp. 130-152). Norwood, NJ: Ablex.

Henriksen, B. (1999). Three dimensions of vocabulary development. Studies in Second Language Acquisition, 21, 303–317.

Hornby, A. S. (1978). Oxford Student's Dictionary of Current English. London: Oxford University Press.

Hornby, A. S., Gatenby, E. V., & Wakefield, H. (1963). The advanced learner's dictionary of current English. London: Oxford University Press.

Johnson, C. J., & Anglin, J. M. (1995). Qualitative development in the content and form of children's definitions. Journal of Speech and Hearing Research, 38 (3), 612-626.

Jones, F. R. (1994). The lone language learner: a diary study. System, 22 (4), 441-454.

Kempen, G., & Huijbers, P. (1983). The lexicalization process in sentence production and naming: indirect selection of words. Cognition 14, 185-209.

Kent, G. H., & Rosanoff, A. J. (1910). A study of association in insanity. American Journal of Insanity, 67, 37-39, 317-390.

Kirkpatrick, E. (1891). The number of words in an ordinary vocabulary. Science, 18, 107-108.

Kirsner, K., Smith, M., Lockhart, R. & Jain, M. (1984). The bilingual lexicon: language specific units in an integrated network. Journal of Verbal Learning and Verbal Behavior, 23, 519-539.

Kittay, E. F., & Lehrer, A. (1992). Introduction. In A. Lehrer, & E. F. Kittay (Eds.), Frames, fields and contrasts (pp.1-20). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kolers, P. A. (1963). Interlingual word associations. Journal of Verbal Learning and Verbal Behavior, 2, 291-300.

Kolln, M., & Funk, R. (2002). Understanding English grammar. Longman.

Колс- Sabo, E., & Lightbrown P. (1999). Students' approaches to vocabulary learning and their relationship to success. Modern Language Learning. (83), 176–192.

Kruse, H, Pankhurst J., & Sharwood Smith, M. (1987). Multiple word association probe in second language acquisition research. Studies in Second Language Acquisition, 9, 141-154.

Lambert, W. E. (1972). Language psychology and culture. Stanford: Stanford University Press.

Landau, S. I. (2001). Dictionaries: the art and craft of lexicography. Cambridge University Press.

Lantolf, J. P. (1999). Second culture acquisition. In E. Hinkel (Ed.), Culture in second language teaching and learning (pp. 28-46). New York: NY. Cambridge University Press.

Laufer, B. (1989) A factor of difficulty in vocabulary learning: deceptive transparency. In P. Nation & R. Carter (Eds.). Vocabulary acquisition. Amsterdam: Free University Press.

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In P. J. Arnaud & H. Béjoint (Eds.), Vocabulary and applied linguistics. London: Macmillan.

Laufer, B. (1997). The lexical plight in second language acquisition: words you don't know, words you think you know, and words you can't guess. In J. Coady and T. Huckin (Eds.), Second language vocabulary acquisition (pp. 20-34). Cambridge: Cambridge University Press.

Laufer, B., & Paribakht, S. (1998). The relationship between passive and active vocabularies: Effects of language learning contexts. Language Learning (48), 365–391.

Laufer-Dvorkin, B. (1991). Similar lexical forms in interlanguage. Тьbingen: G. Narr.

Levelt, W. (1989). Speaking from intention to articulation. Cambridge, MA: MIT Press.

Levelt, W. (1993). Accessing words in speech production: Stages, processes and representations. In W. Levelt (Ed.) Lexical access in speech production (pp. 1–22). Cambridge, MA: Blackwell.

Liu, E. T. K., & Shaw, P. M. (2001). Investigating learner vocabulary. A possible approach to looking at EFL/ ESL learner qualitative knowledge of the word. IRAL (39), 171 – 94.

Longman Dictionary of English Language and Culture. (1992). Essex, England: Longman.

Lorge, I., & Chall, J. (1963). Examining the size of vocabularies of children and adults: an analysis of methodological issues. Journal of Experimental Education 32, 147-157.

McLaughlin, B. (1987). Theories of second language learning. New York, NY: Edward Arnold.

McNiel, D. (1966). A study of word association. Journal of Verbal Learning and Learning behavior, 5, 548-557.

McNiel, D. (1970). The acquisition of language. New York: Harper & Row.

Meara, P. (1978). Learners' associations in French. Interlanguage Studies Bulletin (3), 192-211.

Meara, P. (1982). Word associations in a foreign language. A report on the Birkbeck Vocabulary Report. The Nottingham Linguistic Circular, 11, (2), 29-38.

Meara, P. (1984). The study of lexis in interlanguage. In A. Davies, C. Criper, and A. P. R. Howatt (Eds.), Interlanguage (pp. 225-235). Edinburgh: University of Edinburgh Press.

Meara, P. (1992). Network structures and vocabulary acquisition in a foreign language. In P. Arnaud, & H. Béjoint (Eds.), Vocabulary and applied linguistics (pp. 62-70). London: McMillan.

Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjæ, and J. Williams (Eds.), Competence and performance in language learning. Cambridge: Cambridge University Press.

Meara, P. (1997). Towards a new approach to modeling vocabulary acquisition. In N. Schmitt and M. McCarthy (Eds.), Vocabulary: description, acquisition and pedagogy (pp. 109-121). Cambridge: Cambridge University Press.

Meara, P., & Jones, G. (1988). Vocabulary size as a placement indicator. In P. Grunwell (Ed.), Applied Linguistics in Society (pp. 80-87). London: Center for Information on Language Teaching and Research.

Melka, F. (1997). Receptive vs. productive aspects of vocabulary. In N. Schmitt and M. McCarthy (Eds.), Vocabulary: description, acquisition and pedagogy (pp. 84-102). Cambridge: Cambridge University Press.

Miron, M. S. & Wolfe, S. (1964). A cross-linguistic analysis of the response distribution of restricted word associations. Journal of Verbal Learning and Verbal Behavior, 3, 376-384.

Morton, J. (1969): The interaction of information in word recognition. Psychological Review, 76, 165-178.

Nagy, W., & Anderson, R. (1984). How many words are there in printed school English? Reading Research Quarterly, 19, 304-330.

Nation, I. S. P. (1990). Teaching and learning vocabulary. New York: Newbury House.

Nation, I. S. P. (1993). Using dictionaries to estimate vocabulary size: essential, but rarely followed procedures. Language Testing, 10, 27-40.

Nation, I. S. P. (2001). Learning vocabulary in another language. Cambridge University.

Nelson, K. (1977). The syntagmatic shift revised: a review of research and theory. Psychological Bulletin, 84, 93-116.

Noble, C. E. (1953). The meaning-familiarity relationship. Psychological Review, 60, 89-98.

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 Production Report No. 10). Bloomington, IN: Indiana University.

O'Dell (1997). Incorporating vocabulary into the syllabus. In N. Schmitt and M. McCarthy (Eds.), Vocabulary: description, acquisition and pedagogy (pp. 258-278). Cambridge: Cambridge University Press.

Oldfield, R. C. (1963). Individual vocabulary and semantic currency: a preliminary study. British Journal of Social and Clinical Psychology, 31, 14-38.

Paribakht, T., & Wesche, M. (1993). The relationship between reading comprehension and second language development in a comprehension-based ESL program. TESL Canada Journal, 11, 9-29.

Pedhazur, E. J. (1997). Multiple regression in behavioral research. Wadsworth, Thomson Learning.

Piper, T. H., & Leicester, P. F. (1980). Word association behavior as an indicator of English language proficiency. (ERIC Document Reproduction Service No. ED 227 651)

Plag, I. (1999). Morphological productivity: structural constraints in English derivation. Berlin: Mouton de Guyter.

Politzer, R. L. (1978). Errors of English speakers of German as perceived and evaluated by German natives. Modern Language Journal, 62, 253-261.

Popham, W. J. (1990). Modern educational measurement: A practitioner's perspective. New Jersey: Prentice Hall.

Postman, L., & Keppel, G (Eds.). (1970). Norms of word association. New York: Academic Press.

Radford, A. (1997). Syntax: A minimalist introduction. Cambridge University Press.

Read, J. (1993). The development of new measure of L2 vocabulary knowledge. Language Testing, 10, 355-371.

Read, J. (2000). Assessing vocabulary. Cambridge University Press.

Richards, J. C. (1976). The role of vocabulary teaching. TESOL Quarterly, 10, 77-89.

Riegel, K. F., Ramsey, R. M., & Riegel, R. M. (1967). A comparison of the first and second language of American and Spanish students. Journal of Verbal Learning and Verbal Behavior, 6, 536-544.

Rozenweig, M. R. (1961). Comparisons among word association responses in English, French, German, and Italian. American Journal of Psychology, 74, 347-360.

Saeed, J. (1997). Semantics. Oxford, Cambridge, Mass: Blackwell.

Sandra, D. (1994). The morphology of the mental lexicon: internal word structure viewed from a psycholinguistic perspective. Language and Cognitive Processes 9 (3), 227–269.

Schmit, N. (1998). Quantifying word association responses: what is native-like? System, 26, 389-401.

Schmitt, N. (1999). The relationship between TOEFL vocabulary items and meaning, association, collocation and word class knowledge. Language Testing 16 (2), 189-216.

Schmitt, N., & Meara, P. (1997). Researching vocabulary using a word knowledge framework: word associations and verbal suffix knowledge. Studies in Second Language Acquisition 19, 17-35.

Schwanenflugel P. Blount, B., & Lin, P. (1991). Cross-cultural aspects of word meaning. In P. Schwanenflugel (Ed.), The psychology of word meaning (pp. 72-87). Hillsdale, NJ: Lawrence Erlbaum Associates.

Schwanenflugel, P. J., & Rey, M. (1986). Interlingual semantic facilitation: Evidence for a common representational system in the bilingual lexicon. Journal of Memory and Language, 25, 605-618.

Schwanenflugel, P. J., & White, C. R. (1991). The influence of paragraph information on the processing of upcoming words. Reading Research Quarterly, 26, 160-177.

Schwanenflugel, P., & Gavisk, C. (2002). Psycholinguistic aspects of word meaning. In D. A. Cruse, F. Hudsnurscher, M. Job, & P. R. Lutzeier (Eds.). Lexicology: An international handbook on the nature and structure of words and vocabularies. NY: de Gruyter.

Seashore, R. H., & Eckerson, L. D. (1940). The measurement of individual differences in general English vocabularies. Journal of Educational Psychology 31, 14-37.

Segalowitz, N. (1986). Skilled reading in the second language. In J. Vaid (Ed.), Language processing in bilingualism: Psycholinguistic and neuropsychological perspectives (pp. 3-19). Hillsdale, NJ: Lawrence Erlbaum.

Shore, W. J., & Durso, F. T. (1991). Partial knowledge of word meaning. Journal of Experimental Psychology, 2, 190-202.

Singleton, D. (1999). Exploring the second language mental lexicon. Cambridge, UK: Cambridge University Press.

Smith, M. K. (1941). Measurement of the size of general English vocabulary through the elementary grades and high school. Genetic Psychology Monographs 24, 311-345.

Smith, N. (1999). Chomsky: ideas and ideals. Cambridge: Cambridge University Press.

Söderman, T. (1993). Word associations of foreign language learners and native speakers. The phenomenon of shift od response type and its relevance for lexical development. In H. Ringbom (Ed.), Near-native proficiency in English (pp. 98-182).Abo, Finland: Abo Akademi.

Stolz, W. S., & Tiffany, J. (1972). The production of “child-like” word associations by adults to unfamiliar adjectives. Journal of Verbal Learning and Learning Behavior, 11, 38-46.

Szalay, L. B. (1984). An in-depth analysis of cultural/ ideological belief systems. Mankind Quarterly, 25, 71-100.

Szalay, L. B., & Brent, J. (1967). The analysis of cultural meaning through free verbal associations. Journal of Social Psychology, 3, 247-258.

Szalay, L. B., & Windle, C. (1968). Relative influence of linguistic versus cultural factors on free verbal associations. Psychological Reports, 22, 43-51.

Szalay, L. B., Lynse, D. A., & Bryson, J. A. (1972). Designing and testing cogent communication. Journal of Cross-cultural Psychology, 3, 247-258.

Templin, M. (1957). Certain language skills in children: Their development and interrelationships (Institute of child welfare, Monograph series No. 26). Minneapolis: University of Minnesota Press.

Thorndike, E. L., & Lorge, I. (1944). The teacher's word book of 30,000 words. New York, NY: Teachers College, Columbia University.

Trembly, D. (1966). Laws of learning general and specialized vocabularies [Summary]. Proceedings of the 74th Annual Convention of American Psychological Association, 1, 229-230.

Underwood, B. J. (1959). Verbal meaning in the educative process. Harvard Educational Review, 29, 107-117.

Van Hell, J., & de Groot, A. (1998). Conceptual representation in bilingual memory: effects of concreteness and cognate status in word association. Bilingualism: Language and Cognition, 1 (3), 193-211.

VanPatten, B. (1990). Theory and research in second language acquisition and foreign language learning: on producers and consumers. In B. VanPatten &J. F. Lee (Eds.), Second language acquisition/ foreign language learning (pp. 17-26). Clevedon-Philadelphia: PA, Multilingual Matters.

Vermer, A. (2001). Breadth and depth of vocabulary in relation to L1/ L2 acquisition and frequency if input. Applied Psycholinguistics 22, 217-254.

Vierra, A., & Pollock, J. (1992). Reading educational research. Scottdale: Gorsuch Searisbrick.

Wesche, M., & Paribakht, S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. Canadian Modern Language Review, 53, 13-40.

West, M. (1953). A general service list of English words. London: Longman.

Wispé, L. G. (1954). Psychological need, word frequency, and word association. Journal of Abnormal and Social Psychology, 49, 229-234.

Wolter, B. (2001). Comparing the L1 and L2 mental lexicon. Studies in Second Language Acquisition, 23, 41–69.

Wolter, B. (2002). Assessing proficiency through word associations: is there still hope? System, 30 (3), 315-329.

Xue, G., & Nation, I. S. P. (1984). A university word list. Language Learning and Communication, 3, 215-229.

Yoshida, K. (1990). Knowing vs. behaving vs. feeling: Studies on Japanese bilinguals. In L. A. Arena (Ed.), Language proficiency (pp. 19-40). New York: Plenum Press.

Zechmeister E. B., Chronis, A. M., Cull, W. L., D'Anna, C. A., & Healy, N. A. (1995). Growth of functionally important lexicon. Journal of Reading Behavior, 27 (2), 201-212.

Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). The educator's word frequency guide. Touchstone Applied Science Association (TASA).

APPENDICES

APPENDIX A

LIST OF STIMULUS WORDS USED IN THE STUDY

1. abattoir	20. drawback (n.)	39. masochism	58. settled
2. advantageous	21. edifice	40. middling (adj.)	59. shuttle (n.)
3. amoral	22. entrust	41. monorail	60. sleigh (n.)
4. array (v.)	23. experimentation	42. naïve	61. solstice
5. back (v.)	24. fathom (v.)	43. official (adj.)	62. spotlight (n.)
6. beaten	25. flagstone	44. out (n.)	63. studio
7. blanket (n.)	26. forgo	45. parable	64. sweep (v)
8. bracelet	27. gambol (v.)	46. penance	65. telegraph (n.)
9. bursar	28. glower	47. pillar (n.)	66. throw (n.)
10. cassava	29. griddle (n.)	48. point-blank (adj.)	67. toxic
11. choke (n)	30. hard (adv.)	49. prefect	68. tuner
12. coinage	31. high (adv.)	50. promontory	69. unnerve
13. concede	32. hunger (n.)	51. putative	70. virtuosity
14. contravention	33. inception	52. refuse (v.)	71. weaken
15. crampons	34. instil	53. residence	72. wind (v.)
16. custodian	35. jib (n.)	54. rigidity	73. yield (v.)
17. defensive	36. lackadaisical	55. ruler	
18. diamond	37. livelihood	56. savor (v.)	
19. dissension	38. macaw	57. second (nmrl.)	

APPENDIX B

CLASSIFICATION OF THE STIMULI ACCORDING TO THEIR SYNTACTIC CATEGORY

Nouns	**Verbs**	**Adjectives**	**Adverbs**
abattoir blanket bracelet bursar cassava choke coinage contravention crampons custodian diamond dissension drawback edifice experimentation flagstone griddle hunger inception jib livelihood macaw masochism monorail out (n.) parable penance pillar prefect promontory residence rigidity ruler shuttle sleigh solstice spotlight studio telegraph throw tuner virtuosity	array back concede entrust fathom forgo gambol glower instil refuse savor sweep unnerve weaken wind yield	advantageous amoral beaten defensive lackadaisical middling naïve official point-blank putative second settled toxic	hard high

APPENDIX C

DISTRIBUTION OF THE STIMULI ACCORDING TO WORD FREQUENCY (SFI)

SFI = 20+		SFI = 30+		SFI = 40+		SFI = 50+	
20.8	rigidity	30.7	tuner	40.4	fathom (v.)	50.6	beaten
22.1	abattoir	32.9	entrust	41.7	naïve	50.7	sweep (v)
22.1	bursar	32.9	flagstone (n.)	41.9	custodian	51.5	diamond
22.1	contravention	34.0	lackadaisical	41.9	drawback (n.)	51.6	telegraph (n.)
22.1	gambol (v.)	34.1	jib (n.)	42.2	spotlight (n.)	51.7	yield (v.)
22.1	macaw	34.9	parable	42.6	sleigh (n.)	53.4	hunger (n.)
22.1	masochism	35.0	dissension	42.8	livelihood	54.3	blanket (n.)
22.1	putative	35.0	savor (v.)	43.1	pillar (n.)	54.6	ruler
25.1	virtuosity	35.4	monorail	43.4	advantageous	55.5	official (adj.)
25.5	amoral	35.5	edifice	43.8	bracelet	56.9	throw (n.)
26.9	point-blank (adj.)	35.5	middling (adj.)	44.9	choke (n)	57.1	refuse (v.)
27.7	crampons	36.0	prefect	46.4	toxic	58.7	settled
29.3	glower	36.5	forgo	46.4	weaken	62.4	wind (v.)
		36.6	unnerve	46.6	experimentation	65.3	second (nmrl.)
		37.2	instil	46.8	array (v.)	65.8	hard (adv.)
		37.4	inception	46.8	defensive	66.9	high (adv.)
		37.4	solstice	47.4	shuttle (n.)	70.3	back (v.)
		38.1	penance	48.6	residence	73.7	out
		38.4	cassava	49.7	studio		
		38.7	promontory				
		39.1	griddle (n.)				
		39.4	concede				
		39.6	coinage				

APPENDIX D

INSTRUCTIONS TO THE TEST AND AN EXAMPLE OF A TEST ITEM

The following activity will ask you to rate your degree of familiarity with the list of words below.

Please answer as accurately as you can. **Do not** leave any of the items unanswered.

For item **V** write down as many as **three words** that you associate with the item in bold.

When the lexical category is specified in brackets next to the word (eg. *noun*, *verb*, *adjective*, *adverb*), please, respond to the word as specified.

Note: If you choose III or IV, you should do V as well!

1. abattoir

I. I ***have not seen*** this word before □

II. I ***have seen*** this word before but I ***don't remember*** what it means □

III. I ***think*** this word means _____ (synonym, translation, or brief explanation)

IV. I ***know*** that this word means _____ (synonym, translation, or brief explanation)

V. I ***associate*** this word with _____ , _____, _____

APPENDIX E

ONE-WAY ANOVA SUMMARY TABLE WITH GROUP MEANS AND STANDARD DEVIATIONS OF NS, L2 ADVANCED, AND INTERMEDIATE LEARNERS OF ENGLISH

Group	*N*	*Mean*	*SD*
Native Speakers			
Verified vocabulary knowledge	30	219.90	27.007
Self-reported lexical knowledge	30	229.27	25.899
Vocabulary size	30	9447.53	2054.522
Word frequency	30	170.73	43.918
Native-like typicality of associations	30	369.20	85.103
Number of associations	30	111.57	26.200
Within-group consistency of WA domain	30	369.20	85.103
Advanced learners			
Verified vocabulary knowledge	17	206.18	31.744
Self-reported lexical knowledge	17	218.29	26.770
Vocabulary size	17	8737.65	2747.556
Word frequency	17	156.882	56.696
Native-like typicality of associations	17	198.294	88.527
Number of associations	17	120.647	35.380
Within-group consistency of WA domain	17	345.882	82.449
Intermediate learners			
Vocabulary knowledge	17	164.41	19.606
Self-reported lexical knowledge	17	172.470	19.847
Vocabulary size	17	6033.941	1906.124
Word frequency	17	105.706	38.907
Native-like typicality of associations	17	103.118	70.013
Number of associations	17	65.471	39.030
Within-group consistency of WA domain	17	228.412	127.503

Source	*Sum of Squares*	*df*	*Mean squares*	*F-ratio*
Verified vocabulary knowledge				
Between Groups	33912.696	2	16956.348	23.82*
Error	43425.288	61	711.890	
Self-reported lexical knowledge				
Between Groups	36218.218	2	18109.109	29.68*
Error	37220.220	61	610.166	
Vocabulary size				
Between Groups	129881841.2	2	64940920.597	13.16*
Error	301121680	61	4936420.988	
Word frequency				
Between Groups	47051.777	2	23525.888	10.91*
Error	131587.161	61	2157.167	
Native-like typicality of associations				
Between Groups	837848.906	2	418924.453	61.75*
Error	413852.094	61	6784.461	
Number of associations				
Between Groups	31337.000	2	15668.500	14.86*
Error	64309.484	61	1054.254	
Within-group consistency of WA domain				
Between Groups	224596.552	2	112298.276	11.83*
Error	578910.682	61	9490.339	

$p^* < .001$

APPENDIX F

SUMMARY OF DIRECT SEARCH ON *t* REGRESSION

FOR THE GROUP OF NS OF ENGLISH

Regressor	**Ranked I t I**	**Regressor subset**	**R^2**	**Adj. R^2**	**C_p**
Self-reported vocabulary knowledge (VK)	7.302	Self-reported VK Vocabulary size Native-like typicality	.93	.93	2.01
Native-like typicality of associations	1.007	Self-reported VK Vocabulary size Native-like typicality Word frequency effects	.93	.92	4.00
Vocabulary size	.526	Self-reported VK Native-like typicality	.92	.92	4.35
Word frequency effects	-.066	Self-reported VK Vocabulary size Native-like typicality Word frequency effects Number of associations	.93	.92	6.00
Number of associations	.052	Self-reported VK	.90	.90	9.15

APPENDIX G

SUMMARY OF DIRECT SEARCH ON *t* REGRESSION

FOR THE GROUP OF L2 ADVANCED LEARNERS

Regressor	**Ranked I t I**	**Regressor subset**	**R^2**	**adj. R^2**	**C_p**
Self-reported VK	5.004	Self-reported VK Vocabulary size	.99	.99	1.33
Vocabulary size	1.418	Self-reported VK Vocabulary size Number of associations	.99	.99	2.98
Number of associations	1.119	Self-reported VK Vocabulary size Number of associations Native-like typicality	.99	.99	4.02
Native-like typicality of associations	.936	Self-reported VK Vocabulary size Number of associations Native-like typicality Word frequency effects	.99	.99	6.00
Word frequency effects	.134	Self-reported VK	.95	.95	41.61

APPENDIX H

SUMMARY OF THE DIRECT SEARCH ON *t* REGRESSION

FOR THE GROUP OF L2 INTERMEDIATE LEARNERS

Regressor	**Ranked I t I**	**Regressor subset**	**R^2**	**adj. R^2**	**C_p**
Self-reported VK	8.970	Self-reported VK Vocabulary size	.97	.97	.316
Vocabulary size	.669	Self-reported VK Vocabulary size Word frequency effects	.97	.97	2.04
Word frequency effects	.338	Self-reported VK Vocabulary size Word frequency effects Number of associations	.97	.96	4.02
Number of associations	.196	Self-reported VK Vocabulary size Word frequency effects Number of associations Native-like typicality	.97	.96	6.00
Native-like typicality of associations	.153	Self-reported VK	.93	.92	17.75

APPENDIX I

SW FREQUENCY CATEGORIES ACCORDING TO SFI

Low frequency words		**Mid frequency words**		**High frequency words**	
SFI		**SFI**		**SFI**	
20.8	rigidity	36.0	middling (adj.)	46.4	toxic
22.1	abattoir	35.5	forgo	46.4	weaken
22.1	bursar	36.6	prefect	46.6	experimentation
22.1	contravention	37.2	unnerve	47.4	shuttle (n.)
22.1	gambol (v.)	37.4	instil	48.6	residence
22.1	macaw	37.4	inception	49.7	studio
22.1	masochism	37.4	solstice	50.6	beaten
22.1	putative	38.1	penance	50.7	sweep (v)
25.1	virtuosity	38.7	promontory	51.5	diamond
25.5	amoral	39.1	griddle (n.)	51.6	telegraph (n.)
26.9	point-blank (adj.)	39.4	concede	51.7	yield (v.)
27.7	crampons	39.6	coinage	53.4	hunger
29.3	glower	40.4	fathom (v.)	54.3	blanket (n.)
30.7	tuner	41.7	naïve	54.6	ruler
32.9	entrust	41.9	custodian	55.5	official (adj.)
32.9	flagstone	41.9	drawback (n.)	56.9	throw (n.)
34.0	lackadaisical	42.2	spotlight (n.)	57.1	refuse (v.)
34.1	jib (n.)	42.6	sleigh (n.)	58.7	settled
34.9	parable	42.8	livelihood	62.4	wind (v.)
35.0	dissension	43.1	pillar (n.)	65.3	second (nmrl.)
35.0	savor (v.)	43.4	advantageous	65.8	hard (adv.)
35.4	monorail	43.8	bracelet	66.9	high (adv.)
35.5	edifice	44.9	choke (n)	70.3	back (v.)
38.4	cassava	46.8	defensive	73.7	out (n.)
		46.8	array (v.)		

APPENDIX J

ONE-WAY ANOVA SUMMARY FOR THE QUANTITATIVE CHARACTERISTICS OF NS', L2 ADVANCED, AND INTERMEDIATE LEARNERS' OF ENGLISH

STRUCTURE OF MENTAL LEXICON

Group	*N*	*Mean*	*SD*
Native Speakers			
Number of different responses	30	69.47	17.238
Common responses	30	57.07	13.266
Idiosyncratic responses	30	53.16	13.678
Total number of responses	30	110.13	21.639
The three most popular responses	30	30.07	11.57
Advanced learners			
Number of different responses	17	89.71	17.243
Common responses	17	46.82	12.254
Idiosyncratic responses	17	73.41	13.337
Total number of responses	17	120.41	21.192
The three most popular responses	17	31.47	10.500
Intermediate learners			
Number of different responses	17	48.12	15.580
Common responses	17	26.06	9.437
Idiosyncratic responses	17	39.71	13.963
Total number of responses	17	65.18	15.388
The three most popular responses	17	20.12	7.623

Source	*Sum of Squares*	*df*	*Mean squares*	*F-ratio*
Number of different responses				
Between Groups	14709.349	2	7353.174	25.99*
Within Groups	17258.761	61	282.931	
Total	31965.109	63		
Common responses				
Between Groups	10444.956	2	5222.478	35.67*
Within Groups	8931.278	61	146.414	
Total	19376.234	63		
Idiosyncratic responses				
Between Groups	9840.124	2	4920.062	26.35*
Within Groups	11391.814	61	186.751	
Total	21231.938	63		
Total number of responses				
Between Groups	30724.554	2	15362.277	38.17*
Within Groups	24554.055	61	402.525	
Total	55278.609	63		
The three most popular responses				
Between Groups	1386.493	2	693.246	6.43*
Within Groups	6581.867	61	107.899	
Total	7968.359	63		

$p^* < .001$

APPENDIX K

ONE-WAY ANOVA SUMMARY WITH MEANS AND SD FOR THE QUALITATIVE CHARACTERISTICS OF NS', ADVANCED, AND INTERMEDIATE LEARNERS' PROPORTION OF WA GIVEN TO SW PARTICIPANTS KNOW

Group	*N*	*Mean*	*SD*
Native Speakers			
Paradigmatic associations	73	56.37	30.942
Syntagmatic associations	73	42.27	30.619
Advanced learners			
Paradigmatic associations	73	52.55	30.285
Syntagmatic associations	73	46.35	30.092
Intermediate learners			
Paradigmatic associations	73	44.07	37.773
Syntagmatic associations	73	39.50	36.754

Source	*Sum of Squares*	*df*	*Mean squares*	*F-ratio*
Paradigmatic associations				
Between Groups	5779.903	2	2889.951	2.63
Within Groups	237706.17	216	1100.492	
Total	243286	218		
Syntagmatic associations				
Between Groups	1734.047	2	867.024	.81
Within Groups	229964.63	216	1064.651	
Total	231698.68	218		

$p > .05$

APPENDIX L

ONE-WAY ANOVA SUMMARY WITH MEAN AND SD FOR THE QUALITATIVE CHARACTERISTICS OF NS', ADVANCED, AND INTERMEDIATE LEARNERS' PROPORTION OF WA GIVEN TO SW PARTICIPANTS VAGUELY KNOW

Group	*N*	*Mean*	*SD*
Native Speakers			
Paradigmatic associations	73	19.58	32.597
Syntagmatic associations	73	14.74	28.724
Clang associations	73	2.71	8.823
Advanced learners			
Paradigmatic associations	73	25.96	33.675
Syntagmatic associations	73	32.96	36.98
Clang associations	73	5.49	10.497
Intermediate learners			
Paradigmatic associations	73	15.98	30.334
Syntagmatic associations	73	21.11	33.920
Clang associations	73	6.74	16.575

Source	*Sum of Squares*	*df*	*Mean squares*	*F-ratio*
Paradigmatic associations				
Between Groups	3727.680	2	1863.840	1.79
Within Groups	224405.91	216	1038.916	
Total	228133.59	218		
Syntagmatic associations				
Between Groups	12470.993	2	6235.496	5.60*
Within Groups	240695.94	216	1114.330	
Total	253166.33	218		
Clang associations				
Between Groups	615.725	2	307.862	1.99
Within Groups	33240.136	216	154.605	
Total	33855.861	218		

$p < .05$

APPENDIX M

LIST OF 12 STIMULUS WORDS CLASSIFIED ACCORDING TO FREQUENCY OF OCCURRENCE AND LEXICAL CATEGORY

Frequency	Nouns	Verbs	Adjectives
low	cassava edifice masochism rigidity	entrust gambol unnerve savor	amoral lackadaisical point-blank putative
mid	bracelet coinage drawback pillar	array concede fathom forgo	advantageous defensive middling naïve
high	blanket experimentation hunger studio	back refuse sweep weaken	beaten official toxic second

APPENDIX N

MEANS AND SD FOR THE EFFECTS OF SW LEXICAL CATEGORY AND WORD FREQUENCY ON THE QUALITATIVE CHARACTERISTICS OF THE WA DOMAIN FOR SW PARTICIPANTS KNOW

Types of responses	*Group*	*Lexical category*	*Word frequency*	*Mean*	*Standard deviation*	*Number*
Paradigmatic	NS	nouns	low	8.50	3.786	4
			mid	22.25	10.905	4
			high	29.00	10.165	4
		verbs	low	4.00	3.162	4
			mid	15.25	7.228	4
			high	8.25	5.500	4
		adjectives	low	12.25	7.089	4
			mid	17.50	10.599	4
			high	14.25	6.602	4
	Advanced learners	nouns	low	18.00	11.045	4
			mid	24.00	4.761	4
			high	27.50	1.915	4
		verbs	low	2.00	2.000	4
			mid	5.00	3.464	4
			high	19.00	14.306	4
		adjectives	low	7.25	10.595	4
			mid	11.50	6.191	4
			high	9.25	3.862	4
	Intermediate learners	nouns	low	9.25	8.098	4
			mid	11.00	7.874	4
			high	21.75	4.113	4
		verbs	low	1.25	1.258	4
			mid	1.00	2.000	4
			high	5.50	5.447	4
		adjectives	low	2.75	4.856	4
			mid	6.50	4.655	4
			high	17.50	17.597	4

Syntagmatic	NS	nouns	low	5.00	5.944	4
			mid	11.50	7.937	4
			high	16.75	9.179	4
		verbs	low	15.50	9.147	4
			mid	10.25	5.123	4
			high	33.00	5.7155	4
		adjectives	low	11.50	12.042	4
			mid	14.25	13.598	4
			high	29.00	5.477	4
	Advanced learners	nouns	low	7.75	8.421	4
			mid	9.75	2.217	4
			high	17.50	3.697	4
		verbs	low	20.25	13.598	4
			mid	9.00	2.828	4
			high	24.75	18.554	4
		adjectives	low	10.00	10.614	4
			mid	27.50	7.767	4
			high	36.75	4.787	4
	Intermediate learners	nouns	low	2.25	.957	4
			mid	1.25	1.258	4
			high	10.00	3.742	4
		verbs	low	5.00	7.439	4
			mid	2.50	3.786	4
			high	21.75	3.304	4
		adjectives	low	7.00	10.739	4
			mid	21.25	11.383	4
			high	17.25	13.074	4

APPENDIX O

ANALYSIS OF VARIANCE FOR EFFECTS OF SW LEXICAL CATEGORY AND WORD FREQUENCY ON THE QUALITATIVE CHARACTERISTICS OF THE WA DOMAIN OF STIMULI PARTICIPANTS KNOW

Source	*df*	*F* *paradigmatic*	*syntagmatic*
Lexical category (LC)	2	53.68*	23.82*
Word frequency (WF)	2	1.87	2.94
Proficiency level (PL)	2	3.75*	1.37
LC x WF	4	.85	1.87
LC x PL	4	3.56*	1.10
WF x PL	4	1.87	.76
LC x WF x PL	8	2.29	1.73
Error	81		

$p^* < .05$

APPENDIX P

MEANS AND SD FOR THE EFFECTS OF SW LEXICAL CATEGORY AND WORD FREQUENCY ON THE QUALITATIVE CHARACTERISTICS OF THE WA DOMAIN FOR SW PARTICIPANTS VAGUELY KNOW

Types of responses	*Lexical category*	*Word frequency*	*Group*	*Mean*	*Standard deviation*	*Number*
Paradigmatic	Nouns	low	NS	12.50	25.000	4
			Advanced	.00	.000	4
			Intermediate	.00	.000	4
		mid	NS	33.33	38.486	4
			Advanced	60.94	29.531	4
			Intermediate	33.33	47.140	4
		high	NS	16.67	33.330	4
			Advanced	35.00	47.258	4
			Intermediate	25.00	50.000	4
	Verbs	low	NS	8.33	16.665	4
			Advanced	19.45	31.916	4
			Intermediate	10.10	12.300	4
		mid	NS	16.67	33.330	4
			Advanced	10.83	14.240	4
			Intermediate	3.57	7.145	4
		high	NS	.00	.000	4
			Advanced	.00	.000	4
			Intermediate	.00	.000	4
	Adjectives	low	NS	.00	.000	4
			Advanced	43.75	51.539	4
			Intermediate	.00	.000	4
		mid	NS	25.00	50.000	4
			Advanced	31.50	37.014	4
			Intermediate	23.66	28.472	4
		high	NS	37.50	47.871	4

			Advanced	16.67	33.335	4
			Intermediate	.00	.000	4
Syntagmatic	Nouns	low	NS	25.00	50.000	4
			Advanced	.00	.000	4
			Intermediate	.00	.000	4
		mid	NS	16.67	19.249	4
			Advanced	30.73	35.738	4
			Intermediate	12.50	25.000	4
		high	NS	8.34	16.670	4
			Advanced	15.00	30.000	4
			Intermediate	16.67	33.335	4
	Verbs	low	NS	33.33	47.140	4
			Advanced	67.68	45.735	4
			Intermediate	65.39	31.286	4
		mid	NS	25.00	50.000	4
			Advanced	75.00	17.533	4
			Intermediate	39.29	48.620	4
		high	NS	25.00	50.000	4
			Advanced	75.00	50.000	4
			Intermediate	25.00	50.000	4
	Adjectives	low	NS	.00	.000	4
			Advanced	22.92	31.459	4
			Intermediate	14.29	28.570	4
		mid	NS	.00	.000	4
			Advanced	38.96	42.772	4
			Intermediate	26.34	31.454	4
		high	NS	12.50	25.000	4
			Advanced	25.00	31.915	4
			Intermediate	.00	.000	4
Clang	Nouns	low	NS	12.50	25.000	4
			Advanced	.00	.000	4
			Intermediate	.00	.000	4
		mid	NS	.00	.000	4
			Advanced	8.33	16.665	4

		Intermediate	29.17	34.360	4
	high	NS	.00	.000	4
		Advanced	.00	.000	4
		Intermediate	8.33	16.665	4
Verbs	low	NS	8.34	16.670	4
		Advanced	12.88	16.105	4
		Intermediate	24.52	20.435	4
	mid	NS	8.34	16.670	4
		Advanced	14.17	13.976	4
		Intermediate	7.14	14.285	4
	high	NS	.00	.000	4
		Advanced	.00	.000	4
		Intermediate	.00	.000	4
Adjectives	low	NS	.00	.000	4
		Advanced	8.33	16.665	4
		Intermediate	10.72	21.430	4
	mid	NS	.00	.000	4
		Advanced	4.55	9.090	4
		Intermediate	.00	.000	4
	high	NS	.00	.000	4
		Advanced	8.33	16.665	4
		Intermediate	.00	.000	4

APPENDIX Q

ANALYSIS OF VARIANCE FOR EFFECTS OF SW LEXICAL CATEGORY AND WORD FREQUENCY ON THE QUALITATIVE CHARACTERISTICS OF THE WA DOMAIN OF STIMULI PARTICIPANTS VAGUELY KNOW

		F		
Source	*df*	*paradigmatic*	*syntagmatic*	*clang*
Lexical category (LC)	2	2.87	11.39*	1.13
Word frequency (WF)	2	2.76	.37	2.65
Proficiency level (PL)	2	1.84	4.31*	1.52
LC x WF	4	1.61	.40	1.66
LC x PL	4	.30	1.41	.73
WF x PL	4	.20	.55	.28
LC x WF x PL	8	.91	.42	1.49
Error	81			

$p^* < .05$

APPENDIX R

ANALYSIS OF VARIANCE FOR EFFECTS OF SW LEXICAL CATEGORY AND WORD FREQUENCY ON THE QUANTITATIVE CHARACTERISTICS OF THE WA DOMAIN OF STIMULI PARTICIPANTS KNOW

		F				
Source	*df*	*Different*	*Common*	*Idiosyncratic*	*Total number*	*Three most popular*
Lexical category (LC)	2	3.61*	6.95*	3.52*	4.55*	3.81*
Word frequency (WF)	2	28.28*	32.53*	18.09*	40.24*	17.03*
Proficiency level (PL)	2	18.24*	5.56*	17.05*	17.58*	3.53*
LC x WF	4	4.32*	3.04*	5.78*	2.33	2.10
LC x PL	4	.73	.34	.61	.77	.62
WF x PL	4	.84	.64	1.00	.53	.31
LC x WF x PL	8	.92	.77	.70	.99	.50
Error	81					

$p^* < .05$

APPENDIX S

MEANS AND SD FOR THE EFFECTS OF SW LEXICAL CATEGORY AND WORD FREQUENCY ON THE QUANTITATIVE CHARACTERISTICS OF THE WA DOMAIN FOR SW PARTICIPANTS VAGUELY KNOW

Type of responses	*Group*	*Lexical category*	*Word frequency*	*Mean*	*Standard deviation*	*Number*
number of associations	NS	nouns	low	1.00	1.155	4
			mid	3.00	4.243	4
			high	.75	1.500	4
		verbs	low	2.00	1.414	4
			mid	2.00	2.449	4
			high	.75	1.500	4
		adjectives	low	.00	.000	4
			mid	.75	1.500	4
			high	2.25	2.872	4
	Advanced	nouns	low	1.00	1.414	4
			mid	5.50	7.141	4
			high	1.25	2.500	4
		verbs	low	7.25	3.500	4
			mid	7.75	5.852	4
			high	2.50	2.646	4
		adjectives	low	1.00	1.414	4
			mid	4.50	3.873	4
			high	1.50	1.732	4
	Intermediate	nouns	low	.00	.000	4
			mid	2.50	2.082	4
			high	1.00	1.414	4
		verbs	low	5.00	5.598	4
			mid	2.50	3.3167	4
			high	.50	1.000	4
		adjectives	low	.00	.000	4
			mid	2.75	3.775	4
			high	.00	.000	4