INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

Four Levels of Native-likeness: A New Method of Assessing Learner Lexicons

Sumanthra Govender

A Thesis

In

The TESL Centre

Presented in Partial Fulfilment of the Requirements
for a Master of Arts in Applied Linguistics at
Concordia University
Montreal, Quebec, Canada

April 2002

ABSTRACT

Four Levels of Native-likeness:

A New Method of Assessing Learner Lexicons

Sumanthra Govender

This is a cross-sectional study that focuses on the assessment of the semantic network in the mental lexicon of second language learners. Generally, the number of words a learner claims to know is a reflection of achievement and advances in his vocabulary knowledge. However, the size of a learner's vocabulary knowledge only reveals half of the story about his lexical competence. An understanding of the quality of word knowledge allows for a clearer interpretation of the organisation and integration of lexical items in the mental lexicon. Typically, vocabulary researchers have utilised word association tests to measure vocabulary depth. However, there are several problems with the traditional methodology of this test, which has resulted in some inaccurate claims about a non-native speaker's lexical ability, such as the dichotomous native-like versus non-native-like categorisation of lexical ability. The aim of this thesis study was to test an alternative word association test that claims to capture partial native-like lexical knowledge in non-native speakers. Up to this point, this method had only been developed and tested on native speakers of English. The results of the analysis in this study reveal that this test does capture a range of native-

like levels of lexical ability in non-native speakers of English. The findings suggest that this alternate word association measure of vocabulary depth could be used in conjunction with other language and vocabulary measures for both diagnostic and placement purposes. The information gained from this study provides a new approach to and understanding of vocabulary acquisition research and assessment.

Acknowledgement

So many people were involved in making this thesis a success. First, I would like to thank my supervisors, Dr. Joanna White and Dr. Marlise Horst, for their help in planning and carrying out this thesis. In addition to their practical and moral support, their enthusiasm and dedication to language research inspire me. I would also like to thank Randall Halter for guiding me through what I feel is the "confusing world" of statistics.

I am indebted to Eric Buisson, CELDT/ESL Coordinator, for spending countless hours trying to find CELDT scores for me. I am grateful to Roberto Chen-Rangel, Graduate Program Coordinator, for making sure I always had the right paper work complete. A tremendous thanks goes to all the ESL teachers who kindly allowed me to steal an hour of their class time so I could test their students, and I would also like to thank all the students who participated in this study, which made it all possible.

Last, but certainly not least, I would like to thank my family, friends and fellow classmates for their support and encouragement throughout this past year, especially during the times when it felt as though this ride was more than I could handle. To all of you, I am grateful for your confidence and love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

1.1 VOCABULARY IN SECOND LANGUAGE ACQUISITION RESEARCH

"Vocabulary is central to language", according to Zimmerman (1997, cited in Qian, 1999). However, for many years vocabulary acquisition was largely a "neglected aspect" of language acquisition research, and much of the research that did exist was mainly theoretical, brief, and incomplete (Meara, 1981). This occurred because learners' language proficiency was defined only in terms of their communicative skills and abilities, rather than their knowledge of the meanings of the words used to communicate in the second language. However, by the nineties, a renewed interest in vocabulary occurred in the field of second language acquisition, and several second language acquisition researchers cited very practical reasons why they should focus on vocabulary in addition to other areas of language development (Laufer, 1986; Levenston, 1979; & McCarthy, 1984). The most important reason for this interest was that lexical knowledge has often been described as a fundamental aspect of second language competence.

Many teachers and learners are likely to say that possessing a good vocabulary leads to more efficient communication and comprehension. In fact, students often comment that the primary source of difficulty and frustration in language learning is insufficient vocabulary knowledge (Nation, 1990). Therefore, it is not difficult to conclude that lexical knowledge is an essential component of second language proficiency because vocabulary acquisition plays an active role in the acquisition of a second language. In order to get a better understanding of

the role of vocabulary knowledge in language acquisition research, we will examine how word knowledge is described and measured. We will begin with what it means for someone to possess word knowledge.

1.1.1. Word Knowledge

Traditionally, people claim to know a word when they are able to recognise its form and possess an understanding of its meaning. However, knowing a word entails more than just familiarity with its form and meaning. According to Nation (1990) and Richards (1976), there are "various kinds of word knowledge necessary to master a word completely, including knowledge of its orthographical and phonological form, meanings, grammatical behaviour, associations, collocations, frequency, and register"(Schmitt & McMarthy, 1997, p. 4). Therefore, word knowledge is based on a combination of these elements, and each aspect of word knowledge has its own independent function in its own right.

In order to attain a better understanding of what it means to have "lexical competence", word knowledge should be viewed as falling under two broader dimensions of lexical comprehension: breadth and depth. Research into the dimension of vocabulary breadth deals with the size of the person's mental lexicon; in other words, how many words are known. Research into vocabulary depth focuses on the quality of word knowledge in the mental lexicon; in other words, how well words are known. In general, learners with large vocabularies tend to be more capable in a broader range of language skills than learners with smaller vocabularies. Furthermore, as a learner becomes more proficient in the

language, the dimension of size becomes less important, while the dimension of quality increases in importance (Meara, 1996).

Studying these two interconnected and, to some extent, substantially interdependent dimensions of vocabulary knowledge informs language researchers about how words are acquired and retained in the mental lexicon (Qian, 1999; Schmitt & Meara,1997; Schmitt, 2000), as well as how lexical items are linked together, and if these links change with language proficiency (Read, 1993). It also provides greater insight into the nature of semantic links between words in the mental lexicon.

Today, many language acquisition researchers focus more attention on vocabulary knowledge in an attempt to reconceptualise its role in language acquisition within a broader framework of communicative lexical ability (Schmitt, 1997). The knowledge gained from these investigations has resulted in a clearer understanding of lexical acquisition and the organisation of the mental lexicon, and identified helpful teaching techniques. Furthermore, researchers have investigated whether the acquisition and storage of lexical items is different for native speakers and language learners. The general conclusion is that word knowledge is acquired gradually and incrementally with higher levels of language proficiency for both native speakers and non-native speakers of a language (Meara & Schmitt, 1997; Read, 1993; Schmitt, 1998, 2000). In the following section, research supporting this conclusion is reviewed. First to be considered are the recently developed vocabulary tests.

1.2. Assessing Vocabulary Knowledge

1.2.1 Reasons for Vocabulary Assessment

If vocabulary is considered to be an essential component of language development, then it needs to be assessed to determine how well the language and its lexical items are being acquired, integrated, and retained in the mental lexicon. Schmitt (2000) notes that teachers typically use vocabulary tests as an achievement, diagnostic, or placement measure. Teachers want to find out if their students have learned the words taught in class, or they want to know their students' lexical deficiencies, in order to pay specific attention to their problem areas. As a placement measure, vocabulary tests can be used to place learners into an appropriate class level. Meara (1996) points out that assessing the organisation of lexical items in the mental lexicon is "a useful way of distinguishing between learners at different levels of proficiency"(Meara 1996, p. 48). The type of vocabulary test utilised by vocabulary researchers or teachers depends on which dimension of word knowledge needs to be assessed.

1.2.2. Measures of Vocabulary Knowledge

Many different tests have been designed and used to draw conclusions about the dimensions of vocabulary knowledge and the acquisition process. In exploring the breadth of vocabulary knowledge, a number of measurements exist that allow researchers to closely examine the size of a speaker's vocabulary. The Eurocentres Vocabulary Size Test, developed by Meara and Jones (1988), is a computerised *yes/no* checklist that samples vocabulary frequency zones. This

vocabulary list contains real words and nonsense words. A language learner must indicate whether he knows the word with a check. If he records knowing some of the non-words, this would illustrate that he is over-estimating his lexical knowledge, and his score would be adjusted downwards accordingly.

Another breadth test is the Vocabulary Levels Test designed by Nation (1990) and recently updated by Schmitt (2000). In this test, which also samples frequency zones, a learner is presented with a list of words and short definitions. The stimulus words are presented in isolation, and the definitions are usually synonyms or short phrases. A learner must match the stimulus word with the appropriate definition. Many teachers use this test in their language classes as a quick and practical way of profiling the learners' receptive knowledge. This sketch of learners' receptive knowledge aids in planning vocabulary teaching and designing a learning programme for the whole class or for individual students.

Another measure of vocabulary size can be obtained by administering the Lexical Frequency Profile designed by Laufer and Nation (1995). This test profiles a learner's productive vocabulary knowledge by focusing on the types of words he uses in short compositions. A computer program classifies correctly used words into four frequency levels. A learner who uses a high percentage of low frequency words in his writing is considered to have a large vocabulary size and a high level of language proficiency.

Thus, there are a variety of tests suited to assessing vocabulary breadth. In contrast, measures of vocabulary depth are less well developed. The main depth tool utilised by researchers at the moment is the Vocabulary Knowledge

Scale (Wesche and Paribakht, 1996). This is a selective vocabulary test with words presented in isolation. The testee rates his knowledge of each stimulus word according to a five-point scale of familiarity, which ranges from giving an explanation of the word's meaning to composing a sentence containing the word. This tool provides a clear illustration of a learner's understanding of the stimulus words, but it does not provide information about how words are organised in the lexicon. It is accepted among researchers that the human lexicon is a well-organised network of associations, but they also disagree as to how it is organised and how to explore it (Aitchison, 1994). An instrument that has been used by many researchers to investigate lexical organisation and associational links between lexical items is the word association test. In the next chapter, word association tests will be examined in detail.

CHAPTER 2: WORD ASSOCIATION TESTS

2.1. BACKGROUND INFORMATION

The word association test was primarily used in the field of psychology in the 1960's and 1970's to assess cognitive and behavioural development in individuals. Around the same time, this test began to be used in language research and continues to be used today, in order to explore the "human word-web" and the way words are linked together in the mind (Aitchison, 1994). Figure 1. 1. shows an illustration of a hypothetical "human word-web", or semantic network, for the word *colour*. The darker lines indicate a stronger associational link between the words. In first language acquisition research, this kind of test has been used to determine the association stereotypy of native speakers of a language. For example, mentally healthy speakers of English tend to say *cat* in response to the stimulus word *dog*. In second language acquisition research, this test has been used to determine whether non-native speakers produce association responses similar to those produced by native speakers (Aitchison, 1994; Kruse et al. 1987; Meara, 1980, 1983; Read, 1993, 1998; Sökmen, 1993).

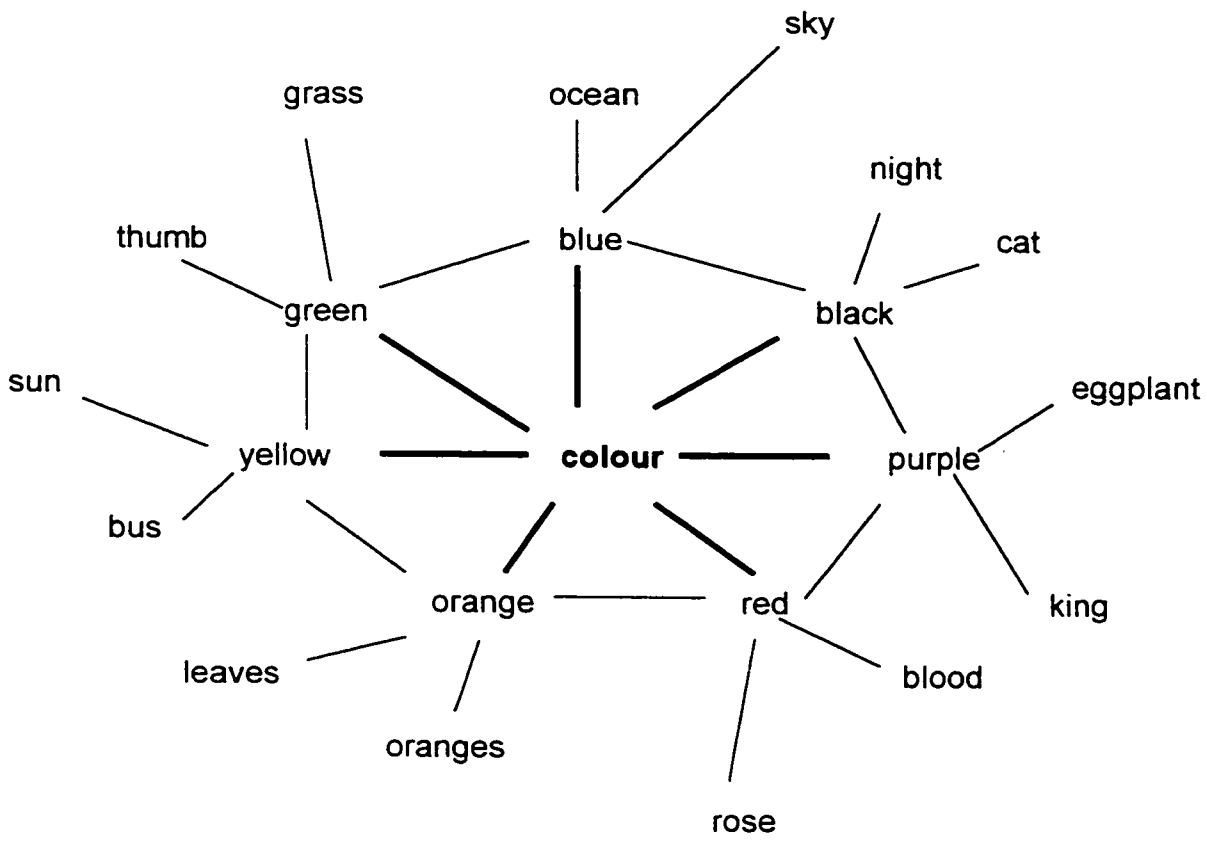


Figure 1.1. Hypothetical semantic network for *colour*

2.1.1. Basic Methodology of the Word Association Test

The basic word association test is relatively simple to use and produces a wealth of data with minimal effort. This test involves presenting a group of participants with a list of stimulus words, to which the participants must produce a single response for each stimulus word. Typically, respondents are instructed to provide the ***first response that comes to mind***. It is assumed that this single response will reveal the strongest mental links between words in the mental lexicon (Aitchison, 1994; Meara, 1980, 1983). In recent years, researchers have modified the format of the test in an attempt to describe the semantic networking of a learner's overall lexicon and not only the individual words within the network.

Read (1993) created the Word Associates Format. In the sample test item shown below, the testee is presented with a stimulus word and eight possible associates and is required to identify the four words that are semantically related to the stimulus word:

edit

arithmetic	film	pole	publishing
revise	risk	surface	text

The Association Vocabulary Test developed by Vives Boix (1995, cited in Schmitt, 2000) is another version of the word association test. The aim of this test is to measure the degree of lexical organisation in the mind, rather than the number of the words known well. As illustrated in the following example, the testee is presented with a number of three-word clusters. In each cluster, two

words are semantically associated and the other is not. The testee must indicate the unrelated word:

creciente (growing)	*veneno* (poison)	*pócima* (nasty drink)

Meara and Fitzpatrick (2000) designed another modified version of the word association test called the Lex 30. This association test was designed to be a simple task in which an index of a learner's productive vocabulary can be determined by tapping into the extent of his lexical knowledge. The testee is required to produce a series of written responses, at least three, for each stimulus word. The response data are analysed for instances of low frequency words. Some of the stimulus words are listed below:

attack *disease* *furniture*

2.1.2. Reasons to use a Word Association Test

An important use of the word association test is to augment vocabulary-teaching techniques. Sökmen (1993) concluded that vocabulary teaching is more effective when learners reflect on personal experiences with new words in order to develop a personal lexical inventory. For instance, she recommends that teachers use project-oriented or communicative activities as a learning technique in which new learned words become emotionally associated with each other. In addition, she stresses that teachers should encourage students to build associations by brainstorming rather than providing them with a definition only. A

student can then be tested on the classification or the coordinate clustering among words, as a means of indicating his understanding of the word and demonstrating the extent of his semantic networking.

Palmberg (1990) used word association findings to stress the necessity of "native-based" exercises in vocabulary teaching. These exercises should be aimed at allowing language learners to incorporate new words into their lexical stock in the same gradual and meaningful manner as native speakers acquire their first language. In other words, he suggests that teachers should train learners to direct their lexical associations more closely to those of native speakers. This can be achieved by using vocabulary activities that focus on word relationships, collocations, as well as activities that focus a learner's attention on discriminating between lexical sets.

The word association test has also allowed researchers to discover the nature of the associative patterning of native speakers' and language learners' responses. Native speakers predominantly produce clang and/or syntagmatic associations in the early stages of language development. Clang associations are responses related to certain phonological features of the stimulus word but bear no obvious semantic relation to the prompt word, for example:

dog log fog

Syntagmatic responses are associative responses that form an obvious sequential link with the stimulus word, such as:

dog bark spot bite

Gradually over time and with gained word knowledge, native speaker associative responses become more paradigmatic in nature. Paradigmatic associations are responses from the same grammatical class as the stimulus word, such as:

dog cat animal

In comparing the associative responses of language learners with those of native speakers, second language researchers have determined that the shift in associative patterning of language learners is similar to that of a child native speaker acquiring their L1 (Meara, 1983; Schmitt, 2000; Söderman, 1993; Sökmen, 1993).

2.1.3. Results of Word Association Tests

As stated earlier in section 2.1.1., there are several modified versions of the word association test. Regardless of the format used, the results from word association tests reveal that native speakers of a language have a sophisticated and stable pattern of semantic networking. In other words, native speakers produce similar association responses to stimulus words, and this kind of knowledge is acquired in gradual incremental steps. These common native responses are compiled to create a norm list of responses. Two well-known

examples are Postman and Keppel (1970) and the Minnesota Word Association Norm (1952). These and other associative norm lists are used in second language research as comparative measures of non-native speaker association responses to determine whether language learners produce typical native associative responses to stimulus words.

This comparison generally shows that the semantic patterning of a language learner is rather unstable at lower levels of proficiency (Meara, 1983). This means that his responses are more diverse and less homogeneous than native speaker association responses. Yet, as a language learner progresses in language proficiency, the association patterning develops toward the native norms (Read, 1993). This development in response types for a language learner is not only a function of overall proficiency, but also a function of the development of individual words. That is, each word passes through a different stage of development for each learner (Söderman, 1993).

Therefore, since non-native speakers' association patterning appears to develop towards native speaker norms, it can be assumed that non-native associative ability may be categorised according to various levels of native-like ability. These would clearly illustrate the progression towards the native norms with increased language proficiency. However, language learners have not been categorised according to this progression in word association experiments so far (Schmitt, 1998, 2000). Instead, a rather simplistic categorisation has been used to describe the native-likeness of a learner's associative ability (Schmitt, 1998; Sökmen, 1993). A learner's word knowledge is classified as being native-like,

typical of the native norm, or non-native-like, not typical of the native norm. In other words, the learner might or might not be demonstrating normative associative performance. Furthermore, this native-like versus non-native-like dichotomy is established with little consideration for partial native-like ability. When partial native-like lexical ability is evident, a learner is considered to exhibit non-native-like ability, because he does not produce typical native associations. This all-or-nothing categorisation is the outcome of the traditional methodology used to score word association tests (Schmitt, 1998). So, while there are many benefits to the traditional method of testing word associations, there are also concerns that this method may lead to an inaccurate account of a language learner's developing native-like ability (Meara, 1983; Read, 1993; Schmitt, 1998). Problems associated with word association test are addressed in the next section.

2.1.4. Weaknesses of Word Association Tests

One problematic aspect of the word association test is the selection of the stimulus words (Meara, 1983). Usually high frequency words, such as *black, dog, hot*, are selected as stimuli. These words are among the first words a learner, native or non-native, acquires when learning a language. For this reason, high frequency words are used to insure that all testees involved in the study know the words, or at least have some knowledge about the words. However, there are three problems with using high frequency stimulus words.

First, the use of these words tends to lead to the production of very similar responses in both the native language and the target language of a language learner (Meara, 1983). This will typically occur in situations where cognates exist between languages. Cognates are words that come from the same linguistic family or derivation. English and French share many homographic cognates. These are words with a similar spelling, such as *construction (English) – construction (French)*. If the native language and target language share cognates, it will be difficult to determine whether or not word knowledge in the second language is actually being tested because the L2 words might be treated as L1 items (Meara, 1983, 1993).

Second, the use of high frequency stimulus words only focuses on the centre of a learner's L2 vocabulary stock, with little concentration on the surrounding areas where new words are being acquired and integrated into the lexicon (Meara, 1983). Wolter (2001) provides a clear description of word knowledge with his Depth of Individual Word Knowledge Model (DIWK) (Appendix A). According to this model, the mental lexicon consists of a core vocabulary of well-known words and several peripheral layers of varying degrees of lexical knowledge. High frequency words, such as *black, dog,* and *hot* are contained in the core area. Since words in this area are well-known, the probability of testing any variation in a learner's word stock is lessened because the chance of producing the typical native response is high. However, by testing words that exist in the outlying areas of the centre of word knowledge, a better indication of how new words are assimilating with other words, already existing in

the lexical stock, can be obtained (Meara, 1983; Schmitt, 1998, 2000; Wolter, 2001). In addition, investigating words in the surrounding layers of word knowledge will decrease the likelihood of accidentally testing the first language instead of the second language. Finally, it is almost impossible to determine the typicality of high frequency words because many of these words have a variety of meanings and uses. For example, *bug*, can mean an insect or a secret listening device (Read, 1993; Schmitt, 1997).

Another shortcoming of the traditional word association test is that the creation of a norm list does not take into account atypical associative responses produced by native speakers (Schmitt, 1998). In other words, uncommon native associative responses are omitted from the norm list because these are not responses typically produced by native speakers. For example, if the stimulus word is *dog* and a native speaker responds with *onion,* while a more typical response might be *cat*, the atypical response *onion* will be excluded from the norm list because only one person gave such a response. Consequently, if a language learner produces the same atypical response for *dog*, he is classified as having a non-native-like understanding of the word because the response given was not on the norm list, even though it was produced by a native speaker. Furthermore, many of the established norm lists are quite old, based on single response elicitations, on children's responses, or a combination of these.

An additional drawback to the traditional word association test arises from the use of a single word response method (Meara and Fitzpatrick, 2000; Schmitt, 1998). A single word response is believed to reveal the strongest mental

connection between the words in the respondent's mind (Aitchison, 1994). For example, if the stimulus word is *black* and it generates the response *white* for a learner, then this is assumed to be the strongest semantic connection for the word *black* in his mental lexicon. Generally, it is assumed that respondents will produce the most typical response for each stimulus word. Therefore, if a learner produces the typical response, he is said to have native-like ability for that word.

Conversely, if a learner produces an alternative response, he is deemed to have non-native-like ability for that word. While the chance of a learner producing the most typical response is high for contrasts (word opposites such as *hot/cold*, and *man/woman*), it is unlikely to occur otherwise. In the case of words like *bear*, a language learner has more of a tendency to produce an idiosyncratic response, such as *hug,* or a clang association, such as *hair* or *fare,* before thinking of a more typical associative response such as *animal* (Aitchison, 1994; Meara, 1983; Schmitt, 1998, 2000). As a result, an uncommon single word response produced by a learner is considered non-native-like. This response method is unsatisfactory because it does not allow alternate responses to be accepted. In addition, it does not allow a learner to demonstrate a progression of thought, from an uncommon response to a common response.

A final weakness with the traditional word association test is its inaccuracy in the native-like versus non-native-like distinction. As stated earlier, non-native speaker responses are usually classified as native-like or non-native-like; in other words, their association responses might or might not be representative of the native speaker association norm. The problem here is that

non-native speakers, demonstrating partial or varying degrees of normative association behaviour, are assumed to be exhibiting non-native-like associative ability. The reason for this generalised assumption is that there is no accurate way of determining whether a second language learner's associative responses are native-like or not because there is no principled manner in deciding whether norming respondent's associations are reasonable (Schmitt, 1998). Typically, native responses are tallied and compiled to create a norming list with little consideration for true representation of native knowledge. If a native speaker happens to associate *apple* with *kangaroo*, this is considered 'legitimate' simply because it was produced. Therefore, there is no established method in which native responses can be used to claim overall native-like ability. As a result, an over-generalised assumption that all norm responses are "natural" exists.

Overall, word association testing has led to many useful conclusions about semantic networking, which has led to a better understanding of the mental lexicon and improved vocabulary teaching techniques. However, the drawbacks presented above bring these conclusions into question. The most questionable claim is the native-like versus non-native-like dichotomy, which is based on word association results that fail to capture a learner's lexical progression towards native-like ability. The next section presents a way of dealing with this shortcoming.

2.2. A VARIATION OF THE WORD ASSOCIATION TEST

Taking into consideration the strengths of the traditional word association method and correcting the weaknesses of this test, Schmitt (1998) proposed an alternate method of quantifying word association responses, which describes a learner's word knowledge according to established degrees of native-likeness. He recommended three changes to the traditional methodology, resulting in a four-level descriptive scale of native-like ability. This alternate method is predicted to capture a learner's native-like ability more accurately. In other words, a learner has a greater opportunity to demonstrate a degree of normative performance, no matter how small the effect.

2.2.1 Modifications of the Test

The first modification consists of using low frequency stimulus words as a means of testing a learner's full lexical knowledge. These words are selected for their degree of polysemy. The more meanings a word has, the greater the opportunity a learner has to produce a typical response because he is likely to know, or at least have some knowledge of, one of the meanings. The second modification requires the adoption of a multiple response method. As stated earlier, the single word response does not appropriately capture the richness of the respondent's association web (Aitchison, 1994; Meara, 1983). A multiple word response method allows a learner to demonstrate the extent of his lexical knowledge of the stimulus words in greater detail, by giving him the chance to produce a typical native response for each stimulus word.

The final change involves creating a method of weighting results to give greater validity to the responses given by a language learner. In the traditional method of word association testing, there is no standardised approach to establishing weighted responses given by native speakers. Generally, the most common response produced by a group of native speakers is considered the norm response. This typical response is then deemed the best response to give, and all other responses are considered atypical. However, if responses were weighted based on strength of their typicality, then uncommon responses produced by a language learner might reflect some degree of native-likeness, no matter how minimal the typicality of the responses. For example, according to Schmitt's 1998 study, the top three native responses given for the stimulus word *brood* were *think (33), sulk (22),* and *chicken (20).* The numbers in the parentheses represent the number of native English speakers that gave that particular response to the stimulus word, out of the total number of participants tested. In other words, out of the100 native speakers that were tested, 33 of them gave the response *think* for the stimulus word *brood,* 22 produced *sulk*, and 20 produced *chicken.* Producing these top three responses would yield a maximum score of 75 (33+22+20 = 75). This is the maximum score of typicality for this particular stimulus word.

Now if a language learner produces the following responses, *book (1), wonder (2),* and *pigs (1)* to the same stimulus word, his responses must be compared to the native speaker norm list to determine how "native-like" they are. Again, the number in the parentheses represents the number of native speakers

that produced that particular response. The maximum typicality score for the learner's responses would be 4 (1+2+1 =4). When this number is divided by the native speaker maximum typicality score (4÷75= 0.053), the result, 0.053, is the strength of the typicality of the learner's associative responses based on the native speaker norms. In other words, 0.053 represents what proportion of all native associative responses this learner has produced for this particular stimulus word. In this hypothetical instance, the learner's typicality is very low.

2.2.2 Schmitt's Study

Schmitt (1998) proposed the method of quantification described above with the intended goal of creating a new norming list and native-likeness scale. He presented 100 native English speakers, of different educational backgrounds, with 17 stimulus words, all of which are low frequency polysemous words. Eleven of the words were selected from the University Word List (Schmitt, 1998). This word list consists of about 800 words which frequently occur in academic texts. The other six words were taken from Schmitt's 1999 study on TOEFL vocabulary items. The native respondents were required to produce three written single-word responses per stimulus word. Using descriptive statistics, Schmitt created a norming list that encompasses both typical and atypical association responses produced by native speakers. This was achieved by taking into account the frequency with which each response was given for each stimulus word by every native speaker.

The result is a four-level description of native-likeness. At level 0, the association proportion is 0, and there are no native-like association responses produced. At level 1, the association proportion approaches 0 but is less than the threshold proportion. Threshold is defined as the level of typicality with which a learner is considered to have a particular level of native-like ability. At this level, one or more association responses, which appear on the norming list, are produced but are not typical responses of the norming group. At level 2, the association proportion approaches the threshold proportion but is less than the native-like mean proportion. With responses at level 3, the association proportion approaches or is equal to the mean association proportion, with native-like productive association performance that is similar to that of the top portion of the native norming group. Level 3 contains more of the most commonly given responses than level 2 (Appendix B).

While this modified methodology is assumed to provide an enhanced way of incorporating word associations into future investigations of vocabulary learning and assessment, it still needs to be tested on language learners to determine whether or not it accurately describes a language learner's native-like ability. This is what this thesis study set out to do.

2.3. RESEARCH STUDY

2.3.1. Research Questions

The general research question addressed in the study was the following: Does Schmitt's (1998) proposed method of quantifying word association responses accurately demonstrate what it claims?

The following specific questions were addressed:

1. Does Schmitt's proposed method of quantifying word association responses capture a range of native-like lexical ability among non-native speakers of English?

2. If so, does this diagnostic measure of vocabulary depth relate to measures of general language proficiency and vocabulary breadth?

3. As a placement measure, does this test predict a non-native speaker's placement level in an ESL class?

It was hypothesized that this measure:

1. would allow learners to demonstrate a range of native-like lexical ability.
2. would relate to a learner's general language proficiency and vocabulary breadth.

3. would also predict a non-native speaker's placement level in an ESL class, and therefore could be used in conjunction with other placement measures for ESL class placement.

This was an exploratory cross-sectional study, utilising correlation statistics to provide a detailed account of the relationship between learners' level of native-like associative lexical ability and their level of language proficiency, vocabulary size, and ESL class level. This hypothesis would be confirmed if a positive correlation that is significant at $p \leq 0.05$ was found, thus revealing a level of native-like ability in learners' word association responses that strongly correlates with their language proficiency, vocabulary size, and class placement. In other words, learners' language proficiency and vocabulary breadth might be a determinant of the native-like quality of their vocabulary knowledge. Furthermore, learners' level of native-like lexical ability could be used as another determinant for their ESL class placement. These results would prove useful for researchers and teachers who wish to use word association tests for diagnostic and placement purposes.

CHAPTER 3. RESEARCH DESIGN

This was a quantitative study, using descriptive and correlation statistics, to investigate the relationship between degrees of native-like associations and language proficiency, vocabulary breadth, and class placement.

3.1. PARTICIPANTS

3.1.1. The ESL Group

Participants in the ESL group were 153 undergraduate students at Concordia University in Montreal. They were all non-native speakers of English from various ethnic and language backgrounds in the academic-based credit ESL program. The students were from class sections at one of four different levels based on language proficiency: pre-intermediate (ESL 298B), intermediate (ESL 207), upper-intermediate (ESL 208), and advanced (ESL 209). Generally, each class section has between 20 and 25 students. Students are placed in these classes based on their CELDT (Concordia English Language Diagnostic Test) placement test results. This is a proficiency test consisting of multiple-choice questions and an essay task used to place students in appropriate levels of English language courses. All non-native speaking students wishing to pursue undergraduate studies at Concordia are required to take the CELDT as part of their admission.

The ESL 207, 208, and 209 courses focus on developing an ESL learner's writing skills for academic purposes, while ESL 298B is a special topic course for

those students who did not pass the entry level for ESL 207. The assumption in this study is that learners in ESL 298B are lower than learners in ESL 207. At the time this study was conducted, ESL 298B was an academic vocabulary development course. The focus of this course was on lexical items from the University Word List, domain specific vocabulary, and acquisition and comprehension strategies. It was designed to develop learners' mastery of academic vocabulary. After completing ESL 298B, the learners take the CELDT test again.

While all 153 students were tested, a final sample of 145 was used because eight students were deleted from the analysis due to missing CELDT scores. The majority of the students have been learning English as a second language for at least five years in both academic and non-academic ESL environments. Their ages ranged from 19 to 55 years. The group was fairly heterogeneous with respect to the participants' first language. The 153 students belonged to twenty different language groups, the three largest groups being speakers of Chinese (84), Arabic (13), and Bengali (9). Appendix C presents a list of the language backgrounds of the non-native speakers in this study.

3.2. INSTRUMENTS

To investigate the first hypothesis of this study, each participant was administered the 17-stimulus word association test from Schmitt's 1998 study. Participants were required to provide three written single word responses for each stimulus word (Appendix D). For example:

plot _____ _____ _____

The word association instrument was originally pilot tested on an intermediate (ESL 207) class to insure the validity of the test design and to determine the length of time the participants needed to complete the test.

To investigate the hypothesis about the relationship between vocabulary depth, language proficiency and vocabulary breadth, the learners' CELDT scores and Vocabulary Levels Test scores were used. The CELDT had been administered previously at each learner's point of entry to the ESL program and was not re-administered at the time this study was conducted. The learners' test scores were used as an indicator of their overall language proficiency. The Vocabulary Levels Test revised by Schmitt (2000), which measures the size of the participants' receptive vocabulary, was administered along with the word association test (Appendix E). This levels test was slightly modified by omitting the 10,000 word level as a time saving factor. Pilot testing showed that omitting this part of the test did not change the usefulness of the results that were obtained. The following is a sample question from the levels test:

1. bulb
2. document
3. legion
4. mare
5. pulse
6. tub

_____ female horse
_____ large group of soldiers or people
_____ a paper that provides information

It was predicted that the results from these two measures would correlate with the results from the word association test, and that the findings would illustrate that a higher level of vocabulary knowledge corresponds to higher levels of language proficiency and a larger lexical stock.

Both the word association and the levels test were administered to intact ESL classes. The participants were given 25 minutes to complete the word association test and 35 minutes to complete the Vocabulary Levels Test. All participants were also required to provide personal information about their age, ethnicity, and first language and English education background (Appendix D). It was thought that this information might eventually be used to help interpret findings. Learners' personal information has remained confidential, and all participants were informed that they were allowed to discontinue the study at any time.

3.3. PROCEDURE

3.3.1 The Scoring Method

To answer the question concerning the participants' level of native-like ability, the learners' association responses were calculated and then compared to the norming results established in Schmitt's 1998 paper. Several different calculations were considered and analysed in order to interpret the participants' native-like ability. First, an association proportion for each participant was calculated by finding the sum of the native-speaker frequencies of the three responses these non-native speakers gave for each stimulus word. In other

words, this score was determined by comparing the participant's responses to the responses given by the native speakers. In the following hypothetical example, suppose 10 native speakers were given this word association test, and the following were their responses to the stimulus word *brood:*

Table 1.

Possible Native Speaker Norm Responses for the Stimulus Word ***brood***

brood	birds (2)	pigs
	baby (9)	wonder (2)
	fret (2)	eggs (4)
	family (8)	group (2)

The maximum typicality score for this stimulus word would be the sum of the top three responses produced. In this case, the top three responses are *baby (9), family (8),* and *eggs (4).* The number in the parentheses indicates the number of native speakers that produced these responses. Therefore the maximum typicality score for this word would be *21 (9+8+4= 21).*

Now suppose a non-native speaker produced the following responses to the word *brood:*

Table 2.

Possible Non-native Speaker's Responses for the Stimulus Word ***brood***

Brood	blood (0)	born (0)	group (2)

This learner's responses would be compared to the native speaker norm list of responses to determine what their maximum association score would be. In this case, the learner produced only one word on the native norm list (*group* with a response tally of *2*); therefore, his maximum score would be *2 (0 for blood + 0 for born + 2 for group = 2)*.

Second, in order to determine the strength of typicality of the learner's association responses, his personal maximum score for this stimulus word was divided by the overall native-speaker maximum score. Therefore, in the hypothetical example presented above, the result would be 0.095 (learner's score of *2* divided by the native typicality score of *21 = 0.095*). This number is the learner's association proportion for *brood,* and it reveals what proportion of all the native speaker association responses this learner produced for this particular stimulus word. Finally, each participant's association proportion, for each stimulus word, was categorised into one of the four native levels devised by Schmitt. In the analysis of the results, the range of native-like lexical ability among the learners was based on each participant's mean native-like level.

To answer the second question of the study concerning the strength of the relationship between the participants' mean native-like level scores and language proficiency and vocabulary size, multiple regression analysis was used. It was expected that the results of this analysis would provide a clearer interpretation of the participants' overall word knowledge capabilities in relation to their overall language capabilities. This would indicate whether or not the word association

measure could be used in conjunction with other diagnostic measures of language and lexical competence.

Finally, to answer the question about the effectiveness of the word association measure as a future instrument for class placement, multiple regression analysis was used to assess the strength of the relationship between the participants' ESL class level, their CELDT scores, and their mean native-like level scores. It was anticipated that the results would support the hypothesis that this word association test could be used in conjunction with other instruments for ESL class placement. If so, the level of word knowledge a learner has would be, in addition to his CELDT scores, an indicator of whether or not the learner would be successful in a particular ESL level.

CHAPTER 4. RESULTS AND DISCUSSION

4.1. CAPTURING DEGREES OF NATIVE-LIKE ABILITY

The first question asked in this study was whether Schmitt's (1998) proposed method of quantifying word association responses captures a range of native-like and lexical ability among non-native speakers of English. In order to determine the range of native-like scores, the participants' word association responses were calculated according to Schmitt's quantification method described in section 3.3.1. of this thesis and were classified according to Schmitt's four-level description of native-likeness (Appendix B). Table 3 below presents a summary of the mean native-like levels for all four ESL levels tested. A full representation of these results for each ESL class's association proportion and native-like level scores are found in Appendices F to P.

Table 3.

Mean Native-like Scores for Each ESL Class Level

ESL Level	n	Mean	SD
298B	16	0.713	0.248
207	40	0.805	0.386
208	24	0.892	0.363
209	73	1.049	0.388

Note: Native-like Level Scores range from 0 to 3

Each individual student within an ESL level produced a range of native-like responses. Moreover, the data illustrate that there is a range of native-like levels for each stimulus word, and that each ESL class, as a whole, demonstrates a different range of mean native-like level scores. These scores increase with each higher ESL level. This indicates that lexical knowledge changes qualitatively with language proficiency, and it can be assumed from these results that the learner's semantic network becomes more intricate at higher language levels.

In addition, this word association test makes it possible to classify a learner's lexical knowledge according to various levels of native-like associative performance. In other words, this test has captured a wide range of abilities in the association responses of non-native speakers of English, including varying levels of partial knowledge. Learners' scores ranged from 0 to 3. None of the learners produced native-like level scores of 0 or 3. Therefore, the native-like versus non-native-like distinction, which has been used to classify a learner's associative responses, is no longer the only way to look at a non-native speaker's lexical ability.

However, there are some instances of stimuli where scores do not increase with higher levels of language proficiency. In other words, the learners in a higher level ESL class are producing the same native-like level scores as learners in a lower level ESL class. This is seen with the stimulus word *brood*. This stimulus word proved to be the hardest lexical item for the learners to interpret on the word association test. Several participants, from all ESL levels, received a native-like level score of zero for this word because their responses

were either illogical or clang associations. These types of responses were not on the norm list because native speakers did not produce them. In addition, several participants left the response area for this word blank. It is hard to determine why the learners produced the responses they did. Possible reasons are that they lacked an understanding of the stimulus word, or misinterpreted it, or only recognised its sound.

In summary then, the answer to the first question is that a range of native-like lexical ability can be captured by this word association test that focuses on low frequency stimulus words. Therefore, measuring partial native-like lexical knowledge is possible. The non-native speakers demonstrate varying degrees of normative associative performance for each stimulus word, individually, within their ESL level, and between ESL levels. There are several possible explanations for the finding that improvement in native-like scores does not co-occur with increased language proficiency, such as the absence of a participant's association response from the norm list. Explanations for this variance will be addressed in more detail in the implications section of this study. Nonetheless, the success of capturing varying native-like associative ability for non-native speakers offers promise for the use of this word association test as a diagnostic measure.

4.2. USEFULNESS AS A DIAGNOSTIC MEASURE

The second question of this study is related to the reliability of this word association test as a diagnostic measure. Based on the findings that this test allows non-native speakers to demonstrate varying degrees of native speaker associative behaviour, the second question focuses on whether this word association test is a sufficient diagnostic measure. It does so by determining whether it relates to other diagnostic measures of general language proficiency and vocabulary breadth. The concern in this section is whether this measure can be used in conjunction with other proficiency and lexical measures.

As stated earlier, the participants were placed into an ESL level according to their CELDT scores. A learner's overall CELDT score is based on the results of two separate tests. One is a multiple-choice test, which measures the learner's language proficiency, and the other is a written test, which measures the learner's level of writing. Both of these CELDT scores were used as a measure of general language proficiency. The participant's score on the 5,000-word level of the Vocabulary Levels Test was used as a measure of vocabulary breadth. This is because a greater range of scores was found for the 5,000-word level than for any other word level of the test. Thus it appears to be the most useful assessment of vocabulary size. Therefore, in this multiple regression the dependent variable was the learners' mean native-like level scores, and the independent variables were the multiple choice CELDT scores, the written scores from the CELDT, and the five thousand word level scores. The results from the analysis are presented in Table 4.

Table 4.

Summary of Hierarchical Regression Analysis for Variables Predicting Mean Native-like Level in Non Native Speakers of English (N= 145)

Variable	*F*	*p*	Adjusted R^2	R^2
Step 1.				
Five Thousand Level	74.13	0.000	0.341	0.341
Step 2.				
CELDT MC	12.98	0.000	0.056	0.397
Five Thousand Level	26.32	0.000		

Note: Multiple R = 0.630
8 cases deleted due to missing data. $p < 0.05$

This forward step regression analysis indicates that the non-native speakers' vocabulary breadth scores and scores on the multiple-choice portion of the CELDT test predict their mean native-like levels. The first step in the multiple regression indicated that the relationship between mean native-like level scores and the five thousand word level scores was strong and positive, and this effect was significant at $p < 0.000$. A scatter plot diagram illustrating this relation is shown in Figure 4.1 below.

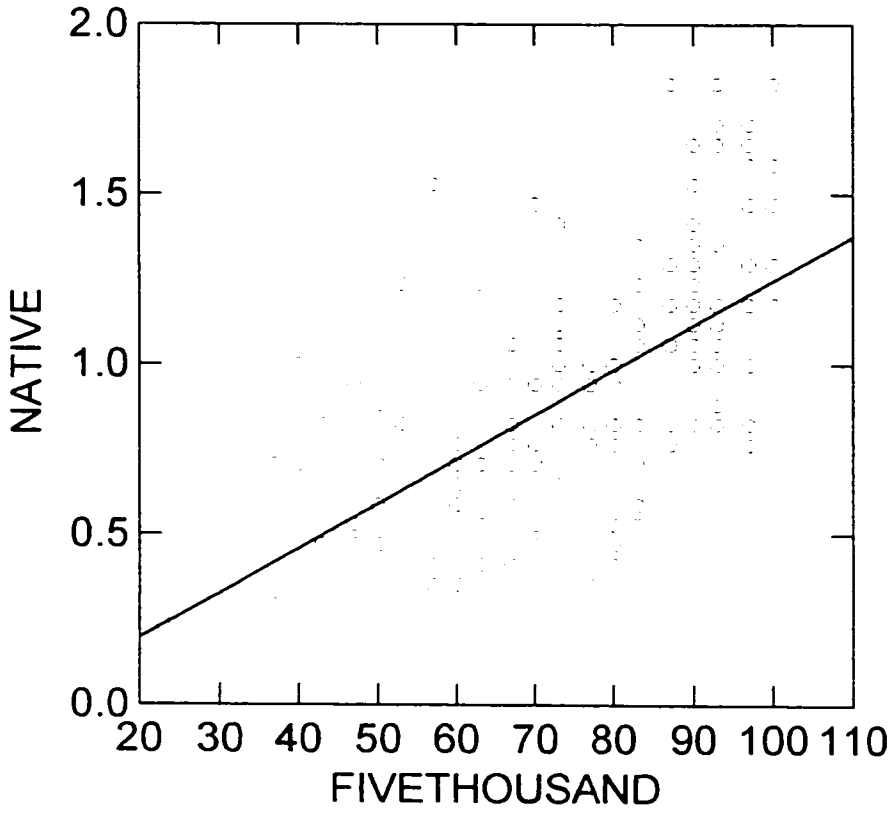


Figure 4.1. Correlation Between Five Thousand Word Level Scores and Mean Native-like Levels

The correlation illustrated in Figure 4.1. above indicates that learners' scores on 5,000 word level predict learners' mean native-like lexical levels. Therefore, the ESL 209 learners, with high scores on the 5,000 word level, are producing higher native-like level scores than the ESL 207 learners who have lower 5,000 word level scores. Indeed, if we look at Table 5, which lists the participants' mean scores on the five thousand word level of the Vocabulary Levels Test, we see that learners at higher levels of language proficiency are producing higher scores than lower level language learners.

Table 5.
Mean Five Thousand Word Level Scores

Class	Mean	SD
298B	65.4%	16.86
207	70.8%	17.64
208	77.8%	19.48
209	78.6%	17.76

The results in this part of the analysis appear to confirm that the quality of a non-native speaker's word knowledge is related to the number of words he claims to know. In other words, learners possessing a large number of words in their mental lexicon appear to have a fairly high level of native-like knowledge of the meanings of the lexical items tested in this study. Furthermore, the number of words a learner knows is larger at higher levels of language proficiency than at lower levels. Therefore, it can be assumed from the results that a non-native

speaker with a high level of language proficiency is more likely to have a high quality of lexical competence and organisation of these lexical items.

The second step of the regression analysis reveals that the learners' multiple-choice CELDT scores also predicted the associative native-like level scores. This relationship, illustrated by a scatter plot diagram shown below in Figure 4.2., is strong and positive and significant at $p < 0.000$.

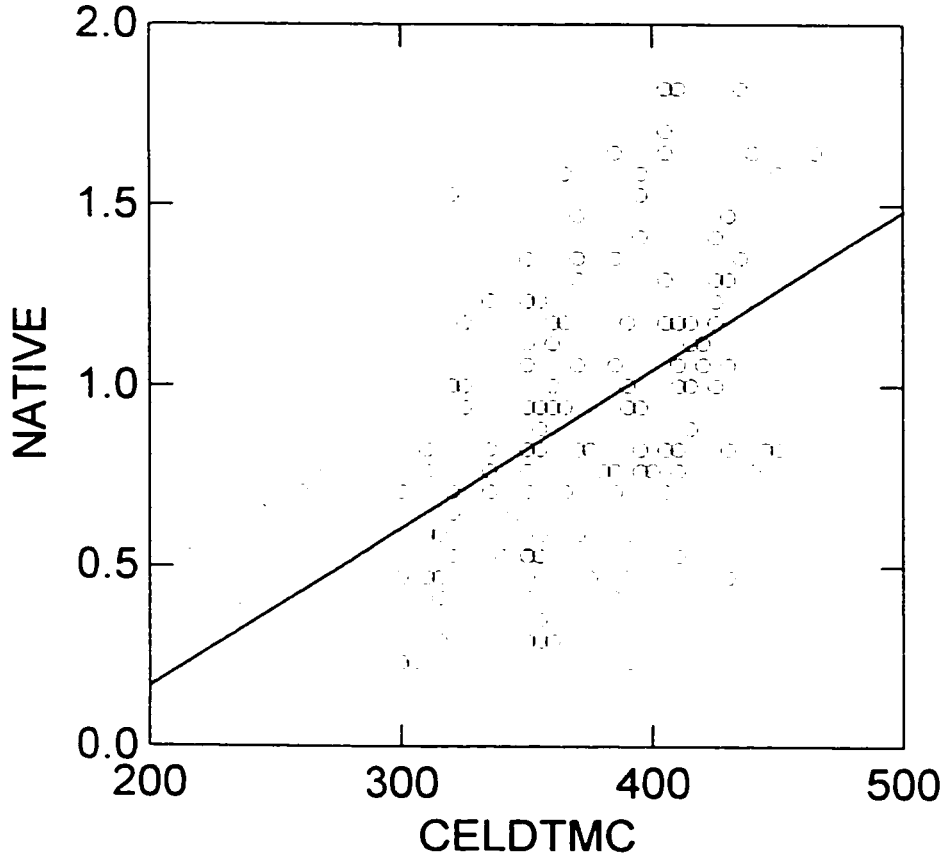


Figure 4.2. Correlation Between Multiple Choice CELDT Scores and Mean Native-like Levels

This correlation illustrates that learners' scores on the multiple-choice section of the CELDT predict learners' mean native-like levels. Therefore, the higher a learner's multiple-choice CELDT score is, the greater the likelihood his mean native-like level will be high. However, the results of this correlation should be viewed with some caution because the nature of these multiple-choice scores is rather questionable.

The non native-speakers' CELDT test scores are based on their original CELDT test performance. This test was not given again for the purposes of this study. Therefore, the CELDT scores are out of date for some participants by a maximum of two years. In that time, a learner could have progressed from his initial ESL level placement, for example 207, to his current ESL level at which he was tested, such as 209. On the other hand, some learners with a two-year old CELDT score were only attending their first ESL class at the time the word association test was administered. The reason for this is that the ESL learners at Concordia have a two year grace period between the time they receive their CELDT score and the deadline to enrol in an ESL class. In this situation, it is hard to account for the lexical gains these students made within the two-years prior to commencing the ESL program. In addition, there are some CELDT scores which are only a year old; participants in this situation could have still progressed from an ESL 207 level to an ESL 209 level within the one year. Finally, some of the scores are from the current term for some of the ESL learners.

Nonetheless, the results from the regression analysis reveal that these scores, no matter when they were obtained, predict learners' mean native-like level scores, and that the strength of the effect is significant. This confirms what vocabulary researchers have been saying for years: lexical ability increases and improves with increased language proficiency. Usually this conclusion is based on the number of words a learner knows, but now we can conclude that the quality of word knowledge may be connected to language proficiency. Therefore, the different levels of the quality of word knowledge are related to the different levels of language proficiency. Interestingly, the learners' CELDT written scores do not correlate with their mean native-like level scores ($p < 0.621$). This suggests that writing scores reflect other kinds of abilities than those tapped by the word association test.

In summary, the results appear to support the hypothesis of a relation between the word association test and other diagnostic measures of vocabulary breadth and language proficiency. Overall, the five thousand level of the Vocabulary Levels Test proves to be the strongest predictor of mean native-like lexical ability, followed by the CELDT multiple choice score. It is surprising that these CELDT scores, which could be as much as two years old, correlate with learners' current native-like lexical levels.

One final note is that, although these two variables predict a learner's native-like lexical level, their shared variance for mean native-like levels is 40%. This means that 40% of the variance in a learner's mean native-like level can be attributed to the learners' multiple-choice CELDT score and their vocabulary size

score. The remainder of the variance must be dependent on other factors not explored in this study. However, the results from this analysis strongly suggest that this word association measure can be used in conjunction with other diagnostic measures as a means of achieving a more global representation of a learner's language and lexical ability. The final concern of this study is whether this word association test can used as a measure of language placement.

4.3. USEFULNESS AS A PLACEMENT MEASURE

The final question in this study sets out to determine the effectiveness of this word association test as a measure of language placement. The dependent variable was the non-native speakers' ESL placement level, and the independent variables were the participants' multiple-choice and written CELDT scores, mean native-like level scores, and the five thousand word level scores. The aim was to determine whether or not this word association test could be an effective tool for predicting a language learner's placement in an appropriate ESL level. The results from the multiple regression analysis are presented in Table 6.

Table 6.

Summary of Hierarchical Regression Analysis for Variables Predicting A Non-Native Speaker's Placement Level in an ESL Class. (N=145)

Variable	*F*	*p*	Adjusted R^2	R^2
Step 1.				
CELDT Writ.	19.59	0.000	0.121	0.121
Step 2.				
Mean Native-like Level	11.00	0.001	0.063	0.184
CELDT Writ.	13.26	0.000		

Note: Multiple R = 0.429
8 cases deleted due to missing data. $p < 0.05$

The forward step regression analysis indicates that learners' scores on the writing portion of the CELDT and mean native-like level scores predicted placement in an ESL level. The other independent variables did not predict placement in an ESL level significantly, and as a result, these variables were not included in the regression analysis. The results from this first step in the multiple regression reveal that a language learner's placement in an ESL level at Concordia University is more closely related to his performance on the written portion of the CELDT test than to the multiple-choice portion of this test, which did not correlate significantly with placement. A scatter plot illustrating the relationship between learners' CELDT written scores and their placement levels in ESL, which is significant at $p<0.000$, is shown in Figure 4.3. below.

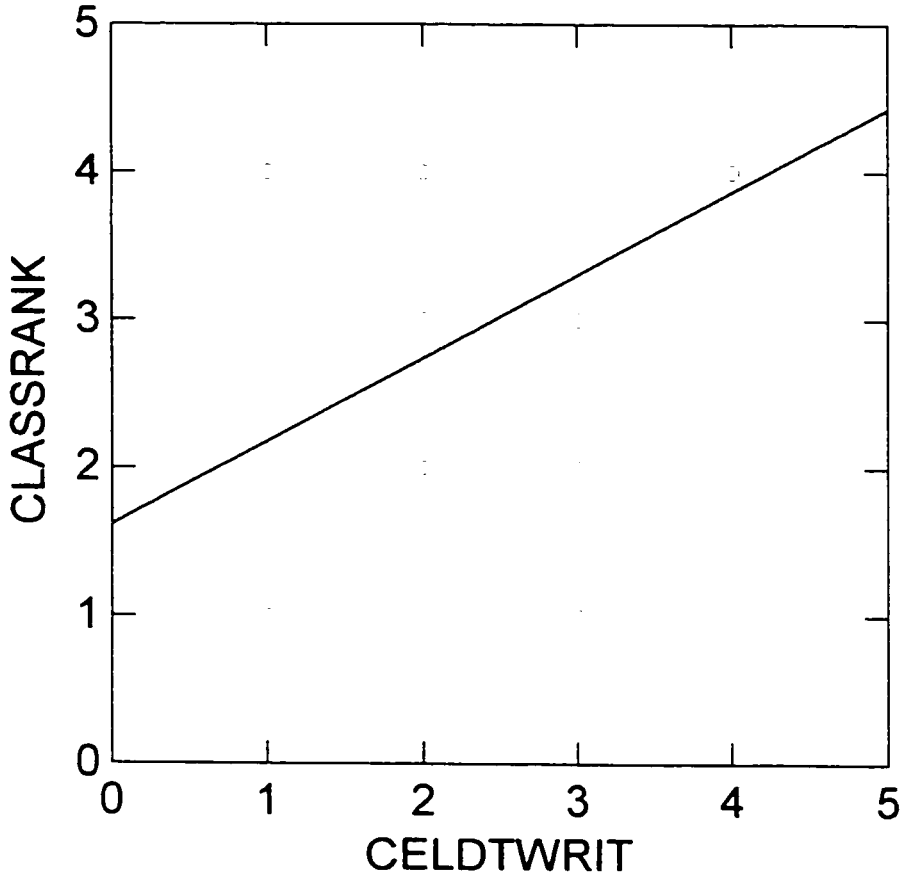


Figure 4.3. Correlation Between Written Test CELDT Scores and ESL Level (Class Rank)

The correlation illustrated in Figure 4.3. shows that learners' scores on the written portion of the CELDT predict their ESL level. Therefore ESL 209 learners are achieving higher writing scores on the CELDT than lower placed learners. Since the ESL program at Concordia University is an academic writing program, which focuses on developing and enhancing a non-native speaker's writing skills for academic purposes, it is not difficult to understand why the CELDT written scores are predicting ESL placement and not the CELDT multiple-choice scores. Although the administrators of the ESL program stress that a learner's class placement is based on a combination of these two scores, the results from this analysis suggest that a greater emphasis is placed on the writing section of the test. This result is surprising because, like the CELDT multiple-choice scores, some CELDT written scores could also be up to two years old. Therefore, these results should also be viewed with some caution.

The second step in this multiple regression reveals that a learner's mean native-like level predicts his ESL class placement. The effect of the relationship between the mean native-like level scores and placement level in an ESL class is significant at $p<0.001$. Figure 4.4. below, shows a scatter plot diagram illustrating this relationship.

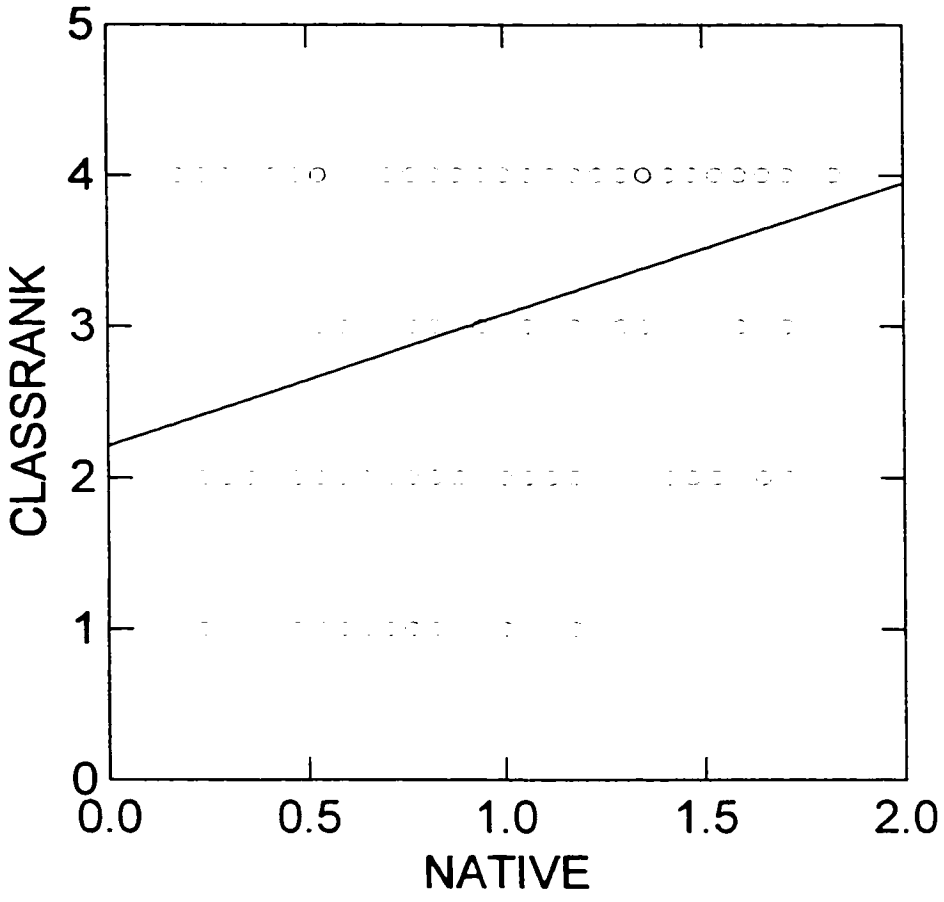


Figure 4.4. Correlation between Mean Native-like Level Scores and ESL Level (Class Rank)

The results show that mean native-like level scores increase with higher levels of ESL ability. Note that ESL 209 learners have higher mean native-like level scores than learners in a lower ESL level. Therefore, learners' placement in an ESL level could also be partially determined by the level of native-like lexical ability they possess. This suggests that the word association test could be used in conjunction with the CELDT as an additional measure for ESL placement, as it could provide a better understanding of learners' word knowledge. For the purposes of the ESL program at Concordia University, this word association test may be helpful in determining the level of vocabulary knowledge a learner needs to possess in order to be competitive and successful within a specific ESL level.

There is one final note. The amount of shared variance for class level predicted by these two variables is 18%. This means that 18% of variance in class placement is accounted for by the learners' mean native-like level scores and their CELDT written scores. The remainder of the variance is due to other factors not investigated in this study; however, as stated earlier, a learner's ESL class placement is also based on his CELDT multiple-choice score. This is likely one of the other factors contributing to the variance in class rank. Nonetheless, this multiple-choice portion of the CELDT failed to correlate with class placement in this study. Therefore, it is assumed that the effect of these scores on class placement is relatively low. In summary, the results from this analysis indicate that the word association test has an important role to play in class placement.

CHAPTER 5: IMPLICATIONS

5.1. ASPECTS OF THE DATA

Although most of the analysis in this study is quantitative, many qualitative inferences can be made about participants' responses. In addition, this study has provided information that could be applicable to future research in vocabulary studies dealing with word association responses. However, there are some limitations of this study that should be addressed before any future investigations can take place.

5.1.1. Qualitative Findings

Many of the findings in this study can only be partially explained by the quantitative analyses. To shed light on the nature of many of the associative responses produced by the participants in this study, it was necessary to look at the data qualitatively. However, it is generally difficult to make conclusions about why a learner produced the responses he did. A number of intrinsic factors, such as the emotional state of the learner, and extrinsic factors, such as a priming effect of the type of one response on another, can affect association responses. Nonetheless, an attempt to explain some of the qualitative aspects of the participants' responses is addressed in this section.

First, the participants who produced the highest word association proportions and, therefore, the highest native-like level scores, were the Chinese

speakers. It is hard to explain these results because the vocabulary strategies these second language learners employed were not investigated. However, many of the ESL teachers interviewed about these students commented that the Chinese speakers have an amazing amount of vocabulary knowledge, and, overall, their writing assignments appear to be more lexically rich than other non-native speakers in the class. These teachers believe that the methodology used in the schools in which the Chinese speakers were taught English as a foreign language in their home country might account for their high lexical ability. Typically, the Chinese speakers employ a rote memorisation technique for acquiring new lexical items. Therefore, we might expect these non-native speakers to have a larger number of low frequency lexical items in their lexical store. If this is the case, then they would have a greater familiarity with different types of lexical items, and a stronger understanding of lexical items different sense relations. This may explain why they would produce higher native-like association scores.

Second, past research with word association tests described low level language learners' association responses as being more diverse and heterogeneous compared to those of native speakers. In addition, researchers found that many low level learners' association responses were clang or syntagmatic associations, while the association responses of high level language learners tended to be more paradigmatic. The participants of this study appeared to be producing association responses in a similar fashion. In other words, the low level learners appeared to be producing more clang and syntagmatic

associations than the high level learners. A typical clang response was *broom* in response to *brood*. A typical syntagmatic response was *rock* in response to *massive*. High level learners appear to be producing more paradigmatic responses. Typical paradigmatic responses were *change* and *alter* in response to *convert*.

However, the high level ESL participants did not produce paradigmatic responses exclusively. Several of these learners produced syntagmatic responses, the odd clang association, as well as a small number of paradigmatic responses. In fact, the number of paradigmatic responses for the high level ESL learners was rather low. This suggests that the lexical familiarity these high level ESL participants have with some of the stimulus words, while significant, is not much stronger than that of the ESL participants. Nonetheless, the variety of responses produced by the high level ESL participants reveal that they are demonstrating a more sophisticated organisation of lexical items in their mental lexicon than low level learners, whose semantic networking is in a state of flux. Overall, the participants' responses support the notion that the associative responses of non-native speakers shift from clang to syntagmatic to paradigmatic associations with increased language proficiency.

An additional observation about the type of association responses produced has to do with how the learners interpreted some of the stimulus words. Words such as *launch* and *surging* were regularly misinterpreted. Some participants interpreted *launch* as *lunch*, and *surging* as having to something to do with *surgeon*. This was determined by looking at some of the responses that

were produced for these words, for example *breakfast, sandwich, dinner,* and *brunch* in response to *launch*, and *hospital, doctor, nurse,* and *surgery* in response to *surging*. These types of responses indicate that the learners relied on their interpretation of the sound of the stimulus word as a means to produce association responses and signify that the learners might not be familiar, orthographically or phonologically, with these stimulus words.

In addition, there are many other qualitative results from this study that remain to be investigated: for example, another look at the data might reveal very interesting findings about the grammatical classes of the learners' associative responses. In other words, it might be determined that the grammatical class of the learners' associative responses are primed by the grammatical class of the corresponding stimulus word. Therefore, the syntactic properties of a stimulus word might be a stronger determinant of the types of associative responses produced than the semantic properties of the stimulus words. Future investigation into the data will reveal the role of grammatical word knowledge. Although, these issues were outside the scope of this study, the data already collected can be used to address these concerns in future research.

5.1.2 Limitations

While this study has resulted in potentially useful quantitative and qualitative findings, the limitations to the word association methodology employed here might help explain some of the peculiar results found. The first limitation of this study pertains to the CELDT test scores. The reader will recall

that the CELDT was not administered a second time for the purposes of this study due to the time limitations imposed by the ESL teachers. Instead, the learners' original CELDT scores were used as a measure of general language proficiency. However, these scores are not an indication of current English language competence, and as a result, it would be assumed that the CELDT scores would not correlate with the current evaluation of the learners' vocabulary depth.

Surprisingly however, the CELDT scores do correlate with current lexical ability and class placement. Specifically, the multiple-choice CELDT scores correlate with the learners' mean native-like level scores, and the CELDT written test scores correlate with the learners' level placement. How this is possible when many of these test scores are as much as two years old is unexplainable. These findings suggest that the language learners, who have advanced from ESL 207 to ESL 208 or ESL 209, have not made any linguistic gains in their language proficiency. This seems highly unlikely.

An additional unexplained finding related to the CELDT scores is that both parts of the test correlate with different questions asked in this study. If the class placement were based on both CELDT scores, then both scores would be expected to correlate with ESL placement. However, the results indicate that this is not the case. Furthermore, if both scores are an indication of general language proficiency, then both scores should have predicted a learner's mean native-like level. As we have seen from the results of the analysis used in this study, this is not the case. This might be due to the manner in which both sections of the

CELDT are weighted for class placement and general language proficiency. Perhaps a more current general language proficiency measure should have been used.

The second limitation deals with the norm list created by Schmitt (1998). While the norm list allows non-native speakers to demonstrate partial degrees of normative native speaker associative performance, it still permits a native-like versus non-native-like distinction for some non-native speakers' associative responses to specific words. One of these is *plot*. According to COBUILD's English Dictionary for Advanced Learners, the stimulus word *plot* has nine different meaningful entries. Therefore, one might expect associations that reflect all these possible entries. However, this is not the case, and the frequency with which associations for some of these entries are produced is lower than for others. Many of the non-native speakers in this study produced logical association responses which indicated understanding of mathematical or scientific meanings. For instance, responses to the stimulus word ***plot*** included *draw, diagram,* and *math*. Even though these answers are very logical "native-like" associative responses, they are not on the norm list. Therefore, the participant would receive a frequency score of zero for association responses like *draw* or *math*, resulting in a low association proportion and overall native-like level score. This is not to say that native speakers did not capture this particular definition of the stimulus word at all. On the contrary, there are some norm responses that relate to this meaning, such as *graph, data,* and *calculate,* but

they were not represented strongly enough in the norm list to affect non-native speaker outcomes.

This problem might be due to the types of native-speakers used in Schmitt's normative group. The norming group was made up of university business students, modern English language students, and French majors, while many of the participants in the current thesis study were students of computer science, general science, maths and engineering. The types of domain-specific words that these two groups have in their lexicon are very different, and since domain specific words may be strongly fixed in the mental lexicon, they could easily be the *first response that comes to mind*, which is what the participants were asked to provide.

A non-native speaker could exhibit native-like knowledge of a stimulus word in his association responses, but if these particular responses are not on the norm list, he is considered to have a non-native-like understanding of the word. This is what appears to have happened to several of the participants in this study. In order to avoid this situation in the future, a larger norm group consisting of native speakers from a broader range of professional and educational domains should be used to develop a more representative norm list.

Finally, it should be kept in mind that the results in this study are based on association responses to a small number of low frequency stimulus words. It could be argued that all the conclusions made thus far are only applicable to the stimulus words used in this test. Even though these stimulus words are believed to be equivalent to other low frequency polysemous words in English, the manner

in which these words were selected needs to be defined. Schmitt did not make this clear in his 1998 study. Nonetheless, the results obtained in this thesis study remain significant, and only future investigations into the nature of the stimulus words used in this word association test will clarify this situation.

5.1.3. Recommendations for Future Research

Researchers using word association tests in future investigations should keep in mind considerations Schmitt raised in his alternative association methodology. Heeding the opinions of many vocabulary researchers, Schmitt used low frequency stimulus words because these items provide more information about what is actually happening to lexical items, in the peripheral layer of word knowledge, as they are being integrated into the mental lexicon. In addition, he utilised a multiple word response method, which allows non-native speakers to exhibit the extent of their understanding on one or more of the meanings of the stimulus words. Furthermore, Schmitt's quantification scheme allows for non-native speakers to demonstrate degrees of normative associative performance. However, if this word association model is used for future research, some other factors should be kept in mind.

First, the norm list should be based on a larger population, with a sample of native speakers from a wider range of occupational and scholarly domains. Only then can a norm list incorporate associations that capture the polysemous nature of these stimulus words. Second, measures of general language proficiency should be current with measures of vocabulary breadth and depth.

A third suggestion for future research deals with a different approach to the norm list. Typically, word association norm lists are based on native speaker associative responses (Schmitt, 1998; Meara, 1983), but the goal of a language learner is not to gain native speaker status in the second language community but rather to be a competent bilingual speaker (Meara, 1983). Therefore, future researchers should create a norm list based on the associative responses of co-ordinate bilingual speakers, bilingual speakers who have learned two languages at the same time from a young age, since the goal of a non-native speaker is to be a bilingual speaker. It is predicted that this would provide a clear illustration of what is going on in the mental lexicon because a bilingual association norm list would allow for a clearer interpretation of the semantic network and lexical organisation of two working languages.

If the aim of the word association test is to investigate and gauge the quality of word knowledge, perhaps the learner's knowledge should be tested in a less superficial manner than usually occurs in a word association test. In other words, instead of asking non-native speakers to respond to a stimulus word with the *first response that comes to mind*, non-native speakers should be asked *what other words could we use instead of this stimulus word?* (Nation, 2001). The association responses of the non-native speaker would then be activated by other words in the mental lexicon. This might allow researchers to evaluate the organisation of a non-native speaker's mental lexicon in more detail, because association responses would reveal whether or not the non-native speaker has an understanding of stereotypical native-like association responses. As a result,

this might allow researchers to better determine the extent to which the semantic network of the non-native speakers is similar to that of native speakers.

CHAPTER 6. CONCLUSION

When Schmitt first devised his method of quantification, he did so because the potential of the word association test had been limited for a long time by an unsophisticated methodology. In the area of second language vocabulary acquisition research and assessment, the word association test holds great promise, and the results of this study have borne this out.

The findings of this study highlight how it is possible to attain a deeper interpretation of how well non-native speakers know various lexical items in their second language. For years, the results of the word association test have categorised non-native speakers' lexical ability as native-like or non-native-like, but this dichotomy is limiting in its conclusions because it does not capture partial levels of native-like lexical knowledge. This is due to three main problems with the original methodology of the test. First, the original format did not account for the differences in the typicality of association responses. Second, associative performance was based on a single unit of information in response to high frequency stimulus words. Third, there was no principled way of determining whether any associative response was native-like or not.

Schmitt (1998) addressed all of these problems when he proposed an alternative method of quantifying word association responses that might allow non-native speakers to demonstrate varying degrees of native speaker associative performance. He achieved this by introducing a different approach to collecting word association responses that would result in a better understanding

of the organisation of the second language mental lexicon. First, he proposed the use of low frequency stimulus words and a multiple word response method, which would allow researchers to obtain a better indication of the networking of the non-native speaker's lexical store. Second, he introduced a new method of weighting associative responses on the basis of their typicality to provide a better understanding of a non-native speaker's level of knowledge of low frequency stimulus words. Third, he devised a four-level description of native-like ability which would allow researchers to categorise non-native speakers' associative ability in a way that accounts for partial native-like lexical ability. Through these proposals, Schmitt laid the groundwork for future investigation into the nature of word association responses, and in this thesis I attempted to build on his work.

The results of this study suggest that Schmitt's quantification scheme makes it possible to replace the dichotomous native-like versus non-native-like classification of non-native speakers' association responses with a categorisation scheme that reflects the varying degrees of native-like lexical ability a non-native speaker possesses. The lexical gains a learner achieves with greater language proficiency and lexical ability is captured by Schmitt's word association test. In addition, the results of this study show that this word association measure correlates with other diagnostic language measures, which reflect a learner's language competence and vocabulary size. Moreover, the results of this study suggest that the word association test plays a useful role in conjunction with other ESL placement measures. Therefore, this measure could possibly be used for both diagnostic and placement purposes in an ESL program.

The results obtained in this study provide promising insights into the usefulness of word association tests in assessing language learners, but this study also contributes to vocabulary acquisition research as a whole. At the end of his 1998 paper Schmitt hoped that his proposed methodology would "prove useful in future enquiries into both vocabulary learning and measurement" (p. 400). The results of this study illustrate the potential strength of word association tests in language acquisition research. The quantification method utilised in this study suggests that it is possible to assess the subtleties of a learner's associative lexical inventory. This paves the way for innovative uses of the word association test in future vocabulary research. The semantic network of a non-native speaker's mental lexicon holds a wealth of information about the nuances of second language lexical organisation, and the development of semantic connections between words with increased language proficiency. Now there is an improved way of investigating this development.

"The mechanics of vocabulary learning are still something of a mystery" (Schmitt, 2000, p. 4) The findings in this study provide another piece of the puzzle, and the next positive step into future investigations in vocabulary acquisition research using word association tests.

References

Aitchison, J. (1994). *Words in the mind: An introduction into the mental lexicon* (2nd ed.). Oxford and New York: Basil Blackwell.

Sinclair, J. (Ed.). (2001) *Collins Cobuild English dictionary for advanced learners* (3rd ed.). Glasgow: Harper Collins Publishers.

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

Laufer, B. (1986). Possible changes towards vocabulary acquisition research. *International Review of Applied Linguistics, 24,* 69-75.

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics 16*(3), 307-322.

Levenston, E.A. (1979). Second language acquisition: Issues and problems. *The Interlanguage Studies Bulletin, 4,* 147-160.

Meara, P. (1981). Vocabulary acquisition: a neglected aspect of language learning. *Language Teaching and Linguistics Abstracts, 13,* 221-246.

Meara, P. (1983). Word associations in a foreign language: a report on the Birkbeck Project. *Nottingham Linguistic Circular, 11,* 29-38.

Meara, P. (1993). The bilingual lexicon and the teaching of vocabulary. In R. Schreuder and B. Weltens (Eds.), *The bilingual lexicon* (p. 279-297). Amsterdam: Benjamins.

Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjaer and J. Williams (Eds.), *Performance and competence in second language acquisition* (p. 36-53). Cambridge: Cambridge University Press.

Meara, P., & Fitzpatrick, T. (2000). Lex30: an improved method of assessing productive vocabulary in an L2. *System, 28,* 19-30.

Nation, I.S.P. (1990). *Teaching and learning vocabulary.* New York: Heinle & Heinle.

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

Palmberg, R. (1990). Improving foreign language learners' vocabulary skills. *RELC Journal, 21*(1), 1-10.

Qian, D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *The Canadian Modern Language Review, 56,* 282-307.

Read, J. (1988). Measuring the vocabulary knowledge of second language learners. *RELC Journal, 19*(2), 12-25.

Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing, 10*(3), 355-371.

Schmitt, N., and Meara, P. (1997). Researching vocabulary through a word knowledge framework: word associations and verbal suffixes. *Studies in Second Language Acquisition 19*(1), 17-35.

Schmitt, N., & McCarthy, M. (Eds.). (1997). *Vocabulary: Description, acquisition, and pedagogy.* Cambridge: Cambridge University Press.

Schmitt, N. (1998). Tracking the incremental acquisition of second language vocabulary: a longitudinal study. *Language Learning, 48* (2), 281-317.

Schmitt, 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. (2000). *Vocabulary in language teaching.* Cambridge: Cambridge University Press.

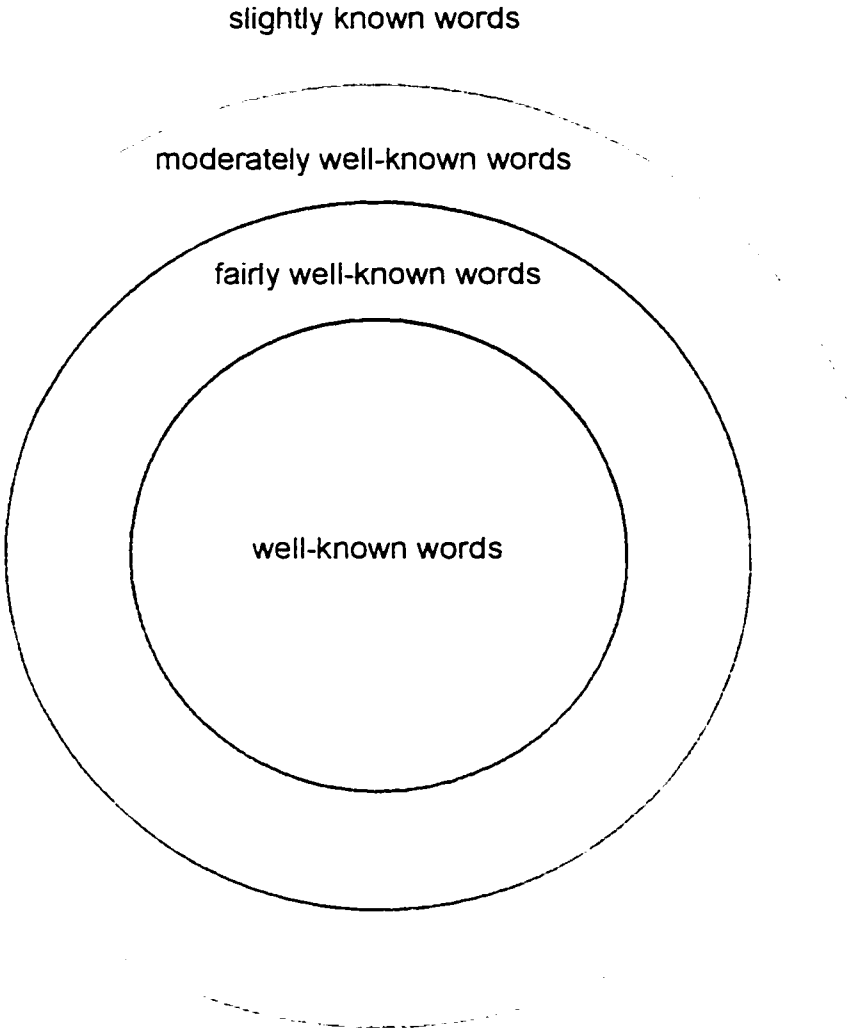
Söderman, T. (1993). Word associations of foreign language learners and native speakers: the phenomenon of a shift in responses types and its relevance to lexical development. In H. Ringbom (Ed.), *Near-native proficiency in English.* Åbo: Åbo Akademi, English Department Publications.

Sökmen, A. (1993). Word association results: a window in the lexicons of ESL students. *JALT Journal 15* (2), 135-150.

Wesche , M., & Paribakht, T.M. (1996). Assessing vocabulary knowledge: Depth vs. breadth. *Canadian Modern Language Review, 53,* 13-40.

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

Appendix A: Depth of Word Knowledge of The Mental Lexicon (Wolter, 2000)



Appendix B: Native Languages of The ESL Participants

Code	Language	No. of Speakers (N=153)
01	Arabic	13
02	Armenian	1
03	Bengali	9
04	Bulgarian	1
05	Chinese	84
06	Dari	1
07	Farsi	4
08	French	8
09	Greek	1
10	Hungarian	1
11	Japanese	3
12	Korean	3
13	Persian	3
14	Romanian	2
15	Russian	2
16	Spanish	9
17	Taiwanese	2
18	Tamil	1
19	Urdu	1
20	Vietnamese	4

Appendix C: Word Association Test

Instructions: **In the spaces provided, write the first three responses that come to mind for each stimulus word.**
You have 25 minutes to complete this test.
Please DO NOT write your name on the test paper.

Personal Information:

1. Were you born in Canada? _____
   If no, what is your nationality? _____
2. How many years have you been in Canada? _____
3. What language(s) do you speak at home? _____
4. How many years have you been studying English? _____
5. Age: _____

---

abandon	_____	_____	_____
brood	_____	_____	_____
circulate	_____	_____	_____
convert	_____	_____	_____
dedicate	_____	_____	_____
illuminate	_____	_____	_____
launch	_____	_____	_____
massive	_____	_____	_____
plot	_____	_____	_____
peak	_____	_____	_____
rare	_____	_____	_____
spur	_____	_____	_____
subtle	_____	_____	_____
surging	_____	_____	_____
suspend	_____	_____	_____
trace	_____	_____	_____
trend	_____	_____	_____

Appendix D: Four-Level Description of Native-likeness (Schmitt, 1998)

Level 0: • Association Proportion = 0 • No native-like responses produced
Level 1: • Association Proportion > 0< Threshold Level • One or more atypical responses on the norm list are produced
Level 2: • Association Proportion =≥ Threshold Level and < Mean Association Proportion • Native-like productive association performance
Level 3: • Association Proportion ≥ Mean Association Proportion • Native-like associations in the top portion of the native norm group

Appendix E: Vocabulary Levels Test (Schmitt, 2000)

Version 1 The 2,000 word level

1 birth	
2 dust	_____ game
3 operation	_____ winning
4 row	_____ being born
5 sport	
6 victory	

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

1 choice	
2 crop	_____ heat
3 flesh	_____ meat
4 salary	_____ money paid regularly for doing a job
5 secret	
6 temperature	

1 bake	
2 connect	_____ join together
3 inquire	_____ walk without purpose
4 limit	_____ keep within a certain size
5 recognize	
6 wander	

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

1 burst	
2 concern	_____ break open
3 deliver	_____ make better
4 fold	_____ take something to someone
5 improve	
6 urge	

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

1 original	
2 private	_____ first
3 royal	_____ not public
4 slow	_____ all added together
5 sorry	
6 total	

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

1 brave	
2 electric	_____ commonly done
3 firm	_____ wanting food
4 hungry	_____ having no fear
5 local	
6 usual	

Version 1 The 3,000 word level

1 belt	
2 climate	_____ idea
3 executive	_____ inner surface of your hand
4 notion	_____ strip of leather worn around the waist
5 palm	
6 victim	

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

1 acid	
2 bishop	_____ cold feeling
3 chill	_____ farm animal
4 ox	_____ organization or framework
5 ridge	
6 structure	

1 encounter	
2 illustrate	_____ meet
3 inspire	_____ beg for help
4 plead	_____ close completely
5 seal	
6 shift	

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

1 assist	
2 bother	_____ help
3 condemn	_____ cut neatly
4 erect	_____ spin around quickly
5 trim	
6 whirl	

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

1 annual	
2 concealed	_____ wild
3 definite	_____ clear and certain
4 mental	_____ happening once a year
5 previous	
6 savage	

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

1 dim	
2 junior	_____ strange
3 magnificent	_____ wonderful
4 maternal	_____ not clearly lit
5 odd	
6 weary	

Version 1 The 5,000 word level

1 balloon
2 federation
3 novelty
4 pail
5 veteran
6 ward

_____ bucket
_____ unusual interesting thing
_____ rubber bag that is filled with air

1 blend
2 devise
3 hug
4 lease
5 plague
6 reject

_____ mix together
_____ plan or invent
_____ hold tightly in your arms

1 alcohol
2 apron
3 hip
4 lure
5 mess
6 phase

_____ stage of development
_____ state of untidiness or dirtiness
_____ cloth worn in front to protect your clothes

1 abolish
2 drip
3 insert
4 predict
5 soothe
6 thrive

_____ bring to an end by law
_____ guess about the future
_____ calm or comfort someone

1 apparatus
2 compliment
3 ledge
4 revenue
5 scrap
6 tile

_____ expression of admiration
_____ set of instruments or machinery
_____ money received by the Government

1 bleed
2 collapse
3 precede
4 reject
5 skip
6 tease

_____ come before
_____ fall down suddenly
_____ move with quick steps and jumps

1 bulb
2 document
3 legion
4 mare
5 pulse
6 tub

_____ female horse
_____ large group of soldiers or people
_____ a paper that provides information

1 casual
2 desolate
3 fragrant
4 radical
5 unique
6 wholesome

_____ sweet-smelling
_____ only one of its kind
_____ good for your health

1 concrete
2 era
3 fiber
4 loop
5 plank
6 summit

_____ circular shape
_____ top of a mountain
_____ a long period of time

1 gloomy
2 gross
3 infinite
4 limp
5 slim
6 vacant

_____ empty
_____ dark or sad
_____ without end

Version 1 Academic Vocabulary

1 benefit
2 labor
3 percent
4 principle
5 source
6 survey

_____ work
_____ part of 100
_____ general idea used to guide one's actions

1 element
2 fund
3 layer
4 philosophy
5 proportion
6 technique

_____ money for a special purpose
_____ skilled way of doing something
_____ study of the meaning of life

1 consent
2 enforcement
3 investigation
4 parameter
5 sum
6 trend

_____ total
_____ agreement or permission
_____ trying to find information about something

1 decade
2 fee
3 file
4 incidence
5 perspective
6 topic

_____ 10 years
_____ subject of a discussion
_____ money paid for services

1 colleague
2 erosion
3 format
4 inclination
5 panel
6 violation

_____ action against the law
_____ wearing away gradually
_____ shape or size of something

1 achieve
2 conceive
3 grant
4 link
5 modify
6 offset

_____ change
_____ connect together
_____ finish successfully

1 convert
2 design
3 exclude
4 facilitate
5 indicate
6 survive

_____ keep out
_____ stay alive
_____ change from one thing into another

1 anticipate
2 compile
3 convince
4 denote
5 manipulate
6 publish

_____ control something skillfully
_____ expect something will happen
_____ produce books and newspapers

1 equivalent
2 financial
3 forthcoming
4 primary
5 random
6 visual

_____ most important
_____ concerning sight
_____ concerning money

1 alternative
2 ambiguous
3 empirical
4 ethnic
5 mutual
6 ultimate

_____ last or most important
_____ something different that can be chosen
_____ concerning people from a certain nation

Appendix F: Association Proportions and Native-like Levels for ESL298-1-IN

Table F1: Association Proportion ESL298B-1-IN

Stimulus Words	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16
abandon (129)	0.659	0.023	0.698	0.876	0	0	0.674	0	0.659	0.682	0	0.124	0.147	0.682	0.659	0.008
brood (75)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
circulate (71)	0.169	0	0.155	0.296	0.423	0	0.31	0.014	0.338	0.155	0	0.014	0.085	0	0.31	0.141
convert (139)	0.619	0.022	0.849	0.619	0.676	0	0.633	0.007	0	0	1	0.194	0	0.647	0.806	0.007
dedicate (63)	0	0.032	0	0.048	0	0	0	0.032	0.397	0.032	0	0	0	0	0.032	0
illuminate (112)	0	0.964	0.768	0.848	0.741	0.063	0	0	0.009	0.054	0	0.732	0.723	0.696	0.196	0
launch (75)	0	0.493	0	0	0	0	0.187	0.28	0	0	0.507	0	0	0.12	0	0.013
massive (187)	0	0.005	0.743	0.38	0.406	0	0.011	0	0	0	0.332	0	0.487	0.332	0.332	0.674
plot (75)	0	0	0.053	0.027	0.013	0	0	0	0	0.053	0	0	0	0	0	0
peak (131)	0	0.427	0.008	0	0.397	0	0.84	0	0	0.397	0	0	0.725	0	0	0.008
rare (76)	0	0	0.053	0	0	0	0.026	0.513	0	0	0.013	0.053	0.04	0	0.04	0
spur (95)	0	0	0	0	0	0	0	0	0	0.021	0.011	0	0	0	0	0
subtle (56)	0	0	0.018	0.018	0	0	0.018	0	0	0	0.125	0	0	0	0	0
surging (62)	0	0	0.016	0	0	0.016	0	0.016	0	0	0	0	0	0	0	0
suspend (104)	0	0	0.01	0	0.154	0.135	0	0	0	0	0.75	0.135	0	0	0.115	0
trace (99)	0.343	0.081	0	0	0.01	0	0	0	0	0	0	0	0	0	0.03	0.02
trend (114)	0	0.026	0	0.018	0.053	0	0.719	0	0	0.053	0	0.009	0	0	0	0.018

Table F2: Native-like Levels ESL298B-1-IN

Stimulus Words	Threshold	Mean Prop	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16
abandon (129)	0.306	0.68	2	1	3	3	0	0	2	0	2	3	0	1	1	3	2	1
brood (75)	0.094	0.45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
circulate (71)	0.148	0.43	2	0	2	2	2	0	2	1	2	2	0	1	1	0	2	1
convert (139)	0.307	0.67	2	1	3	2	3	0	2	1	0	0	3	1	0	2	3	1
dedicate (63)	0.040	0.36	0	1	0	2	0	0	0	1	3	1	0	0	0	0	1	0
illuminate (112)	0.253	0.67	0	3	3	3	3	1	0	0	1	1	0	3	3	3	1	0
launch (75)	0.107	0.45	0	3	0	0	0	0	2	2	0	0	3	0	0	2	0	1
massive (187)	0.343	0.71	0	1	3	2	2	0	1	0	0	0	1	0	2	1	1	2
plot (75)	0.059	0.44	0	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0
peak (131)	0.266	0.58	0	2	1	0	2	0	3	0	0	2	0	0	3	0	0	1
rare (76)	0.096	0.47	0	0	1	0	0	0	1	3	0	0	1	1	1	0	1	0
spur (95)	0.033	0.41	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0
subtle (56)	0.021	0.36	0	0	1	1	0	0	1	0	0	0	2	0	0	0	0	0
surging (62)	0.094	0.41	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0
suspend (104)	0.129	0.59	0	0	1	0	2	2	0	0	0	0	3	2	0	0	1	0
trace (99)	0.121	0.47	2	1	0	0	1	0	0	0	0	0	0	0	0	0	1	1
trend (114)	0.250	0.67	0	1	0	1	1	0	3	0	0	1	0	1	0	0	0	1
Average Native-like Level:			0.471	0.824	1.176	1	1	0.235	1	0.529	0.471	0.706	0.824	0.588	0.647	0.647	0.765	0.529

Appendix G: Association Proportions and Native-like Levels for ESL207-1-DL

Table G1: Association Proportion ESL207-1-DL

Stimulus Words	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	D16	D17	D18	D19	D20	D21
abandon (129)	0	0.008	0.031	0.031	0	0	0.023	0	0	0	0.008	0.008	0.031	0	0.698	0.659	0	0	0.023	0	0
brood (75)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.013	0	0	0	0
circulate (71)	0	0.141	0.155	0.141	0	0	0.056	0	0	0	0.423	0.141	0.014	0	0	0	0.155	0	0.028	0.141	0
convert (139)	0.194	0	0.014	0.647	0	0	0.619	0.64	0	0.022	0.619	0.014	0.007	0.676	0	0.655	0	0	0.05	0	0.647
dedicate (63)	0	0	0.032	0	0	0.016	0.032	0	0	0	0.778	0	0	0	0.095	0.19	0	0	0.016	0	0
illuminate (112)	0.75	0	0.018	0.045	0.045	0.009	0.045	0.696	0.036	0	0	0.696	0.696	0	0.938	0	0.054	0.768	0.768	0	0.018
launch (75)	0	0.013	0.067	0.013	0	0.067	0	0	0.027	0	0	0.067	0.067	0	0	0	0.067	0	0.133	0.013	0.187
massive (187)	0.406	0.005	0	0	0	0	0.005	0	0.005	0	0.738	0.262	0.406	0.005	0.016	0.032	0.016	0	0.005	0.594	0
plot (75)	0	0.013	0	0	0.027	0	0	0.133	0	0	0	0	0.027	0	0	0.027	0	0	0	0.227	0
peak (131)	0.427	0.084	0.122	0.443	0.427	0.824	0	0.832	0	0.427	0.557	0.427	0.122	0	0.519	0	0.947	0	0.519	0	0
rare (76)	0.105	0	0.013	0	0	0	0	0.237	0.145	0	0	0.132	0	0	0.013	0.526	0.184	0	0.092	0.184	0.184
spur (95)	0	0	0	0.011	0	0	0	0	0.011	0.011	0.621	0.011	0	0.011	0	0	0.011	0	0.074	0	0
subtle (56)	0	0.036	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
surging (62)	0.435	0.065	0	0	0	0	0	0	0	0	0.016	0	0	0	0	0	0	0	0.645	0	0
suspend (104)	0	0.01	0.087	0.019	0.106	0	0	0	0	0	0.798	0.01	0	0	0.029	0	0.01	0	0	0	0
trace (99)	0	0	0	0.01	0	0	0	0.03	0	0	0.212	0	0.01	0.545	0.01	0.02	0.202	0	0	0	0.071
trend (114)	0	0	0	0	0	0.018	0	0	0	0	0.833	0	0.018	0	0	0	0	0.711	0	0.061	0

Table G2: Native-like Levels ESL207-1-DL

Stimulus Words	Threshold	Mean Prop.	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	D16	D17	D18	D19	D20	D21
abandon (129)	0.306	0.68	0	1	1	1	0	0	1	0	0	0	1	1	1	0	3	2	0	0	1	0	0
brood (75)	0.094	0.45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
circulate (71)	0.148	0.43	0	1	2	1	0	0	1	0	0	0	2	1	1	0	0	0	2	0	1	1	0
convert (139)	0.307	0.67	1	0	1	2	0	0	2	2	0	1	2	1	1	3	0	2	0	0	1	0	2
dedicate (63)	0.040	0.36	0	0	1	0	0	1	1	0	0	0	3	0	0	0	2	2	0	0	1	0	0
illuminate (112)	0.253	0.67	3	0	1	1	1	1	1	3	1	0	0	3	3	0	3	0	1	3	3	0	1
launch (75)	0.107	0.45	0	1	1	1	0	1	0	0	1	0	0	1	1	0	0	0	1	0	2	1	2
massive (187)	0.343	0.71	2	1	0	0	0	0	1	0	1	0	3	1	2	1	1	1	1	0	1	2	0
plot (75)	0.059	0.44	0	1	0	0	1	0	0	2	0	0	0	0	1	0	0	1	0	0	0	2	0
peak (131)	0.266	0.58	2	1	1	2	2	3	0	3	0	2	2	2	1	0	2	0	3	0	2	0	0
rare (76)	0.096	0.47	2	0	1	0	0	0	0	2	2	0	0	2	0	0	1	3	2	0	1	2	2
spur (95)	0.033	0.41	0	0	0	1	0	0	0	0	1	1	3	1	0	1	0	0	1	0	2	0	0
subtle (56)	0.021	0.36	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
surging (62)	0.094	0.41	3	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	3	0	0
suspend (104)	0.129	0.59	0	1	1	1	1	0	0	0	0	0	3	1	0	0	1	0	1	0	0	0	0
trace (99)	0.121	0.47	0	0	0	1	0	0	0	1	0	0	2	0	1	3	1	1	2	0	0	0	1
trend (114)	0.250	0.67	0	0	0	0	0	1	0	0	0	0	3	0	1	0	0	0	0	3	0	1	0
Average Native-like Level:			0.765	0.588	0.588	0.647	0.294	0.412	0.412	0.765	0.353	0.235	1.471	0.824	0.765	0.471	0.824	0.706	0.882	0.353	1.059	0.529	0.471

Appendix H: Association Proportions and Native-like Levels for ESL207-1-JC

Table H1: Association Proportion ESL207-1-JC

Stimulus Words	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19
abandon (129)	0.659	0.023	0	0.008	0.016	0	0.008	0.721	0.008	0	0.217	0.023	0.023	0	0	0	0.667	0.023	0
brood (75)	0	0	0	0	0	0	0	0	0	0	0.56	0.707	0	0	0	0	0	0.04	0
circulate (71)	0.324	0.141	0.141	0.197	0	0.197	0.141	0.014	0.338	0.141	0.141	0.07	0.141	0	0.155	0.169	0	0.465	0.451
convert (139)	0.619	0.633	0.647	0.662	0.647	0.072	0.676	0.619	0.647	0.626	0.655	0.683	0.619	0.619	0	0	0.676	0.619	0
dedicate (63)	0	0	0	0	0	0.095	0	0.397	0.19	0	0.333	0.19	0.429	0	0	0	0	0.19	0
illuminate (112)	0.732	0.045	0	0	0	0.768	0.018	0.696	0.696	0	0.938	0	0.045	0	0	0	0.893	0	0.018
launch (75)	0	0.08	0.267	0	0.213	0.04	0	0.493	0.52	0	0.067	0	0	0.133	0.027	0	0	0	0
massive (187)	0	0.38	0	0.332	0.684	0.262	0.594	0.406	0.267	0	0.332	0	0.016	0.738	0.021	0	0.738	0.262	0
plot (75)	0	0	0.013	0.533	0.52	0	0.533	0	0.52	0.053	0	0	0	0.093	0	0	0	0.56	0
peak (131)	0	0	0	0.573	0.58	0.504	0.824	0.397	0.947	0	0.534	0	0.519	0.55	0	0	0	0.443	0.824
rare (76)	0	0.013	0.671	0	0	0.105	0.184	0.132	0.184	0	0	0.145	0.184	0	0	0	0	0.145	0.513
spur (95)	0	0	0	0.011	0	0	0.011	0	0.032	0	0.379	0	0	0.379	0	0.411	0	0.053	0
subtle (56)	0.018	0	0	0.25	0	0	0.089	0	0.125	0	0	0	0	0	0	0	0	0.054	0
surging (62)	0	0	0	0.016	0	0.274	0	0	0	0	0	0	0.016	0	0	0	0	0.484	0
suspend (104)	0	0	0	0.019	0	0	0.163	0.663	0.135	0	0	0	0	0.663	0.029	0	0.01	0.087	0.087
trace (99)	0.202	0.01	0	0.202	0.01	0.03	0.545	0.202	0.01	0	0	0.02	0.374	0.202	0.202	0.101	0.04	0.232	0
trend (114)	0	0	0	0.053	0.816	0.702	0	0	0.053	0.053	0.061	0.018	0.044	0.754	0	0	0	0.061	0.702

Table H2: Native-like Levels ESL207-1-JC

Stimulus Words	Threshold	Mean Prop.	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19
abandon (129)	0.306	0.68	2	1	0	1	1	0	1	3	1	0	1	1	1	0	0	0	2	1	0
brood (75)	0.094	0.45	0	0	0	0	0	0	0	0	0	0	3	3	0	0	0	0	0	1	0
circulate (71)	0.148	0.43	2	1	1	2	0	2	1	1	2	1	1	1	1	0	2	2	0	3	3
convert (139)	0.307	0.67	2	2	2	2	2	1	3	2	2	2	2	3	2	2	0	0	3	2	0
dedicate (63)	0.040	0.36	0	0	0	0	0	2	0	3	2	0	2	2	3	0	0	0	0	2	0
illuminate (112)	0.253	0.67	3	1	0	0	0	3	1	3	3	0	3	0	1	0	0	0	3	0	1
launch (75)	0.107	0.45	0	1	2	0	2	1	0	3	3	0	1	0	0	1	1	0	0	0	0
massive (187)	0.343	0.71	0	2	0	1	2	1	2	2	1	0	1	0	1	3	1	0	3	1	0
plot (75)	0.059	0.44	0	0	1	3	3	0	3	0	3	1	0	0	0	2	0	0	0	3	0
peak (131)	0.266	0.58	0	0	0	2	3	2	3	2	3	0	2	0	2	2	0	0	0	2	3
rare (76)	0.096	0.47	0	1	3	0	0	2	2	2	2	0	0	2	2	0	0	0	0	2	3
spur (95)	0.033	0.41	0	0	0	1	0	0	1	0	1	0	2	0	0	2	0	3	0	2	0
subtle (56)	0.021	0.36	1	0	0	2	0	0	2	0	2	0	0	0	0	0	0	0	0	2	0
surging (62)	0.094	0.41	0	0	0	1	0	2	0	0	0	0	0	0	1	0	0	0	0	3	0
suspend (104)	0.129	0.59	0	0	0	1	0	0	2	3	2	0	0	0	0	3	1	0	1	1	1
trace (99)	0.121	0.47	2	1	0	2	1	1	3	2	1	0	0	1	2	2	2	1	1	2	0
trend (114)	0.250	0.67	0	0	0	1	3	3	0	0	1	1	1	1	1	3	0	0	0	1	3
Average Native-like Level:			0.706	0.588	0.529	1.118	1	1.176	1.412	1.529	1.706	0.294	1.118	0.824	1	1.176	0.412	0.353	0.765	1.647	0.824

Appendix I: Association Proportions and Native-like Levels for ESL 208-1-JJ

Table I1: Association Proportion ESL208-1-JJ

Stimulus Words	J1	J2	J3	J4	J5	J6	J7	J8	J9
abandon (129)	0.736	0.659	0.085	0	0	0.023	0.674	0.016	0
brood (75)	0	0	0	0	0	0	0	0	0
circulate (71)	0.141	0.141	0.465	0.042	0.028	0.028	0.479	0.423	0.028
convert (139)	0.022	0.619	0	0.626	0.014	0	0.813	0.209	0.619
dedicate (63)	0.016	0	0.032	0.571	0.524	0	0.048	0.127	0
illuminate (112)	0.768	0	0.893	0	0	0.009	0	0.902	0
launch (75)	0	0.013	0.12	0	0.467	0.013	0.08	0.2	0.08
massive (187)	0.824	0	0	0.005	0.353	0	0.786	0	0
plot (75)	0.013	0.04	0.267	0	0	0.027	0.04	0.733	0.013
peak (131)	0.511	0.397	0.122	0.55	0	0.427	0	0.137	0
rare (76)	0	0	0.276	0.197	0.197	0	0.013	0.671	0.092
spur (95)	0	0	0	0	0	0	0	0.147	0
subtle (56)	0.357	0	0	0	0	0	0	0.625	0
surging (62)	0	0	0	0	0	0	0	0	0
suspend (104)	0.077	0	0.135	0.01	0.01	0.798	0	0	0.019
trace (99)	0.051	0	0.202	0.384	0	0.343	0.343	0.081	0.273
trend (114)	0.746	0	0.737	0	0	0.026	0	0.754	0

Table I2: Native-like Levels ESL208-1-JJ

Stimulus Words	Threshold	Mean Prop.	J1	J2	J3	J4	J5	J6	J7	J8	J9
abandon (129)	0.306	0.68	3	2	1	0	0	1	2	1	0
brood (75)	0.094	0.45	0	0	0	0	0	0	0	0	0
circulate (71)	0.148	0.43	1	1	3	1	1	1	3	2	1
convert (139)	0.307	0.67	1	2	0	2	1	0	3	1	2
dedicate (63)	0.040	0.36	1	0	1	3	3	0	2	2	0
illuminate (112)	0.253	0.67	3	0	3	0	0	1	0	3	0
launch (75)	0.107	0.45	0	1	2	0	3	1	1	2	1
massive (187)	0.343	0.71	3	0	0	1	2	0	3	0	0
plot (75)	0.059	0.44	1	1	2	0	0	1	1	3	1
peak (131)	0.266	0.58	2	2	1	2	0	2	0	1	0
rare (76)	0.096	0.47	0	0	2	2	2	0	1	3	1
spur (95)	0.033	0.41	0	0	0	0	0	0	0	2	0
subtle (56)	0.021	0.36	2	0	0	0	0	0	0	3	0
surging (62)	0.094	0.41	0	0	0	0	0	0	0	0	0
suspend (104)	0.129	0.59	1	0	2	1	1	3	0	0	1
trace (99)	0.121	0.47	1	0	2	2	0	2	2	1	2
trend (114)	0.250	0.67	3	0	3	0	0	1	0	3	0
Average Native-like Level:			1.294	0.529	1.294	0.824	0.765	0.765	1.059	1.588	0.529

Appendix J: Association Proportions and Native-like Levels for ESL 208-1-MM

Table J1: Association Proportion ESL208-1-MM

Stimulus Words	MM1	MM2	MM3	MM4	MM5	MM6	MM7	MM8	MM9	MM10	MM11	MM12	MM13	MM14
abandon (129)	0.054	0.039	0	0.031	0	0	0.682	0.008	0.039	0.008	0.054	0.023	0.023	0.023
brood (75)	0	0	0	0	0	0	0	0	0	0.12	0	0	0	0
circulate (71)	0.338	0.225	0.07	0.141	10.71	0.014	0.141	1.211	0.493	0.268	0	0.169	0	0.183
convert (139)	0.647	0.647	0.647	0.647	0.633	0	0	0	0.194	0.626	0.079	0.022	0.619	0.007
dedicate (63)	0.111	0.016	0.095	0	0	0.016	0	0.095	0.317	0	0.048	0.032	0	0
illuminate (112)	0.741	0.741	0.205	0.741	0	0.107	0.045	0.018	0	0.696	0	0	0.009	0.705
launch (75)	0	0	0	1	0	0.013	0	0	0.453	0.107	0.853	0.013	0.027	0.013
massive (187)	1	0.053	0.406	0.824	0	0	0.016	0	0.07	0.412	0.893	0.005	0.016	0.011
plot (75)	0	0	0.013	0.013	0	0.52	0	0	0.173	0	0	0	0.04	0
peak (131)	0.099	0.42	0.122	0.099	0	0	0	0	0.832	0.824	0.435	0.824	0.55	0.412
rare (76)	0	0.013	0.184	0.132	0.184	0	0.039	0	0.132	0	0.013	0.184	0	0
spur (95)	0	0	0	0.053	0.379	0	0	0	0	0	0	0	0	0
subtle (56)	0.054	0	0	0.054	0	0.018	0.036	0	0	0.125	0	0	0.161	0
surging (62)	0	0	0	0	0	0	0	0	0	0	0	0	0.016	0
suspend (104)	0.663	0.077	0.029	0.663	0	0	0.01	0	0.048	0.798	0	0	0.019	0.663
trace (99)	0	0	0	0.374	0.081	0	0	0	0.03	0	0.081	0.081	0.222	0.343
trend (114)	0.868	0	0.746	0.754	0	0.009	0	0	0	0.061	0	0	0.825	0

Table J2: Native-like Levels ESL208-1-MM

Stimulus Words	Threshold	Mean Prop.	MM1	MM2	MM3	MM4	MM5	MM6	MM7	MM8	MM9	MM10	MM11	MM12	MM13	MM14
abandon (129)	0.306	0.68	1	1	0	1	0	0	3	1	1	1	1	1	1	1
brood (75)	0.094	0.45	0	0	0	0	0	0	0	0	0	2	0	0	0	0
circulate (71)	0.148	0.43	2	2	1	1	1	1	1	1	3	2	0	2	0	2
convert (139)	0.307	0.67	2	2	2	2	2	0	0	0	1	2	1	1	2	1
dedicate (63)	0.040	0.36	2	1	2	0	0	1	0	2	2	0	2	1	0	0
illuminate (112)	0.253	0.67	3	3	1	3	0	1	1	1	0	3	0	0	1	3
launch (75)	0.107	0.45	0	0	0	3	0	1	0	0	3	2	3	1	1	1
massive (187)	0.343	0.71	3	1	2	3	0	0	1	0	1	2	3	1	1	1
plot (75)	0.059	0.44	0	0	1	1	0	3	0	0	2	0	0	0	1	0
peak (131)	0.266	0.58	1	2	1	1	0	0	0	0	3	3	2	3	2	2
rare (76)	0.096	0.47	0	1	2	2	2	0	1	0	2	0	1	2	0	0
spur (95)	0.033	0.41	0	0	0	2	2	0	0	0	0	0	0	0	0	0
subtle (56)	0.021	0.36	2	0	0	2	0	1	2	0	0	2	0	0	2	0
surging (62)	0.094	0.41	0	0	0	0	0	0	0	0	0	0	0	0	1	0
suspend (104)	0.129	0.59	3	1	1	3	0	0	1	0	1	3	0	0	1	3
trace (99)	0.121	0.47	0	0	0	2	1	0	0	0	1	0	1	1	2	2
trend (114)	0.250	0.67	3	0	3	3	0	1	0	0	0	1	0	0	3	0
Average Native-like Level:			1.294	0.824	0.941	1.706	0.471	0.529	0.588	0.294	1.176	1.353	0.824	0.765	1.059	0.941

Appendix K: Association Proportions and Native-like Levels for ESL209-1-BB

Table K1:

Association Proportion ESL209-1-BB

Stimulus Words	BB1	BB2	BB3	BB4
abandon (129)	0.016	0.023	0.023	0.008
brood (75)	0	0.107	0	0
circulate (71)	0.197	0.141	0.197	0.127
convert (139)	0.676	0.633	0.647	0.072
dedicate (63)	0.19	0.397	0.19	0.048
illuminate (112)	0.196	0.902	0	0.669
launch (75)	0.453	0.12	0.547	0.52
massive (187)	0.016	0.738	0.235	0
plot (75)	0.147	0.52	0.053	0.56
peak (131)	0.573	0.557	0.427	0.527
rare (76)	0.184	0.013	0	0.132
spur (95)	0.074	0	0	0
subtle (56)	0.054	0	0	0
surging (62)	0	0.645	0.048	0.048
suspend (104)	0.144	0.135	0.029	0
trace (99)	0.202	0	0.081	0
trend (114)	0	0.702	0.763	0

Table K2:

Native-like Levels ESL209-1-BB

Stimulus Words	Threshold	Mean Prop.	BB1	BB2	BB3	BB4
abandon (129)	0.306	0.68	1	1	1	1
brood (75)	0.094	0.45	0	2	0	0
circulate (71)	0.148	0.43	2	1	2	1
convert (139)	0.307	0.67	3	2	2	1
dedicate (63)	0.040	0.36	2	3	2	2
illuminate (112)	0.253	0.67	1	3	0	2
launch (75)	0.107	0.45	3	2	3	3
massive (187)	0.343	0.71	1	3	1	0
plot (75)	0.059	0.44	2	3	1	3
peak (131)	0.266	0.58	2	2	2	2
rare (76)	0.096	0.47	2	1	0	2
spur (95)	0.033	0.41	2	0	0	0
subtle (56)	0.021	0.36	2	0	0	0
surging (62)	0.094	0.41	0	3	1	1
suspend (104)	0.129	0.59	2	2	1	0
trace (99)	0.121	0.47	2	0	1	0
trend (114)	0.250	0.67	0	3	3	0
Average Native-like Level:			1.588	1.824	1.176	1.059

Appendix L: Association Proportions and Native-like Levels for ESL 209-2-BB

Table L1: Association Proportions ESL209-2-BB

Stimulus Words	2B1	2B2	2B3	2B4	2B5	2B6	2B7	2B8	2B9	2B10	2B11	2B12	2B13	2B14	2B15	2B16
abandon (129)	0.395	0.047	0.031	0.008	0	0	0	0.023	0.023	0.721	0.023	0.008	0	0.124	0	0.008
brood (75)	0	0	0	0	0	0	0	0.44	0	0	0	0	0	0	0	0
circulate (71)	0.296	0.423	0.085	0.437	0.592	0.225	0.155	0.085	0.141	0.028	0.606	0.155	0	0.141	0	0.127
convert (139)	0.813	0.036	0.662	0.619	0.64	0.014	0.619	0.029	0.662	0.626	0.043	0	0.194	0	0.647	0.64
dedicate (63)	0.429	0.476	0.032	0.429	0.016	0.016	0	0	0	0.429	0	0	0	0	0.397	0
illuminate (112)	0	0.009	0	0	0	0.714	0.036	0.696	0.045	0.696	0	0.893	0	0	0	0.107
launch (75)	0.173	0.773	0.28	0	0.053	0.013	0.52	0	0	0	0.493	0	0.107	0	0	0
massive (187)	0.406	0	0.016	0.738	0.262	0	0.754	0.422	0	0.005	0.059	0	0.743	0.332	0	0.406
plot (75)	0.52	0.053	0.173	0	0.04	0	0.52	0.533	0	0	0	0.027	0	0	0	0
peak (131)	0.008	0.427	0.397	0.405	0.527	0.176	0.397	0.176	0	0.397	0	0	0.519	0	0	0.084
rare (76)	0.118	0	0.224	0.197	0.342	0.092	0.329	0.184	0	0.013	0.184	0	0.013	0	0	0
spur (95)	0.158	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
subtle (56)	0.125	0	0	0	0	0.018	0.018	0.429	0	0.125	0	0	0	0	0	0
surging (62)	0	0	0	0	0	0.065	0	0	0.484	0.919	0	0	0	0	0	0
suspend (104)	0	0	0	0	0	0	0	0.308	0	0.077	0	0	0	0.019	0.135	0.692
trace (99)	0.343	0	0	0.232	0.545	0.081	0.04	0.081	0	0.202	0	0.03	0.202	0	0	0.081
trend (114)	0.825	0	0	0.711	0.728	0.035	0	0	0	0	0.114	0	0	0	0	0

Table L2: Native-like Levels ESL209-2-BB

Stimulus Words	Threshold	Mean Prop.	2B1	2B2	2B3	2B4	2B5	2B6	2B7	2B8	2B9	2B10	2B11	2B12	2B13	2B14	2B15	2B16
abandon (129)	0.306	0.68	2	1	1	1	0	0	0	1	1	3	1	1	0	1	0	1
brood (75)	0.094	0.45	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0
circulate (71)	0.148	0.43	2	2	1	3	3	2	2	1	1	1	3	2	0	1	0	1
convert (139)	0.307	0.67	3	1	2	2	2	1	2	1	2	2	1	0	1	0	2	2
dedicate (63)	0.040	0.36	3	3	1	3	1	1	0	0	0	3	0	0	0	0	3	0
illuminate (112)	0.253	0.67	0	1	0	0	0	3	1	3	1	3	0	3	0	0	0	1
launch (75)	0.107	0.45	2	3	2	0	1	1	3	0	0	0	3	0	2	0	0	0
massive (187)	0.343	0.71	2	0	1	3	1	0	3	2	0	1	1	0	3	1	0	2
plot (75)	0.059	0.44	3	1	2	0	1	0	3	3	0	0	0	1	0	0	0	0
peak (131)	0.266	0.58	1	2	2	2	2	1	2	1	0	2	0	0	3	0	0	1
rare (76)	0.096	0.47	2	0	2	2	2	1	2	2	0	1	2	0	1	0	0	0
spur (95)	0.033	0.41	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
subtle (56)	0.021	0.36	2	0	0	0	0	1	1	3	0	2	0	0	0	0	0	0
surging (62)	0.094	0.41	0	0	0	0	0	1	0	0	3	3	0	0	0	0	0	0
suspend (104)	0.129	0.59	0	0	0	0	0	0	0	2	0	1	0	0	0	1	2	3
trace (99)	0.121	0.47	2	0	0	2	3	1	1	1	0	2	0	1	2	0	0	1
trend (114)	0.250	0.67	3	0	0	3	3	1	0	0	0	0	1	0	0	0	0	0
Average Native-like Level:			1.706	0.824	0.824	1.235	1.118	0.824	1.176	1.294	0.471	1.412	0.706	0.471	0.706	0.235	0.412	0.706

Appendix M: Association Proportions and Native-like Levels for ESL209-1-LM

Table M1:

Association Proportion ESL209-1-LM

Stimulus Words	L1	L2	L3	L4
abandon (129)	0.023	0.008	0.039	0.86
brood (75)	0.267	0	0	0
circulate (71)	0.366	0.197	0.197	0.268
convert (139)	0.633	0.633	0	0.806
dedicate (63)	0.397	0	0.619	0
illuminate (112)	0	0	0.054	0.009
launch (75)	0.013	0.507	0.573	0.387
massive (187)	0.262	0.262	0.455	0.406
plot (75)	0	0	0.173	0.133
peak (131)	0.924	0.427	0.824	0.595
rare (76)	0.053	0	0.645	0.026
spur (95)	0.379	0	0	0
subtle (56)	0.554	0	0.125	0
surging (62)	0.484	0.065	0	0
suspend (104)	0.173	0	0.135	0
trace (99)	0.202	0.081	0.101	0
trend (114)	0.702	0.061	0.035	0.061

Table M2:

Native-like Levels ESL209-1-LM

Stimulus Words	Threshold	Mean Prop	L1	L2	L3	L4
abandon (129)	0.306	0.68	1	1	1	3
brood (75)	0.094	0.45	2	0	0	0
circulate (71)	0.148	0.43	2	2	2	2
convert (139)	0.307	0.67	2	2	0	3
dedicate (63)	0.040	0.36	3	0	3	0
illuminate (112)	0.253	0.67	0	0	1	1
launch (75)	0.107	0.45	1	3	3	2
massive (187)	0.343	0.71	1	1	2	2
plot (75)	0.059	0.44	0	0	2	2
peak (131)	0.266	0.58	3	2	3	3
rare (76)	0.096	0.47	1	0	3	1
spur (95)	0.033	0.41	2	0	0	0
subtle (56)	0.021	0.36	3	0	2	0
surging (62)	0.094	0.41	3	1	0	0
suspend (104)	0.129	0.59	2	0	2	0
trace (99)	0.121	0.47	2	1	1	0
trend (114)	0.250	0.67	3	1	1	1
Average Native-like Level:			1.824	0.824	1.529	1.176

Appendix N: Association Proportions and Native-like Level for ESL209-1-ML

Table N1: Association Proportion ESL209-1-ML

Stimulus Word	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17
abandon (129)	0.008	0.721	0.667	0.047	0.388	0	0	0.667	0	0.008	0.659	0.682	0	0	0.698	0	0.124
brood (75)	0	0.027	0.147	0.133	0	0	0	0	0	0	0	0	0	0	0	0	0
circulate (71)	0.127	0.07	0.465	0.282	0.69	0.493	0	0.338	0.141	0.056	0.028	0.028	0	0.507	0.408	0.028	0.141
convert (139)	0.64	0.633	0.194	0.64	0.619	0	0.647	0.655	0.619	0	0.633	0.619	0.619	0.655	0.619	0.647	0.64
dedicate (63)	0.175	0	0.111	0.063	0.238	0.127	0.143	0	0	0.19	0.397	0	0.016	0.032	0	0.19	0.016
illuminate (112)	0.696	0.821	0.902	0.696	0.83	0	0.696	0	0	0	0	0	0	0.732	0.813	0	0
launch (75)	0.16	0.733	0.013	0.04	0.133	0.373	0	0.52	0.067	0.067	0.533	0.493	0.493	0	0.013	0	0.453
massive (187)	0.273	0.348	0.07	0.487	0	0.016	0.738	0.011	0.668	0	0.005	0.749	0.005	0.332	0.594	0.738	0.332
plot (75)	0.16	0.173	0.68	0.067	0	0.133	0.053	0.267	0	0.147	0.04	0.027	0	0	0	0	0.133
peak (131)	0.832	0.427	0.183	0.55	0.427	0.214	0	0.107	0.122	0.176	0	0.412	0.427	0.435	0.221	0	0
rare (76)	0.092	0.026	0	0.039	0.013	0	0	0.513	0	0.013	0.526	0.039	0	0.105	0.026	0	0.053
spur (95)	0	0	0.137	0	0	0	0	0.137	0	0	0	0	0.379	0	0	0	0
subtle (56)	0.161	0	0	0	0	0	0	0	0	0.482	0	0	0	0.018	0.321	0	0
surging (62)	0	0.016	0	0	0	0	0	0	0	0	0	0	0	0	0	0.048	0.226
suspend (104)	0	0.135	0.144	0.038	0	0	0	0	0.135	0	0.087	0.01	0.01	0	0.663	0	0.087
trace (99)	0.02	0.232	0.414	0.283	0.545	0	0.414	0.202	0.424	0.071	0	0.01	0.061	0.03	0.091	0	0.081
trend (114)	0.009	0.711	0	0	0	0.702	0	0	0.746	0	0	0.061	0.702	0	0	0	0

Table N2: Native-like Levels ESL209-1-ML

Stimulus Words	Threshold	Mean Prop.	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17
abandon (129)	0.306	0.68	1	3	2	1	2	0	0	2	0	1	2	3	0	0	3	0	1
brood (75)	0.094	0.45	0	1	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0
circulate (71)	0.148	0.43	1	1	3	2	3	3	0	2	1	1	1	1	0	3	2	1	1
convert (139)	0.307	0.67	2	2	1	2	2	0	2	2	2	0	2	2	2	2	2	2	2
dedicate (63)	0.040	0.36	2	0	2	2	2	2	2	0	0	2	3	0	1	1	0	2	1
illuminate (112)	0.253	0.67	3	3	3	3	3	0	3	0	0	0	0	0	0	3	3	0	0
launch (75)	0.107	0.45	2	3	1	1	2	2	0	3	1	1	3	3	3	0	1	0	3
massive (187)	0.343	0.71	1	2	1	2	0	1	3	1	2	0	1	3	1	1	2	3	1
plot (75)	0.059	0.44	2	2	3	2	0	2	1	2	0	2	1	1	0	0	0	0	2
peak (131)	0.266	0.58	3	2	1	2	2	1	0	1	1	1	0	2	2	2	1	0	0
rare (76)	0.096	0.47	1	1	0	1	1	0	0	3	0	1	3	1	0	2	1	0	1
spur (95)	0.033	0.41	0	0	2	0	0	0	0	2	0	0	0	0	2	0	0	0	0
subtle (56)	0.021	0.36	2	0	0	0	0	0	0	0	0	3	0	0	0	1	2	0	0
surging (62)	0.094	0.41	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2
suspend (104)	0.129	0.59	0	2	2	1	0	0	0	0	2	0	1	1	1	0	3	0	1
trace (99)	0.121	0.47	1	2	2	2	3	0	2	2	2	1	0	1	1	1	1	0	1
trend (114)	0.250	0.67	1	3	0	0	0	3	0	0	3	0	0	1	3	0	0	0	0
Average Native-like Level:			1.294	1.647	1.471	1.353	1.176	0.824	0.765	1.176	0.824	0.765	1	1.118	0.941	0.941	1.235	0.529	0.941

Appendix O: Association Proportions and Native-like Levels for ESL209-2-ML

Table O1: Association Proportion ESL209-2-ML

Stimulus Words	2M1	2M2	2M3	2M4	2M5	2M6	2M7	2M8	2M9	2M10	2M11	2M12	2M13	2M14	2M15	2M16	2M17	2M18
abandon (129)	0	0.698	0.023	0.163	0.023	0.682	0.698	0.023	0.783	0	0.016	0.008	0.023	0	0.031	0.031	0	0.054
brood (75)	0	0	0	0	0	0.013	0.027	0	0.027	0	0	0	0	0	0	0	0	0
circulate (71)	0.408	0.169	0	0.437	0.141	0.493	0.07	0.028	0.127	0	0.437	0	0.394	0.141	0	0.028	0.254	0.141
convert (139)	0	0.619	0.849	0.655	0.633	0.194	0	0	0.619	0	0.813	0	0.619	0.619	0.64	0.662	0.619	0
dedicate (63)	0.19	0	0.048	0.143	0.397	0.111	0.397	0	0	0	0	0	0	0	0	0.095	0	0
illuminate (112)	0.741	0.696	0	0	0	0.741	0	0.696	0.25	0	0.705	0	0.009	0.696	0	0	0.741	0
launch (75)	0.453	0.24	0	0.12	0.307	0.36	0.013	0.467	0	0	0.533	0	0.067	0	0.733	0	0	0.013
massive (187)	0.337	0.519	0.048	0.749	0	0.016	0.406	0.406	0	0.332	0	0	0.594	0.332	0	0	0	1
plot (75)	0.187	0	0.147	0.04	0.027	0	0	0	0	0.267	0.053	0.013	0	0	0.213	0.04	0	0.027
peak (131)	0.725	0.015	0.26	0.008	0.573	0.427	0.427	0.397	0.397	0.397	0.824	0	0.397	0.397	0	0.435	0.397	0.824
rare (76)	0.026	0.039	0	0	0.132	0	0	0.184	0.013	0	0.184	0	0.184	0	0	0.553	0	0
spur (95)	0.021	0	0.011	0	0	0.021	0	0	0.274	0	0	0	0	0	0	0.379	0	0.042
subtle (56)	0	0	0.054	0.071	0	0.125	0.054	0.036	0	0	0	0	0	0	0	0	0	0.018
surging (62)	0.339	0	0	0	0	0.129	0	0	0.016	0	0.145	0.016	0	0.048	0	0	0	0.919
suspend (104)	0.087	0.663	0.135	0	0.135	0	0.087	0	0.183	0	0.01	0	0.308	0.135	0.663	0	0.837	0.135
trace (99)	0.03	0	0.071	0.061	0	0.343	0	0	0.232	0	0.596	0.01	0.081	0	0	0.232	0	0.202
trend (114)	0.825	0.018	0.018	0.763	0	0.702	0.053	0.728	0	0	0.763	0	0	0.711	0.07	0	0	0

Table O2: Native-like Levels ESL209-2-ML

Stimulus Words	Threshold	Mean Prop.	2M1	2M2	2M3	2M4	2M5	2M6	2M7	2M8	2M9	2M10	2M11	2M12	2M13	2M14	2M15	2M16	2M17	2M18
abandon (129)	0.306	0.68	0	3	1	1	1	3	3	1	3	0	1	1	1	0	1	1	0	1
brood (75)	0.094	0.45	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0
circulate (71)	0.148	0.43	2	2	0	3	1	3	1	1	1	0	3	0	2	1	0	1	2	1
convert (139)	0.307	0.67	0	2	3	2	2	1	0	0	2	0	3	0	2	2	2	2	2	0
dedicate (63)	0.040	0.36	2	0	2	2	3	2	3	0	0	0	0	0	0	0	0	2	0	0
illuminate (112)	0.253	0.67	3	3	0	0	0	3	0	3	1	0	3	0	1	3	0	0	3	0
launch (75)	0.107	0.45	3	2	0	2	2	2	1	3	0	0	3	0	1	0	3	0	0	1
massive (187)	0.343	0.71	1	2	1	3	0	1	2	2	0	1	0	0	2	1	0	0	0	3
plot (75)	0.059	0.44	2	0	2	1	1	0	0	0	0	2	1	1	0	0	2	1	0	1
peak (131)	0.266	0.58	3	1	1	1	2	2	2	2	2	2	3	0	2	3	0	2	2	3
rare (76)	0.096	0.47	1	1	0	0	2	0	0	2	1	0	2	0	2	0	0	3	0	0
spur (95)	0.033	0.41	1	0	1	0	0	1	0	0	2	0	0	0	0	0	0	2	0	2
subtle (56)	0.021	0.36	0	0	2	2	0	2	2	2	0	0	0	0	0	0	0	0	0	1
surging (62)	0.094	0.41	2	0	0	0	0	2	0	0	1	0	2	1	0	1	0	0	0	3
suspend (104)	0.129	0.59	1	3	2	0	2	0	1	0	2	0	1	0	2	2	3	0	3	2
trace (99)	0.121	0.47	1	0	1	1	0	2	0	0	2	0	3	1	1	0	0	2	0	2
trend (114)	0.250	0.67	3	1	1	3	0	3	1	3	0	0	3	0	0	3	1	0	0	0
Average Native-like Level:			1.471	1.176	1	1.235	0.941	1	1	1.118	1.059	0.294	1.647	0.235	0.941	0.941	0.706	0.941	0.706	1.176

Appendix P: Association Proportions and Native-like Levels for ESL209-1-ZG

Table P1: Association Proportion ESL209-1-ZG

Stimulus Words	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9	Z10	Z11	Z12	Z13	Z14	Z15
abandon (129)	0.039	0.031	0.062	0.023	0.016	0.016	0	0.031	0.884	0.023	0	0.233	0.023	0	0.023
brood (75)	0	0	0	0	0.013	0.013	0	0.56	0	0	0.013	0.013	0	0	0
circulate (71)	0.437	0.183	0.352	0.437	0.141	0	0.324	0.549	0.141	0.197	0.028	0.268	0	0.014	0
convert (139)	0.223	0.014	0.007	0.842	0.619	0.813	0.626	0.043	0.633	0.662	0.619	0.619	0	0	0.647
dedicate (63)	0.016	0.032	0	0.19	0	0	0	0	0.556	0.19	0	0.048	0	0	0.429
illuminate (112)	0.036	0.893	0.705	0.018	0	0.009	0.045	0.75	0.152	0.696	0.696	0.045	0	0.045	1
launch (75)	0.04	0.56	0	0	0.013	0.067	0	0.453	0.2	0	0.027	0.493	0	0	0.027
massive (187)	0.332	0.348	0	0	0.337	0.011	0	0.278	0	0.332	0	0.374	0	0.332	0
plot (75)	0.04	0.013	0.04	0	0.04	0.173	0	0.52	0.52	0.027	0.053	0.053	0	0.013	0.733
peak (131)	0.122	0.122	0.527	0.427	0	0.397	0	0.427	0.443	0.435	0.824	0.466	0	0	0.519
rare (76)	0	0.224	0.184	0	0	0.184	0.513	0.039	0.184	0	0	0.013	0	0.013	0
spur (95)	0	0.379	0.379	0.011	0	0	0	0	0.011	0.011	0	0	0	0	0.379
subtle (56)	0	0	0.018	0.196	0	0.107	0	0.321	0	0.196	0.018	0	0	0.018	0
surging (62)	0	0.161	0	0.484	0	0	0	0.5	0.048	0.016	0.274	0.484	0	0.048	0
suspend (104)	0	0.183	0	0.067	0	0.135	0	0.154	0.692	0	0	0.01	0	0.231	0.135
trace (99)	0	0.343	0.202	0.01	0	0	0.02	0.202	0.121	0	0.01	0	0.253	0.354	0.051
trend (114)	0.702	0.763	0.053	0.702	0	0	0	0.754	0.132	0	0	0.07	0	0.018	0.07

Table P2: Native-like Levels ESL209-1-ZG

Stimulus Words	Threshold	Mean Prop.	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9	Z10	Z11	Z12	Z13	Z14	Z15
abandon (129)	0.306	0.68	2	1	1	1	1	1	0	1	3	1	0	1	1	0	1
brood (75)	0.094	0.45	0	0	0	0	1	1	0	3	0	0	1	1	0	0	0
circulate (71)	0.148	0.43	3	2	2	3	1	0	2	3	1	2	1	2	0	1	0
convert (139)	0.307	0.67	1	1	1	3	2	3	2	1	2	2	2	2	0	0	2
dedicate (63)	0.040	0.36	1	1	0	2	0	0	0	0	3	2	0	2	0	0	3
illuminate (112)	0.253	0.67	1	3	3	1	0	1	1	3	1	3	3	1	0	1	3
launch (75)	0.107	0.45	1	3	0	0	1	1	0	3	2	0	1	3	0	0	1
massive (187)	0.343	0.71	1	2	0	0	1	1	0	1	0	1	0	2	0	1	0
plot (75)	0.059	0.44	1	1	1	0	1	2	0	1	3	1	1	1	0	1	3
peak (131)	0.266	0.58	1	1	2	2	0	2	0	2	2	2	3	2	0	0	2
rare (76)	0.096	0.47	0	2	2	0	0	2	3	1	2	0	0	1	0	1	0
spur (95)	0.033	0.41	0	2	2	1	0	0	0	0	1	1	0	0	0	0	2
subtle (56)	0.021	0.36	0	0	1	2	0	2	0	2	0	2	1	0	0	1	0
surging (62)	0.094	0.41	0	2	0	3	0	0	0	3	1	1	2	3	0	1	0
suspend (104)	0.129	0.59	0	2	0	1	0	2	0	2	3	0	0	1	0	2	2
trace (99)	0.121	0.47	0	2	2	1	0	0	1	2	2	0	1	0	2	2	1
trend (114)	0.250	0.67	3	3	1	3	0	0	0	3	1	0	0	1	0	1	1
Average Native-like Level:			0.882	1.647	1.059	1.353	0.471	1.059	0.529	1.824	1.588	1.059	0.941	1.353	0.176	0.706	1.235