SSLA, **27**, 567–595. Printed in the United States of America.
DOI: 10.1017/S0272263105050254

RELATIONSHIP BETWEEN LEXICAL COMPETENCE AND LANGUAGE PROFICIENCY

Variable Sensitivity

Alla Zareva
Northern Arizona University

Paula Schwanenflugel
University of Georgia

Yordanka Nikolova
Bourgas Free University

The purpose of the present study was to determine what features associated with the macrolevel of lexical competence vary as a function of an increase in second language (L2) proficiency. The macrolevel of participants' word knowledge was described with respect to six variables that are commonly associated with three proposed macrolevel dimensions, namely quantity, quality, and metacognitive awareness. Sixty-four participants (native speakers of English, L2 advanced learners, and intermediate learners of English) self-rated their familiarity with 73 lexical items and were asked to generate word associations to the words they identified in a verifiable way as known. The data analyses showed that some measures, such as vocabulary size, word frequency effects, number of associations, and within-group consistency of participants' associative domain, are more sensitive to L2 learners' increasing proficiency than others (e.g., nativelike commonality of associations). We thus conclude that some aspects, such as quality and quantity of L2 lexical competence, develop as the proficiency of the L2 learners increases, whereas others, such as learn-

Address correspondence to: Alla Zareva, Northern Arizona University, Department of English, Applied Linguistics Program, Flagstaff, AZ 86011; e-mail: Alla.Zareva@nau.edu.

© 2005 Cambridge University Press 0272-2631/05 $12.00

https://doi.org/10.1017/S0272263105050254 Published online by Cambridge University Press

ers' metacognitive awareness, are not proficiency dependent. We also suggest that the measures that were identified as sensitive to capturing the overall state of L2 learners' vocabularies would also be reliable indexes of learners' proficiency development.

The potential of second language (L2) vocabulary knowledge to serve as a predictor of L2 learners' proficiency has recently attracted much interest among researchers. As the interest in L2 vocabulary has grown, an increasing awareness that vocabulary knowledge is multidimensional has emerged. One of the goals of the study of L2 vocabulary is to capture important lexical dimensions with respect to the language proficiency of the L2 learners so that test instruments that are theory-driven, empirically tested, and practical in measuring different aspects of word knowledge can be developed. In this regard, Read (2000) suggested that an issue that has not received sufficient attention in L2 lexical research is the relationship between language proficiency and the overall state of a learner's vocabulary. It is generally assumed that as the proficiency level increases, so does the general state of lexical knowledge of the L2 learners. However, fundamental issues remain. For example, it is unclear what the similarities and differences in lexical knowledge are between native speakers (NSs) and nonnative speakers (NNSs) at various levels of proficiency. Consequently, we do not currently know which measures are sensitive to capturing differences in the vocabulary knowledge of language users at different levels of language proficiency.

As Meara (1996) noted, "knowing a word" is a multifaceted task for language users and the assessment of that knowledge is a "mammoth task for the test constructor" (p. 46). Part of the issue resides in the large number of aspects associated with the notion of a *word* and, respectively, with the criteria for what knowing a word entails. Some researchers (e.g., Gass & Selinker, 2001; Nation, 1990, 2001; Richards, 1976) have proposed models of lexical knowledge consisting of separate traits (known as *separate trait models*). For example, Nation (2001) listed the following aspects of lexical knowledge: (a) *form* (spoken and written form), (b) *position* (grammatical behavior and collocational patterns), (c) *function* (word frequency and appropriateness), and (d) *meaning* (conceptual content and word associations). The problem with the proposed descriptive criteria is mostly practical. As noted elsewhere in the literature, although it is theoretically possible to describe what it means to know a word, it is hardly possible in practice to design a test that would measure all of the previously mentioned traits for all of the words an individual knows. This practical concern calls for a distinction between assessment of how well a particular set of words is known and assessment of the overall state of a learner's vocabulary. In this regard, as Meara proposed, having a small number of measurable dimensions that reflect properties of the lexicon as a whole—rather than properties of individual items—would make it possi-

https://doi.org/10.1017/S0272263105050254 Published online by Cambridge University Press

ble for researchers to examine language users' general state of the lexicon in its entirety.

Currently, existing models that promote several global dimensions for description of the lexicon as a whole are referred to as *global trait models* (e.g., Henriksen, 1999). In L2 research, global trait models of vocabulary knowledge suggest two or three dimensions that might be important for examining L2 learners' lexicon. Among these, there seems to be a great deal of agreement on the importance of vocabulary size—often referred to as *breadth*—as a key dimension of lexical competence, and for a long time, it was the only dimension to receive considerable research attention. However, in the last couple of decades, L2 researchers began to realize that knowledge of words should not be studied unidimensionally (i.e., with respect to size only because other possible dimensions might play as important a role as size in distinguishing meaningful differences in L1 and L2 vocabulary knowledge). Thus, a second global dimension of the lexicon that is often posited is quality—or *depth*—of lexical knowledge (Dolch & Leeds, 1953, for L1 vocabulary knowledge; Greidanus & Nienhuis, 2001; Henriksen; Qian & Schedl, 2004; Read, 1993, 2000; Vermeer, 2001; Wesche & Paribakht, 1996; Wolter, 2001). However, there is little agreement among L2 researchers about what quality or depth entails. One proposal is that depth of vocabulary knowledge refers to knowledge of vocabulary aspects such as pronunciation, orthography, morphosyntactic and semantic features, register, collocations, and so forth that describe properties of lexical items as outlined by Richards (1976) and Nation (2001) (see Chapelle, 1998; Henriksen; Qian & Schedl). One serious question that this understanding of depth raises is how this dimension is different from what might be subsumed under *size*; that is, when we make claims concerning the vocabulary size of L2 learners, do we not also mean that they have knowledge of some properties, such as pronunciation, orthography, morphosyntactic features, semantic features, among others, for a given number of words? Put differently, if knowledge of word properties is the depth dimension, what does knowing a particular number of words entail?

Some recent attempts in examining quality or depth of word knowledge employ associative behavior as an indicator of depth, which assumes that the study of the connectivity among words in the L2 mental lexicon is an approach to studying the way semantic information is structured in memory (Nelson, 1977). The researchers who have used association tests share the common understanding that the associations provided by L2 learners are revealing of their associative patterns of meaning connections. Yet, they differ in how these association patterns can be used to infer quality of vocabulary knowledge. For example, Schmitt (1998) proposed that the attribute on which quality of L2 responses should be judged is the degree of nativelikeness of L2 learners associations. There is a high level of commonality among NSs' responses and—to the extent that a similar degree of associative commonality is revealed by L2 speakers—this is viewed as indicative of NNSs' nativelikeness of associative behavior; hence, this commonality is seen as an indicator of their depth

https://doi.org/10.1017/S0272263105050254 Published online by Cambridge University Press

of lexical knowledge. Another way of interpreting associative patterns is linked to association studies conducted by Meara (1978, 1984; Wilks & Meara, 2002). In this view, it is important to examine how words are organized in L2 learners' mental lexicon with respect to whether semantic connections are organized similarly to the ones that NSs have or whether language learners simply tag L2 words into their L1 lexicons. Several L2 researchers (e.g., Greidanus & Nienhuis, 2001; Read, 1993; Vermeer, 2001; Wolter, 2001) have adopted Meara's point of view in their research and have examined comparatively the type of associative or meaning connections (e.g., paradigmatic, syntagmatic, phonological) that NSs and L2 learners of English tend to build among the words in their mental lexicon.

Finally, there is a third dimension that is sometimes posited in the literature (e.g., Henriksen, 1999): the receptive-productive dimension, which in some frameworks is used as a bridging dimension between lexical competence and performance. In general, L2 researchers agree that there is a difference between vocabulary used for comprehension (receptive vocabulary) and vocabulary used for production (productive vocabulary). They agree that word comprehension does not automatically predict its (correct) use (Henriksen; Nation, 1990). However, it is generally acknowledged that the receptive-productive distinction is not a dichotomous one but a continuum, where a word passes a threshold from receptive into productive use (e.g., Henriksen; Melka, 1997; Read, 2000). In this regard, Melka suggested *word familiarity* as the variable that distinguishes where a particular word will be on the receptive-productive continuum. Relatedly, other researchers (e.g., Meara, 1996) have proposed degree of *automaticity* as a determinant of receptive versus productive knowledge. Laufer and Paribakht (1998), on the other hand, identified *word frequency* as the factor that motivates movement of lexical items from the receptive to the productive domain. Finally, Read offered a narrower distinction by arguing that reception and production are too broad terms; therefore, researchers should be specific about whether it is recognition, recall, comprehension, or use that is studied.

In summary, the discussion on lexical competence has come to emphasize that lexical knowledge is a complex phenomenon that falls on a continuum in all of the previously discussed dimensions. Researchers within the separate trait paradigm proposed studying word knowledge with regard to all aspects of knowing a word; hence, they put forward a set of descriptive criteria for what it means to know a word. On the other hand, researchers who favored a global approach (i.e., examining lexical competence in its entirety) proposed two (breadth and depth; Greidanus & Nienhuis, 2001; Read, 1993, 2000; Vermeer, 2001; Wesche & Paribakht, 1996; Wolter, 2001), or three (quantity, quality, and receptive-productive control; Henriksen, 1999) or, possibly, four global dimensions (quantity, quality, receptive-productive control, and lexical organization; Chapelle, 1998; Qian & Schedl, 2004) for describing the overall state of L2 learners' vocabulary knowledge. According to Meara (1996), the first two dimensions offer a rich framework for describing different aspects of lexical

https://doi.org/10.1017/S0272263105050254 Published online by Cambridge University Press

competence and point to some interesting relationships between vocabulary growth and connectivity of the lexicon. He also hypothesized that the size dimension becomes less important as the size of vocabulary gets larger and that the quality dimension takes precedence after a threshold of 5,000–6,000 words is passed. Therefore, given that L2 learners' vocabulary size is considered to be generally smaller compared to the vocabulary size of NSs, the importance of it being structured increases considerably. However, although the two-dimensional model has received some empirical support, the three- and four-dimensional models are largely hypotheses at this time, and their effectiveness in accounting for vocabulary growth as proficiency increases has not been empirically tested as a whole.

In this paper, we use a three-dimensional model to study the L2 lexicon and to examine the potential of this model to account for differences in participants' word knowledge as a function of their proficiency. The expectation is that by examining three dimensions of word knowledge, we can determine whether it is necessary to obtain information about all dimensions of lexical competence or whether it is possible that some dimensions are more revealing of the overall state of learners' vocabulary than others. In our framework, there are two major distinctions that we consider important in the study of lexical competence. The first major distinction is between receptive and productive vocabulary (i.e., examining word knowledge with regard to learners' ability to recognize a word as well as with respect to their ability to use it productively). The second major distinction is between the microlevel and the macrolevel of word knowledge, which applies to both the productive and the receptive domains. The microlevel refers to knowledge of various aspects of individual words, which are related to the linguistic levels of word description (e.g., knowledge of a word as a phonetic, orthographic, phonological, morphological, syntactic, or semantic entry). The macrolevel, on the other hand, is the level that describes the overall state of an individual's lexicon within a framework of three proposed dimensions: quantity, quality, and metacognitive awareness. In our framework, quantity refers to language users' breadth of vocabulary, quality refers to their patterns of lexical organization, and metacognitive awareness relates to their ability to monitor how much they know of the microlevel characteristics of words when they say that they know a word. It should be noted here that the receptive and productive domains as well as the microlevel and the macrolevel are not viewed as independent of each other. On the contrary, they are viewed as constantly interacting, exchanging information, and informing each other; therefore, the proposed distinctions should be seen as serving the single purpose of distinguishing between what assessment of knowledge of individual words entails at the microlevel and how it differs from assessment of the overall lexicon at the macrolevel, in receptive and productive terms.

Although the role of quantity and quality of word knowledge has been extensively discussed in the L2 literature, little mention has been made about the place of metacognitive awareness in a framework of L2 learners' overall

https://doi.org/10.1017/S0272263105050254 Published online by Cambridge University Press

state of lexical competence. The psychological and psycholinguistic literature abound with terms such as *language awareness*, *control*, *metalinguistic awareness*, *explicit* or *learned knowledge*, and so forth, which overlap to such an extent that the subtle differences between them are almost impossible to disentangle (Ellis, 2004). Yet, most researchers agree that there is a difference between what one knows and the extent to which one is consciously aware and in control of what he or she knows about a phenomenon. Bialystok and Sharwood Smith (1985), for example, drew a distinction between knowledge and control of vocabulary. Knowledge was defined as "the way in which the language system is represented in the mind of the learner," whereas control was defined as "the processing system for controlling that system during actual performance" (p. 104). Gass (1988) rightly pointed out that this distinction is very useful when applied to vocabulary "since it crosses the boundaries of more traditional notions of productive and receptive knowledge. Both production and reception include information regarding knowledge and control" (p. 95). We use the term *metacognitive awareness* as a more general term that refers to "one's knowledge concerning one's own cognitive processes and products or anything related to them" (Flavell, 1976, p. 232). In our framework, this dimension performs a bridging function between the microlevel and the macrolevel of knowledge; that is, it allows language users to monitor what aspects of words they know when they say they know them.

The present study empirically tests the sensitivity of several measures commonly associated with the proposed macrolevel dimensions of lexical knowledge (i.e., quantity, quality, and metacognitive awareness) to capture differences in the lexical knowledge of L2 learners at different levels of language proficiency. The study aims at determining which macrolevel features vary as a function of L2 proficiency and at what proficiency level L2 learners' knowledge of words starts to resemble NSs' macrolevel of lexical competence. Several prior studies have found that the number of words learners know is linked to their level of proficiency (e.g., Meara & Jones, 1988; Read, 2000). However, vocabulary size accounts for only one dimension in the proposed framework of lexical knowledge. Vocabulary size by itself might say little about L2 learners' patterns of lexical organization or their ability to monitor how much they know about lexical items in comparison to NSs of a language. Thus, the present study aims at investigating how L2 learners' breadth, quality, and metacognitive awareness differ from that of NSs as a function of their increased language proficiency.

METHOD

Participants

Sixty-four NSs and NNSs of English, representative of a sample of normal educated adults, participated in the study. The NSs of English ($n = 30$) were under-

https://doi.org/10.1017/S0272263105050254 Published online by Cambridge University Press

graduate students at a North American university, enrolled in an introductory course in linguistics, both male (n = 9) and female (n = 21). They also rated their knowledge of a L2 on a 5-point scale (0 = no knowledge, 1 = beginner, 2 = lower intermediate, 3 = intermediate, 4 = upper intermediate, 5 = advanced). The mean ratings ranged from 0 to 1.65, which showed a low level of knowledge of another language that would not have influenced their performance in the experiment.

The NNS group consisted of Bulgarian L2 learners of English at two levels of language proficiency: advanced (n = 17) and intermediate (n = 17). At the time of the experiment, these learners were enrolled in advanced or intermediate certificate preparation courses in English in a school for foreign languages in Bulgaria. Both genders—males (n = 14) and females (n = 20)—were represented in the sample, and the participants were age 18 and older. All NNS participants were Bulgarian learners of English who had completed their high school education in Bulgaria and had received formal instruction in English.

The placement of students at these proficiency levels was determined on the basis of two tests that the school regularly uses to assess the progress of the students attending their certificate preparation classes, the Cambridge Certificate in Advanced English (CAE) test (for the advanced learners) and the Cambridge First Certificate of English (FCE) test (for the intermediate learners). Additionally, seven participants were placed in the L2 advanced group on the basis of their official TOEFL scores. Five participants had a TOEFL score over 560 (M = 581; range: 561–620) on a pencil-and-paper test format, and two participants had a mean score of 259 on a computer-based test format. All tests for determining language proficiency were taken in controlled settings.

Only the scores on the reading and use of English sections of the CAE and FCE were taken into consideration, with a lowest passing grade of D (between 50% and 60% correct answers) on both sections cumulatively. For the advanced group, the mean on the CAE reading section was 70.1% (range: 62–85%), and the mean on the CAE use of English section was 62.7% (range: 52–80%). The results of the intermediate group were as follows: The FCE reading mean was 72.5% (range: 66–82%), and the FCE use of English mean was 56.2% (range: 44–67%). In no case did the scores of any intermediate student on those assessments indicate that he or she should have been placed in the advanced class or vice versa.

Procedure

The three groups completed the vocabulary knowledge assessment used in the experiment in writing as a take-home test. The instructors of the three groups took special care to explain how the test should be completed by going over the instructions in class and giving several examples of how items should be approached. They also made sure that the participants understood that

https://doi.org/10.1017/S0272263105050254 Published online by Cambridge University Press

they were expected to complete the test honestly, without using a dictionary or any other reference material.

Materials

Word Sampling Procedure. The selection of stimulus words (SWs) was carried out with a spaced sampling procedure frequently used in L1 (e.g., Anglin, 1993; Johnson & Anglin, 1995) and L2 vocabulary studies (e.g., Goulden, Nation, & Read, 1990). The procedure involves selecting items from a randomly determined starting point in a dictionary and subsequently selecting words at a specific interval. The words were selected from *Oxford Student's Dictionary of Current English* (OSDOCE) (Hornby, 1978), which contains 35,000 entries. The dictionary is designed to be used by learners of English as a L2 and contains both British English and American English usage. Although concise, it is comprehensive and includes a wide range of vocabulary usage—formal, informal, slang, poetic, technical, foreign, and so forth.

Word Count Exclusion Criteria. As Nagy and Anderson (1984) convincingly argued, the sampling and evaluation of tests that aim at measuring vocabulary size, among other variables, largely depends on the answers to three questions: (a) What words are included or excluded from the count? (b) What words are grouped together and what words are treated as separate entries? (c) What is the size of the dictionary from which the sample is obtained? With these fundamental questions in mind, an independent count of the dictionary entries was performed based on specific criteria for exclusion of entries from the count (see Goulden et al., 1990). The following entries were excluded from the dictionary count as well as from testing when listed as separate boldface main entries: (a) letters of the alphabet and names of letters, (b) affixes, (c) capitalized proper words and their derivations, (d) abbreviations, (e) contracted forms, (f) interjections, (g) words with an alternate spelling, and (h) regularly inflected forms of words.

An examination of the organization of the OSDOCE information showed that the macrostructure contained lemmatized headwords, letters, affixes, proper words, abbreviations, cross-references of contracted and irregular forms, interjections, and multiword entries. Homonymy and polysemy were treated in a consistent manner: Homonyms were granted separate headword status, whereas definitions of polysemous entries were listed under the headword entry. The irregularly inflected forms were given separate entries in the dictionary, which is very much in line with experimental evidence that supports the notion of obscured morphemes (e.g., suppletive forms like *go* and *went*) having a separate representation in the mental lexicon (Sandra, 1994, p. 238). In the count, we included lemmatized headwords that were granted a headword status as well as boldface run-on entries, including hyphenated multiword entries. As for the derivational affixes, the practice observed in OSDOCE seemed to be based on the principle that if a derivative had a mean-

https://doi.org/10.1017/S0272263105050254 Published online by Cambridge University Press

ing that is not covered by the senses or the definitions listed under the main entry to which it is appended, it should be entered and defined as a separate entry. Therefore, we decided to include in the word count boldface run-on entries derived by means of derivational affixation or compounding because their formation is not paradigmatically determined and, as such, it cannot be easily predicted in production (Sandra).

In summary, the notion of a word adopted in this study was taken to include: (a) a word (including its other variants; e.g., spelling, abbreviations) and (b) its regularly inflected forms. Regularly realized inflectional suffixes were assumed to produce mere paradigmatic variants of the base, and psycholinguistic research in modeling the mental lexicon has convincingly argued that inflectional suffixes should be treated outside the lexical domain because they are linked to a particular class form (e.g., verbs, nouns, adjectives) and produce different realizations of a single word rather than different words (Sandra, 1994, p. 231). Thus, following the suggested specifications, knowing the noun *motor* (meaning the word family of the headword), for example, will include knowing its inflected forms (*motors* and *motor's*) but will not include, for example, *motorcade*, *motorway*, *motorist*, *motorize*, *motoric*, and others listed under the main entry.

All words in the OSDOCE were counted, according to the criteria just discussed by one of the researchers. The total number of words that occurred as boldface new main entries in the dictionary was found to be 23,996. Then, 100 pages of the 762-page dictionary were randomly sampled and a second researcher used the established criteria to make an independent count of the words on these pages. Of the 3,131 words that occurred on the 100 randomly selected pages, there were only 25 disagreements with the original count (less than 0.8%), which means that there was a high degree of agreement (99.2%) between the two independent counts.

Sampling of Words. The test items or SWs were randomly selected from the dictionary by using a spaced sampling procedure once the dictionary sampling procedure had been completed. The sampling interval took every first new boldface main entry from every 20th column, starting from the right-hand column on the first page of the dictionary, regardless of whether the entry was a homograph or not. A selected item was listed with its syntactic category in the test if it occurred as a subsequent homograph in the dictionary. Thus, the resulting list of 75 words in the test sample became representative of the 23,996-word content of the dictionary in terms of word frequency, lexical categories, and morphological word types.

In summary, the approach taken to the dictionary count—as well as to the sampling of the test items—was to provide a meaningful operational definition of a word that would describe the lexicon of educated adult NNSs of English in terms that are close to the ones most educated people would apply when referring to the content of their vocabularies. The test sample covered a broad range of lexical usage, including specialized vocabulary, foreign words,

https://doi.org/10.1017/S0272263105050254 Published online by Cambridge University Press

slang, formal and informal usage, old and current usage, as well as poetic and technical words from a wide range of word frequencies.

Description of the Test Items. The selected SWs were listed alphabetically in the test in the order they were selected from the dictionary. Our sample pretesting results identified two items as outliers (*nonagenarian* and *stern* [v.]), whose frequencies were also not found in *The Teacher's Word Book of 30,000 Words* (Thorndike & Lorge, 1944) or in *The Educator's Word Frequency Guide* (Zeno, Ivens, Millard, & Duvvuri, 1995). This suggested that the items were most probably not responded to in the pretesting trial because of their very low frequency of occurrence, and they were subsequently excluded from testing. As such, the final version of the test contained 73 items as SWs (see Appendix A) of different syntactic categories (see Table 1) and frequency of occurrence.

A problem—briefly mentioned previously—that stems from using spaced sampling for vocabulary testing concerns the representation of different frequencies in the test sample. In other words, it has been suggested in the literature that a direct consequence of the application of this procedure is the likelihood of high-frequency words to be overrepresented in the sample because they tend to occupy more dictionary space than the lower-frequency words (e.g., Goulden et al., 1990; Lorge & Chall, 1963; Nation, 1993). To ensure that several frequency bands were well represented in the test sample, we examined the frequencies of the selected SWs in *The Educator's Word Frequency Guide* (Zeno et al., 1995), which is based on over 60,000 text samples (over 17 million words) from a wide range of texts that students in the United States are likely to encounter throughout their school and college years. The

Table 1. Syntactic categories and frequency of occurrence of the SWs used in the study

SWs	No. of items
Syntactic categories	
Nouns	42
Verbs	16
Adjectives	13
Adverbs	2
SFI	
20.0+	13
30.0+	23
40.0+	19
50.0+	11
60.0+	4
70.0+	3

https://doi.org/10.1017/S0272263105050254 Published online by Cambridge University Press

frequency of each word included in the test was identified by its Standard Frequency Index (SFI), which is a logarithmic transformation of the *u*-value (the frequency of a word per million tokens in a corpus of an infinite size; Zeno et al.). The decision to use the SFI was motivated by the fact that it provides a more compressed range of values than the *u*-value while completely retaining all its advantages. Overall, based on the frequency of distribution of the test items, it can be said that low-, mid-, and high-frequency words were fairly evenly distributed in the sample and that the sample was not biased toward any frequency band.

Description of the Test Format. The test format required that the participants in the experiment demonstrate in a verifiable way that they knew what each of the words meant. To this end, we used a modified version of the Vocabulary Knowledge Scale (VKS), introduced by Paribakht and Wesche (1993), in a L2 context. The instrument used in the experiment consisted of two scales. The first, which showed a degree of participants' familiarity with the lexical items in a verifiable way, followed the categories (I through IV) suggested by the designers of the original scale. The second scale was modified in accordance with the goals of the present study and required the participants to produce as many as three associations to the SWs in place of a sentence production in the original VKS. The two scales accompanied every SW and were presented simultaneously to the participants (see Appendix B).

Scoring Procedures. Six dependent variables frequently associated with the posited three dimensions of lexical competence were used to compare the macrolevel of participants' vocabulary knowledge. The scoring procedures used for quantifying each variable were as follows.

Self-perceived knowledge versus verified vocabulary knowledge. It was assumed that the difference between participants' self-perceived knowledge and their actual knowledge of the SWs would reflect their degree of metacognitive awareness (i.e., how much they actually know when they claimed that they know a word). Therefore, each option on the VKS was assigned a numerical value (i.e., option I = 1 point, option II = 2 points, option III = 3 points, option IV = 4 points) and, depending on how a participant self-reported and demonstrated his or her familiarity with the SWs, a corresponding score was assigned to each stimulus. To capture how participants' self-perceived familiarity corresponded to their actual knowledge, a scoring scale similar to the one proposed by Wesche and Paribakht (1996) was used. Options I and II yielded a score of 1 and 2, respectively. Option III could lead to a maximum score of 3, if a participant had provided an acceptable brief definition, synonym, or translation (for the NNSs) of a SW, or 2, if a participant claimed some familiarity, but the response had shown that he or she did not recognize it correctly. Option IV yielded a maximum of 4 points, if knowledge of a SW was demonstrated as required, or 2 points, if the response revealed some sort of a meaning misinterpretation of the SW or did not reflect its lexical category.

https://doi.org/10.1017/S0272263105050254 Published online by Cambridge University Press

For a synonym or a translation of a SW to be considered a correct response, it had to be semantically and syntactically relevant to the SW. When a brief definition or explanation was provided as a response in option III or IV, the decision about its acceptability was based on the criteria used in lexicography for defining words by part of speech, as outlined by Landau (2001). For instance, the explanation or definition of a noun should immediately answer the question "What is it?" and in order to answer that question, it should contain a noun (quantified or not) in the first part of the definition, which shows the part of speech. The definition of adjectives was expected to contain some of the introductory words and phrases used for defining adjectives, such as *able to, being, belonging to, full of, having the quality of, pertaining to*, and so forth. The definition of verbs was anticipated to begin with another verb, with or without the infinitival particle "to." Adverbs were considered correctly defined if another adverb was included in the definition or if the definition contained a phrase with the meaning "in the manner of." The requirement that a response should adequately reflect the semantic and the syntactic properties of the SWs was also applied to the translations.

Vocabulary size. Estimates of vocabulary size were derived by multiplying the number of known words (obtained from the correct responses to option III or IV in the test) by the number of words counted in the dictionary (23,996), divided by the number of SWs used in the test (73).

Word frequency effects. The SWs in the test represented six word frequency bands, which were defined by reference to Zeno et al. (1995). The assumption behind considering word frequency as a factor that influences knowledge of words was that vocabulary knowledge is cumulative across several frequency bands. To examine the relationship between word frequency and vocabulary knowledge, a numerical value from 1 to 6 was assigned to each SW that the participants knew; the lower the frequency of occurrence of a word, the higher the numerical value. As a result, responses supported by demonstrated knowledge of a SW with a SFI = 70+ (e.g., *back* [v.]) were assigned 1 point, responses to a SW with SFI = 60+ (e.g., *hard* [adv.]) were awarded 2 points, responses to a SW with SFI = 50+ (e.g., *diamond* [n.]) were given 3 points, and so forth. A major consideration in the quantification of the responses was to give more credit to participants who knew words from the lower frequencies in addition to the high frequency words.

Nativelike commonality of associative responses. The procedure frequently used for quantifying NNSs' associative responses (e.g., Kruse, Pankhurst, & Sharwood Smith, 1987; Schmitt, 1998) is to match them to a norm list of associations elicited from a group of NSs. The elicitation process from the control group resulted in a total of 2,084 different associations, with a few illegible responses that were not considered in the count. The norm responses were lemmatized and tallied on a list, and the following items were listed as separate associations: (a) base words and some of the inflections were lemmatized and combined as one item—for example, inflectional *-s* for nouns (e.g., *car* and *cars*), inflectional *-s* for verbs (e.g., *forget* and *forgets*), inflectional *-er*

https://doi.org/10.1017/S0272263105050254 Published online by Cambridge University Press

and *-est* for adjectives (e.g., *tall–taller–tallest*) and adverbs (e.g., *soon–sooner*); (b) multiword responses (e.g., *auspicious situations, good opportunities, marching line, give up, spread out*) were listed and scored as one item based on the head of the phrase; (c) all derivations were treated as separate items (e.g., *cleaning, cleaned, breathless, immoral, resignation, disagreement, scrutinize, unfamiliar*)[1]; and (d) irregularly inflected forms (e.g., *children, better, went*) were treated as separate forms.

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

As a result, a list of all associative responses to each SW that was shown to be familiar to the NS participants was obtained along with a tally of how often each response was given. Then, each NS and NNS response was assigned a numerical value of commonality dependent on its frequency of occurrence in the NS data. For example, for the SW *advantageous*, the top three responses were *helpful* ($n = 7$), *beneficial* ($n = 4$), and *good* ($n = 4$). As such, the best performance for either a NS or NNS would be to produce the three most frequently given associations, which would translate into a score of 15 points ($7 + 4 + 4 = 15$).

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

Number of associations. A numerical value reflecting the absolute number of associations to any SW by a participant was assigned to every SW. In other words, given that familiarity with a SW was demonstrated in an acceptable way, each familiar SW could carry a maximum score of 3 because the participants were prompted to give as many as three associations to the words they could verifiably identify as known. Then, the total number of generated associations by a participant was considered to reflect the size of his or her associative domain.

Reliability. Because the SWs were not selected from commonly used word lists such as Nation's (1990) Vocabulary Levels Tests, the University Word List

https://doi.org/10.1017/S0272263105050254 Published online by Cambridge University Press

(Xue & Nation, 1984), or the General Service List (West, 1953), which are organized around word frequency counts (e.g., Thorndike & Lorge, 1944), it was of paramount importance to assess the internal consistency reliability of the test as a measurement instrument. Reliability for internal consistency of the test was calculated for each group by using the Kuder-Richardson 21 (K-R 21) formula. The results were as follows: For the NS group, a value of K-R 21 = .88 was obtained; for the group of advanced learners of English, K-R 21 = .80, and for the L2 intermediate group, K-R 21 = .85. These results were comparable with the internal consistency reliabilities of other instruments for assessment of lexical knowledge (e.g., forms A and B from the University Word Level Test; Xue & Nation) and in the acceptable range (Popham, 1990; Vierra & Pollock, 1992) for the purposes of this study.

RESULTS

A series of one-way analyses of variance (ANOVAs) was conducted to evaluate the relationship between language proficiency and the variables associated with lexical knowledge of educated adults. The between-subjects variable (language proficiency) had three levels: NSs of English, L2 advanced learners, and L2 intermediate learners. The dependent variables used in the experiment were (a) vocabulary size, (b) word frequency effects (or knowledge of words from different frequency bands)—both variables commonly associated with breadth of vocabulary, (c) difference between verified and self-perceived knowledge of vocabulary, which in our framework is a measure linked to metacognitive awareness, and, finally, (d) nativelike commonality of associations, (e) number of associations, and (f) within-group consistency of the WA domain—three variables frequently used to describe the quality of language users' meaning connections with respect to the commonality and size of their WA domain. With three relatively small groups of participants, power to detect Cohen's (1988) large effect size ($d = .8$) at $\alpha = .5$ exceeded .80, which left us with a probability of a Type II error within acceptable values. Means and standard deviations are presented in Table 2.

The analyses yielded a significant group effect on several dependent variables, indicating statistically significant mean differences among the NS, L2 advanced, and intermediate learners on vocabulary size, $F(2, 61) = 13.22$, $p < .001$, $\omega^2 = .28$, word frequency, $F(2, 61) = 10.91$, $p < .001$, $\omega^2 = .25$, number of associations, $F(2, 61) = 14.86$, $p < .001$, $\omega^2 = .32$, nativelike commonality of associations, $F(2, 61) = 61.75$, $p < .001$, $\omega^2 = .66$, and within-group consistency of the WA domain, $F(2, 61) = 18.10$, $p < .001$, $\omega^2 = .35$. The difference between participants' verified and self-perceived vocabulary knowledge was found to be statistically nonsignificant, $F(2, 61) = 1.47$, $p = .24$, $\omega^2 = .01$. The calculated practical significance of the measured relationships (ω^2) revealed that the dependent variables provided practically meaningful measures of the overall state of participants' lexicons when proficiency was used as a grouping variable.

https://doi.org/10.1017/S0272263105050254 Published online by Cambridge University Press

https://doi.org/10.1017/S0272263105050254 Published online by Cambridge University Press

Table 2. Group means and standard deviations for responses given by NSs and L2 advanced and intermediate learners of English for the variables used in the experiment

	NS (n = 30)		Advanced (n = 17)		Intermediate (n = 17)	
Measures	*M*	*SD*	*M*	*SD*	*M*	*SD*
Vocabulary size	14,320	3,114	13,245	4,164	9,134	2,887
Word frequency	170.73	43.92	156.88	56.70	105.71	38.91
Difference between verified and self-perceived vocabulary knowledge	9.37	8.40	12.41	8.23	8.06	5.44
Nativelike commonality of associations	369.20	85.10	198.29	88.53	103.12	70.01
Number of associations	111.57	26.20	120.65	35.38	65.47	39.03
Within-group consistency of WA domain	369.20	85.10	307.12	76.77	202.55	112.83

Overall, the results of the analyses support the conclusion that the NSs', L2 advanced learners', and intermediate learners' lexicons are different on two of the dimensions (quality and quantity) but showed nonsignificant differences in the degree of their metacognitive awareness. To further explore the nature of the differences among the participants' lexical knowledge and their relationship to language proficiency, post hoc pairwise comparisons were conducted to compare the three groups on all measures used in the study. The post hoc comparisons (in which the significance level was adjusted to .05 based on Bonferroni rationale) revealed that there were statistically significant differences between the intermediate learners and the NSs in all subsets of comparisons except the metacognitive measure—that is, vocabulary size, 95% confidence interval (CI) = (2669.79, 7703.47); word frequency effects, 95% CI = (30.08, 99.97); nativelike commonality of associations, 95% CI = (204.53, 327.64); number of associations, 95% CI = (21.83, 70.36); and within-group consistency of associative responses, 95% CI = (98.45, 234.85). Similar significant differences were detected between the intermediate and the advanced learners, again with the exception of the metacognitive measure: vocabulary size, 95% CI = (1267.37, 6954.75); word frequency effects, 95% CI = (12.28, 91.25); nativelike commonality of associations, 95% CI = (25.52, 164.73); number of associations, 95% CI = (27.76, 82.59); and within-group consistency of associative responses, 95% CI = (27.51, 181.63). Interestingly, the intermediate learners showed as heightened a metacognitive awareness as the NSs, 95% CI = (−4.44, 7.05), and the advanced learners, 95% CI = (−2.14, 10.84). Not surprisingly, no statistically significant differences were found between the advanced learners' and the NSs' breadth of vocabulary: vocabulary size, 95% CI = (−1441.26, 3592.41); word frequency effects, 95% CI = (−21.68, 48.21); metacognitive awareness, 95% CI = (−8.79, 2.70); or between advanced NNSs' and NSs' quality of vocabulary knowledge: number of associations, 95% CI = (−33.35, 15.19); and within-group consistency of associative responses, 95% CI = (−9.12, 130.28). However, the difference between NSs' and advanced learners' degree of nativelike commonality of associations was found to be statistically significant, 95% CI = (109.35, 232.46), which showed that even at an advanced level of language proficiency, the L2 learners have not reached the extent of associative commonality characteristic of NSs.[2]

On the other hand, the fact that the advanced learners were not statistically different from the NSs on any of the other variables except nativelike commonality of associations was intriguing because it directly challenged the usefulness of WA tests as a measure of degree of shared nativelikeness of associations on which conclusions about the lexical knowledge of NNSs can be based. Moreover, when each group's associative commonality was compared by using the within-group consistency of the WA domain as a measure, it was found that there were no statistically significant differences between the advanced learners and the NSs. This suggests that, just as NSs tend to show a high degree of commonality of associative connections, L2 learners at an

https://doi.org/10.1017/S0272263105050254 Published online by Cambridge University Press

advanced level of proficiency also show stable patterns of commonality that do not resemble those of NSs.

It has become axiomatic in the L2 research field that "most nonnative speakers never reach a point of being indistinguishable from native speakers of a particular community" (Gass, 1990, p. 37), although it is questionable whether this is the ultimate goal of L2 learning. Nonetheless, the conclusion itself presupposes detectable differences between the two broadly defined groups, especially when the measuring stick is nativelikeness of commonality of associations. Therefore, we find it more meaningful to study whether, for example, NNSs' strength and diversity of relationships among words in their mental store compare to those of NSs instead of exhausting a research area that does not hold a potential for novel findings. We will further explore this idea in a subsequent section.

GENERAL DISCUSSION

The research questions examined in this study concerned three dimensions of L2 vocabulary knowledge, posited to account for its macrostructure and the sensitivity of measures associated with each of the three dimensions in distinguishing between levels of L2 proficiency. Previous research has focused on studying one or two dimensions of the lexical competence of L2 learners, either not paying much attention to the relative proficiency of the subjects or focusing on the characteristics of the words themselves (e.g., frequency, lexical category). The current study examined three groups of participants: A NS group served as a control against which the performance of L2 advanced and intermediate learners was compared. The proficiency level of the L2 learners was established independently on the basis of their performance on three proficiency tests (CAE, FCE, and TOEFL).

It was argued that quantity or breadth of vocabulary, quality or connectivity among lexical items in language users' mental lexicon, and learners' metacognitive awareness were the dimensions that capture the overall state of L2 learners' lexicons. Participants' breadth of vocabulary was evaluated by obtaining a measure of their knowledge of words from different frequency bands and by estimating their vocabulary size. The quality of their lexicons was established by examining several features of their meaning connections, such as the size of their WA domain (determined by the number of responses generated to the SWs) and its consistency (determined by the degree of nativelike and within-group associative commonality of responses). Finally, it is logical to assume that when we are dealing with adult learners' lexical competence, we are also dealing with their metacognitive awareness, which can be measured by the extent to which their consciously self-perceived degree of familiarity with lexical items reflected their actual knowledge of words, which might be dependent on their level of language proficiency. Thus, the proposed dimen-

https://doi.org/10.1017/S0272263105050254 Published online by Cambridge University Press

sions were measured by six variables that have frequently been used by L2 researchers to explore the lexical competence of L2 learners in an attempt to investigate their relationship to the language proficiency of the participants in the study. It was also of interest to the researchers to find out what variables were sensitive indicators of growth in L2 learners' proficiency and what dimensions of learners' lexical competence were subject to development as a result of an increase in proficiency. In this section, we will focus on the way the results of the experiment support a three-dimensional framework for studying the overall state of L2 learners' lexical competence. To this end, the major findings of the experiment will be discussed within the proposed framework and with respect to the research questions addressed in the study.

Breadth of Vocabulary Knowledge

We first asked how the participants' breadth of lexical knowledge changed as a result of an increase in language proficiency. Breadth or quantity of word knowledge is probably the best researched dimension in L2 lexical research. The quantity of participants' vocabulary was examined by measuring their vocabulary size and their knowledge of words from a range of several frequency bands. To estimate the vocabulary size of the participants, we selected the test items from a dictionary that (based on our specific research criteria) was found to contain approximately 24,000 words; this number served as a baseline lexicon for estimation of participants' vocabulary size.

Although providing an estimate of NSs' vocabulary size was beyond the focus of this study, a brief mention needs to be made about how the results obtained in the experiment compare with other studies that have specifically focused on determining the number of words NSs have in their lexicons. Such a comparison is necessary because it will show whether the estimated size of the NSs' lexicon in the present study provides a reasonable baseline against which NNS' quantity of word knowledge can be matched. By and large, estimates of how many words NSs have in their lexical repertoire abound, ranging from unbelievably high (e.g., 216,000 words; Diller, 1978) to more conservative current estimates of 14,000–20,000 words (e.g., D'Anna, Zechmeister, & Hall, 1991; Goulden et al., 1990; Nusbaum, Pisoni, & Davis, 1984; Zechmeister, Chronis, Cull, D'Anna, & Healy, 1995) or 17,000-word families (e.g., Meara, 1996).

Our results were consistent with previous research on this issue and indicated that—based on a 24,000-word baseline lexicon—educated adult NSs of English might be estimated to know on average about 14,320 words (range: 8,547–20,709 words). As a proportion, this number showed that the NS participants in this study knew, on average, 59% of the estimated total size of the dictionary from which the SWs were selected. It should be mentioned here that we used conservative criteria for what counted as knowing a word: (a) The

https://doi.org/10.1017/S0272263105050254 Published online by Cambridge University Press

participants had to provide verified responses; (b) a response was considered acceptable only if it accounted for the lexical category of the SW in addition to at least one of its senses; and (c) partial responses were not counted in the number of words rated as familiar by the participants. Given these stringent criteria, the proportion of words known by the NSs to the estimated size of the dictionary compared very well with L1 studies that also used abridged dictionaries (e.g., D'Anna et al., 1991; Zechmeister et al., 1995). For example, based on a dictionary count of 26,091 words and a 5-point scale of rated SW familiarity, D'Anna et al. found that college students knew the meaning of 16,785 words, which was 62% of their dictionary count. Several years later, using the same dictionary, Zechmeister et al. concluded that although the size obtained earlier was relatively small compared to other estimates in the literature, it likely overestimated the students' vocabulary size because the test-takers simply rated their knowledge on a scale from 1 to 5 without being required to verify it. When asked to select definitions for the words in a multiple-choice format, the researchers estimated that undergraduate college students knew 15,872 words on average, which was 59% of the overall dictionary count. On the whole, the estimated vocabulary size of the NSs in our study is in line with estimates derived by other researchers (e.g., Zechmeister et al.), who specifically focused on measuring the size of vocabulary of educated adult NSs of English. Additionally, when the means obtained in this experiment were compared with the average sizes across other studies, the results revealed that they were compatible, although nonsignificantly smaller. Thus, we concluded that the experiment provided a meaningful estimate of NSs' vocabulary size against which NNSs' vocabulary could be matched.

Although L2 learners' and NSs' lexicons have been found to share certain similarities, some of the differences are often attributed to L2 learners' generally smaller vocabulary size, which is often the reason for their problems in successfully handling L2 tasks. Our experiment had the advantage of involving L2 learners at two proficiency levels, which allowed for finer generalizations concerning the quantity of their vocabulary knowledge. The average lexicon size of the advanced learners was estimated at 13,245 words (range: 7,889–20,709 words), which indicated that the advanced learners could recognize and define or translate approximately as many words as the NSs could. However, the average intermediate student's vocabulary was found to contain 9,134 words (range: 4,931–14,792 words) and was significantly smaller than that of both the NSs and the advanced learners. The same differences among the three groups were detected when their knowledge of words from different frequency bands was compared, which confirmed that variation in the breadth of vocabulary knowledge closely reflects the overall proficiency of L2 learners. On a final note, we should say that the conclusions regarding the size of the participants' lexicon are strictly limited to recognition of words in isolation, which renders the lexical task discrete in nature. However, as far as productive vocabulary or performance on communicative tasks is concerned, the

https://doi.org/10.1017/S0272263105050254 Published online by Cambridge University Press

estimated quantity of lexical repertoire might not be the same. The best prediction we can make, based on Laufer (1997) and the derived vocabulary size approximations, is that intermediate learners should be expected to have reached the lexical threshold that, for example, would allow them to successfully transfer L1 reading strategies onto L2 reading tasks. Laufer also suggested that knowledge of 6,000-word families (9,600 lexical items) is likely to result in 77% text comprehension, which was to a great extent supported by the reading scores (M = 72.5%; see description of participants) of the intermediate participants in this study. However, in terms of text coverage, especially for academic purposes, the proposed figure of 3,000-word families (about 5,000 lexical items) as sufficient for comprehension of 95% text coverage (e.g., Laufer, 1992; Nation, 1990) might need revision.

Quality of Lexical Knowledge

Quality of lexical knowledge was the second dimension of lexical competence proposed in the framework. Various attempts have been made to explore the idea of measuring the depth of lexical knowledge. Many lexical researchers (e.g., Greidanus & Nienhuis, 2001; Kruse et al., 1987; Meara, 1978, 1996; Nation, 1993; Read, 1993; Schmitt, 1999; Wesche & Paribakht, 1996; Wilks & Meara, 2002; Wolter, 2001, 2002) seem to agree that using WA tests to probe L2 learners' depth of vocabulary knowledge can be a valuable instrument for collecting data that are revealing of the way language users organize their mental lexicon. However, many of these attempts have been devoted to studying the extent to which NNSs' associative behavior resembles the degree of commonality of associations that NSs maintain. This line of research has been largely promoted by L1 lexical studies that have repeatedly found that NSs tend to cluster their responses to any SW around a small number of commonly given associations. Consequently, this high degree of associative commonality has come to be interpreted as indicative of the stability of semantic connections that NSs build among the words in their mental store, which, in turn, reflects the way they organize their meaning networks. Similarly, some L2 researchers (e.g., Kruse et al., 1987; Schmitt, 1999; Wolter, 2002) adopted the approach of comparing the degree of NNSs' associative commonality to that of NSs' associative behavior in an attempt to account for the effects of language proficiency on the semantic stability of the L2 organization of the mental lexicon. This line of research was further encouraged by findings that suggested that L2 learners tend to have much more diverse (interpreted as unstable) patterns of responses, although, as their proficiency increases, some aspects of their associative connections develop toward the NS norms (e.g., Meara, 1978; Vermeer, 2001). Consequently, several relatively distinct lines of L2 WA research developed, which differ primarily in the perspective from which L2 association patterns are evaluated: from a sociocultural perspective, from a language

https://doi.org/10.1017/S0272263105050254 Published online by Cambridge University Press

proficiency point of view, as an indicator of depth of vocabulary knowledge, and so forth. Unfortunately, the relationship between these lines of research and the implications they have for each other has never been commented upon in the L2 literature. The present study offers such an opportunity, which we will explore in the context of the discussed experiment.

In this study, we used three variables associated with the analysis of the quantitative features of language users' WA domain: nativelike commonality of responses, number of associations, and within-group consistency of the WA domain. The rationale behind using three variables was to find which ones would account for the effects of language proficiency on the quality of participants' lexicon organization as well as the stability of connections that they build among the words they know. The comparison among the three groups of participants on their shared nativelike commonality of associative responses revealed that neither the advanced nor the intermediate learners came close to the relatively high degree of associative commonality that the NSs maintained. Similar findings are traditionally interpreted in the L2 WA literature to mean that L2 learners produce diverse and unstable patterns of associations, which implies a lesser degree of connectivity among the words they know, interpreted as an indication of a lower quality of their meaning connections and lexicon organization. However, a closer look at the association data as well as at the sociocultural WA research prompted us to challenge the usefulness of this approach in the analysis of WAs, particularly when used to draw conclusions about the quality of L2 learners' organization of their lexical knowledge. To begin with, positing that the quality of L2 lexical organization depends on the degree of shared nativelike commonality of associative meaning is, in effect, positing that semantic organization is entirely and solely language based.

Some recent WA studies from the sociocultural line of research have reached a different conclusion. Yoshida (1990) noticed that Japanese college students tend to produce language-dependent associative responses to SWs from the culture category but not as much to culturally neutral stimuli. Other researchers from the cultural paradigm (e.g., Szalay, 1984; Szalay & Brent, 1967; Szalay & Windle, 1968) further argued that the acquisition of cultural concepts is accompanied by cognitive restructuring of the conceptual system. Szalay and Windle and Szalay, Lynse, and Bryson (1972), for example, uncovered differences in the response patterns of Korean participants that were traceable to the influence of the culture in which the L2 was acquired and not to the language of response itself. Overall, the conclusions drawn by the researchers working with WA data in the sociocultural paradigm convincingly show that meaning organization is not only linguistically driven, but also strongly influenced by a host of extralinguistic factors. This general conclusion was supported by numerous examples in our data, which showed that using nativelike commonality to measure quality of L2 learners' knowledge is an approach that is sensitive only to the broad distinction between NSs and NNSs of English and that it does not capture differences related to

https://doi.org/10.1017/S0272263105050254 Published online by Cambridge University Press

proficiency or learners-related influences. The results of our analysis suggested that even at an advanced level of proficiency, L2 learners do not—and probably cannot—reach the extent of commonality characteristic of NSs, which does not necessarily mean that their degree of lexical connectivity is of a lesser quality than that of NSs. This was further supported by the comparison of the average number of associations the participants generated to the SWs used in the study. The analysis revealed that, on average, the advanced learners could generate as many associations (M = 121) as the NSs (M = 112), whereas the intermediate learners were able to produce a significantly smaller number of responses (M = 65). In light of the well-established positive relationship between vocabulary size and size of the WA domain, one possible conclusion is that L2 vocabularies of approximately 6,000 words—or intermediate learners' vocabularies—are characterized by relatively few connections among words. Larger vocabularies—containing 8,700 words or more—show greater connectivity and allow language users to associate every lexical item with a much greater number of things and ideas. In other words, language users with larger vocabularies have considerably richer connections, both in size and heterogeneity.

Finally, a closer look at the association data suggested that, just as most NSs tended to maintain common word association networks, the L2 learners also had stable patterns of commonality of associative responses that did not, however, resemble those of NSs. For example, the three associations most commonly given by NSs to the SW *ruler* were *school*, *inch(es)*, and *math*, whereas the three most common associations in the NNSs' WA data were *authority*, *monarch*, and *king*. It is apparent that the L2 learners' responses were as relevant to the SW as the NSs' responses; yet, they reflected experience with the stimulus in contexts that were not commonly shared by the NSs and NNSs. This and many other similar instances prompted us to compare the degree of each group's commonality of associative responses as a measure of how well a word was integrated in the lexicon of the NSs and the L2 learners. The assumption was that degree of commonality of associative behavior, rather than degree of nativelike commonality, could be a more sensitive index of language proficiency. This assumption was confirmed by the analysis of each group's commonality of associations. Not surprisingly, the advanced learners' consistency of WA responses was as strong as that of the NSs (although of different nature), whereas the intermediate learners' strength of the WA domain was noticeably less consistent than that of the NSs and advanced learners' semantic stability of associations. This finding was in full agreement with the comparison of participants' size of the associative domain (number of responses), which led to the conclusion that these two variables provide a more meaningful estimate of quality of lexical knowledge than nativelikeness of the L2 associations. More importantly, as an approach to the analysis of WA data, the two measures accounted equally well for the difference in the level of proficiency of the L2 learners as well as for the linguistic and the extralinguistic factors that influence associative relationships.

https://doi.org/10.1017/S0272263105050254 Published online by Cambridge University Press

Metacognitive Awareness

The third dimension of lexical competence that was proposed in the framework was metacognitive awareness: We were interested in determining whether or not metacognitive awareness of vocabulary knowledge was dependent on proficiency. It was argued that this dimension performs a bridging function between the microlevel and the macrolevel of lexical knowledge in that it accounts for language users' ability to constantly monitor how much they know about a word (as a phonetic or orthographic, phonological, morphological, syntactic, and semantic entry) when they say that they know it. Along these lines, Ellis (2004, p. 242) rightly pointed out that although knowledge of a word form and its collocations can generally be regarded as implicit, knowledge of word meaning is explicit in nature because it shows an individual's conscious awareness of what it means through their ability to define it. Although studies that specifically focus on examining L2 learners' metacognitive awareness of word features are scarce, a method that successfully elicits analyzed knowledge requires participants to define or explain the meaning of words either orally or in writing. The underlying assumption is that a definition or brief explanation reflects knowledge of the properties of a word over and above its meaning, which need not be expressed in metalanguage (Ellis). The VKS that we used in the study offered several advantages in this regard. For example, using a rating scale along with a requirement for verifiable responses allowed the participants to rate their knowledge of words on a scale that reflects partial as well as precise lexical knowledge. On the other hand, participants' verifiable short definitions, explanations, or synonyms and translations of the SWs allowed for a comparison of their triggered conscious analyzed knowledge of the words they think or claim they know with the actual properties of these words. Thus, the test format made it possible to draw a number of conclusions about participants' degree of metacognitive awareness regarding the overall state of their vocabulary, which a multiple-choice format, for example, does not allow.

In discussing a metacognitive framework for the development of language skills, Bialystok and Ryan (1985) suggested that it encompasses two components—analyzed knowledge and cognitive control—which represent two interrelated aspects of metalinguistic performance that should nonetheless be distinguished from each other. Their proposed framework has a specific relevance to the present study. In the experiment, the participants were required to exercise both analyzed knowledge, given that the SWs had to be recognized and responded to in isolation, as well as cognitive control, as they were required to coordinate several pieces of information about the properties of the SWs before responding to them. Therefore, it was assumed that the point of intersection between these two metalinguistic skills could be accounted for by examining the difference between participants' self-reported knowledge (in the form of a definition, brief explanation, or a synonym or translation) and their actual knowledge of the SW. Additionally, it was impor-

https://doi.org/10.1017/S0272263105050254 Published online by Cambridge University Press

tant to ascertain whether L2 learners' language proficiency would have specific effects on their ability to exercise adequate control and perform well on the lexical task, to the extent permitted by their analyzed knowledge. Interestingly, the results revealed that in spite of the different levels of language proficiency and differing quality and quantity of lexical knowledge, all participants were equally skilled at making relatively difficult metalinguistic decisions when responding to the SWs. Put differently, although the participants' vocabulary size, degree of knowledge of words from various frequency bands, and their connectivity of the lexicon were proficiency dependent, the difference between their actual and self-perceived knowledge of words was practically the same, with a slight tendency for overestimation most often attributable to a failure to account for the intended syntactic category of the SW.

Thus, on the basis of our analysis, we can confidently say that the three groups of participants showed a heightened degree of metacognitive awareness, regardless of their level of language proficiency. This can be explained to a large extent by their overall cognitive development (inasmuch as all participants were adults), which triggers heightened metalinguistic awareness. This finding is also in agreement with the hypothesis advanced by Bialystok and Ryan (1985), which states that if a metalinguistic skill is taken to be the control required to analyze a structured representation of a language, then adult L2 learners—who have already had an experience of acquiring one language—should find it relatively easy to solve metalinguistic problems in the new language. The only restriction suggested by the authors is the extent of the learners' analyzed knowledge. Apparently, it should be expected that adult L2 learners at an intermediate or higher level of language proficiency can exercise as heightened a degree of analyzed knowledge and cognitive control in the L2 as NSs can in their L1, especially when the task itself puts that specific demand on their performance. Yet, further research is needed to examine whether the same dependency will hold true for the lower-proficiency L2 learners.

Variable Sensitivity

The last research question asked what measures linked to the three dimensions were proficiency related and, as such, had the potential to account for the variance in the overall state of participants' lexical knowledge in advanced stages of language learning. The analysis showed that some measures were more sensitive to the L2 learners' proficiency than others. This means that some aspects, such as quality and quantity of the L2 lexical competence, develop as the proficiency of the L2 learners increases, whereas others, such as learners' metacognitive awareness, are not proficiency dependent. It was found that vocabulary size, knowledge of words from different frequency bands, number of associations, and within-group associative commonality were the

https://doi.org/10.1017/S0272263105050254 Published online by Cambridge University Press

macrolevel measures that can distinguish well between advanced and intermediate levels of proficiency. Therefore, we suggest that, for diagnostic purposes, these measures would be reliable indexes of proficiency development and sensitive to the overall state of L2 learners' vocabularies. In the context of the whole experiment, the analysis of the nativelike commonality of the WAs generated by the L2 learners proved to be unproductive because this variable, as in other studies, was not sensitive to an increase in their proficiency. In fact, it only distinguished NSs from NNSs, which is of little practical value as a diagnostic tool. Instead, it was argued that a measure of the overall commonality of L2 learners' associative domain as well as its size better reflect their lexical organization in terms of its overall consistency and breadth of connections. Further, mere metacognitive awareness of lexical knowledge was also deemed insensitive to distinguishing differences among L2 learners in the more advanced stages of L2 proficiency, probably because the participants were adult L2 learners who acquired English through formal instruction; in general, formal instruction is thought to promote the development of explicit knowledge (Ellis, 2004). In any event, it is of value to have experimental support for the idea that at the higher levels of proficiency, L2 learners' metacognitive awareness is as strong as NSs', which only means that at the higher proficiency levels, L2 learners exercise conscious control over their lexical knowledge as efficiently as NSs, which might be a characteristic of development rather than of language learning.

In conclusion, Read (2000) pointed out that L2 vocabulary assessment needs measures that are practical to use and valid indicators of word knowledge. Similarly, Laufer and Nation (1999) argued that lexical research should draw on a range of well thought out and empirically tested measures to gain a better understanding of L2 vocabulary knowledge (i.e., measures that are sensitive to proficiency and practically useful as diagnostic tools). To this end, we found that measures of quantity and quality of vocabulary knowledge (particularly vocabulary size, word frequency effects, number of associations, and within-group consistency of associative responses) are effective in distinguishing proficiency differences between intermediate and advanced L2 learners. Thus, the present study was an attempt to contribute to the collective effort in this area by proposing a framework for examining L2 lexical knowledge of individual items, a sampling procedure with relatively high reliability, and a set of measures that showed sensitivity to capturing the variation in the overall state of L2 learners' lexical knowledge at higher levels of proficiency.

(*Received 2 February 2005*)

NOTES

1. The set of the inflectional morphemes of English can be controversial, particularly regarding *-ing* and *-ed* suffixes, which can be both inflectional and derivational (for more details, see Bauer, 1983; Bauer & Nation, 1993). To maintain consistency in classifying the associative responses, these

https://doi.org/10.1017/S0272263105050254 Published online by Cambridge University Press

two suffixes were treated as derivational because the type of task (single-word associations) did not allow for fine distinctions.

2. It might be argued that including each participant's own responses in the estimates of native-like commonality of associations inflated the estimate for NSs. To correct for this, we removed each NS's associations from the base from which the NS's commonality of associations was estimated. We continued to find a main effect of language proficiency, $F(2, 61) = 24.77$, $p < .001$, with native English speakers demonstrating greater commonality of associations, $M = 260.80$, $SD = 66.66$, than advanced learners, $M = 198.29$, $SD = 88.53$, and intermediate learners, $M = 103.12$, $SD = 70.01$. Thus, the findings that we report for this variable cannot be attributed to the way in which these estimates were tabulated.

REFERENCES

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

Bauer, E. (1983). *English word-formation*. New York: Cambridge University Press.

Bauer, L., & Nation, I. S. P. (1993). Word families. *International Journal of Lexicography*, *6*, 253–279.

Bialystok, E., & Ryan, E. (1985). A metacognitive framework for the development of first and second language skills. In D. L. Forrest-Pressley, G. E. MacKinnon, & T. G. Waller (Eds.), *Metacognition, cognition, and human performance* (pp. 207–246). San Diego, CA: Academic Press.

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

Chapelle, C. (1998). Construct definition and validity inquiry in SLA research. In L. F. Bachman & A. D. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 32–70). New York: Cambridge University Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Mahwah, NJ: Erlbaum.

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

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

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

Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning*, *54*, 227–275.

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

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

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

Gass, S. M., & Selinker, L. (2001). *Second language acquisition: An introductory course*. Mahwah, NJ: Erlbaum.

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

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

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

Hornby, A. S. (1978). *Oxford student's dictionary of current English*. Oxford: Oxford University Press.

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

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

Landau, S. I. (2001). *Dictionaries: The art and craft of lexicography*. New York: Cambridge University Press.

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In P. J. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 126–132). London: Macmillan.

https://doi.org/10.1017/S0272263105050254 Published online by Cambridge University Press

Laufer, B. (1997). The lexical plight in second language acquisition: Words you don't know, words you think you know, and words you can't guess. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 20–34). New York: Cambridge University Press.
Laufer, B., & Nation, I. S. P. (1999). A vocabulary size test of controlled productive ability. *Language Testing*, *16*, 33–51.
Laufer, B., & Paribakht, S. (1998). The relationship between passive and active vocabularies: Effects of language learning contexts. *Language Learning*, *48*, 365–391.
Lorge, I., & Chall, J. (1963). Examining the size of vocabularies of children and adults: An analysis of methodological issues. *Journal of Experimental Education*, *32*, 147–157.
Meara, P. (1978). Learners' associations in French. *Interlanguage Studies Bulletin*, *3*, 192–211.
Meara, P. (1984). The study of lexis in interlanguage. In A. Davies, C. Criper, & A. P. R. Howatt (Eds.), *Interlanguage* (pp. 225–235). Edinburgh: University of Edinburgh Press.
Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjær, & J. Williams (Eds.), *Competence and performance in language learning* (pp. 35–53). New York: Cambridge University Press.
Meara, P., & Jones, G. (1988). Vocabulary size as a placement indicator. In P. Grunwell (Ed.), *Applied linguistics in society* (pp. 80–87). London: Center for Information on Language Teaching and Research.
Melka, F. (1997). Receptive vs. productive aspects of vocabulary. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition, and pedagogy* (pp. 84–102). New York: Cambridge University Press.
Nagy, W., & Anderson, R. (1984). How many words are there in printed school English? *Reading Research Quarterly, 19*, 304–330.
Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Rowley, MA: Newbury House.
Nation, I. S. P. (1993). Using dictionaries to estimate vocabulary size: Essential, but rarely followed procedures. *Language Testing*, *10*, 27–40.
Nation, I. S. P. (2001). *Learning vocabulary in another language*. New York: Cambridge University Press.
Nelson, K. (1977). The syntagmatic shift revised: A review of research and theory. *Psychological Bulletin*, *84*, 93–116.
Nusbaum, H. C., Pisoni, D. B., & Davis, C. K. (1984). *Sizing up the Hoosier mental lexicon*: *Measuring the familiarity of 20,000 words*. (Research on Speech Production Rep. No. 10). Bloomington: Indiana University.
Paribakht, T., & Wesche, M. (1993). The relationship between reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal*, *11*, 9–29.
Popham, W. J. (1990). *Modern educational measurement: A practitioner's perspective*. Englewood Cliffs, NJ: Prentice-Hall.
Qian, D., & Schedl, M. (2004). Evaluation of an in-depth vocabulary knowledge measure for assessing reading performance. *Language Testing*, *21*, 28–52.
Read, J. (1993). The development of new measure of L2 vocabulary knowledge. *Language Testing*, *10*, 355–371.
Read, J. (2000). *Assessing vocabulary*. New York: Cambridge University Press.
Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quarterly*, *10*, 77–89.
Sandra, D. (1994). The morphology of the mental lexicon: Internal word structure viewed from a psycholinguistic perspective. *Language and Cognitive Processes*, *9*, 227–269.
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*, 189–216.
Szalay, L. B. (1984). An in-depth analysis of cultural/ideological belief systems. *Mankind Quarterly*, *25*, 71–100.
Szalay, L. B., & Brent, J. (1967). The analysis of cultural meaning through free verbal associations. *Journal of Social Psychology*, *3*, 247–258.
Szalay, L. B., Lynse, D. A., & Bryson, J. A. (1972). Designing and testing cogent communication. *Journal of Cross-cultural Psychology*, *3*, 247–258.
Szalay, L. B., & Windle, C. (1968). Relative influence of linguistic versus cultural factors on free verbal associations. *Psychological Reports*, *22*, 43–51.
Thorndike, E. L., & Lorge, I. (1944). *The teacher's word book of 30,000 words*. New York: Teachers College, Columbia University.
Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics*, *22*, 217–254.

https://doi.org/10.1017/S0272263105050254 Published online by Cambridge University Press

Vierra, A., & Pollock, J. (1992). *Reading educational research*. Englewood Cliffs, NJ: Prentice-Hall.
Wesche, M., & Paribakht, S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review*, *53*, 13–40.
West, M. (1953). *A general service list of English words*. London: Longman.
Wilks, C., & Meara, P. (2002). Understanding word webs: Graph theory and the notion of density in second language word association networks. *Second Language Research*, *18*, 303–324.
Wolter, B. (2001). Comparing the L1 and L2 mental lexicon. *Studies in Second Language Acquisition*, *23*, 41–69.
Wolter, B. (2002). Assessing proficiency through word associations: Is there still hope? *System*, *30*, 315–329.
Xue, G., & Nation, I. S. P. (1984). A university word list. *Language Learning and Communication*, *3*, 215–229.
Yoshida, K. (1990). Knowing vs. behaving vs. feeling: Studies on Japanese bilinguals. In L. A. Arena (Ed.), *Language proficiency* (pp. 19–40). New York: Plenum Press.
Zechmeister, E. B., Chronis, A. M., Cull, W. L., D'Anna, C. A., & Healy, N. A. (1995). Growth of a functionally important lexicon. *Journal of Reading Behavior*, *27*, 201–212.
Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The educator's word frequency guide*. New York: Touchstone Applied Science Association.

APPENDIX A

Appendix A1. List of the SWs used in the study grouped by frequency of occurrence

SFI = 20.0–29.9		SFI = 30.0–30.9		SFI = 40.0–49.9		SFI = 50.0–70.0+	
Freq.	SW	Freq.	SW	Freq.	SW	Freq.	SW
20.8	rigidity	30.7	tuner	40.4	fathom (v.)	50.6	beaten
22.1	abattoir	32.9	entrust	41.7	naïve	50.7	sweep (v)
22.1	bursar	32.9	flagstone	41.9	custodian	51.5	diamond (n.)
22.1	contravention	34.0	lackadaisical	41.9	drawback (n.)	51.6	telegraph (n.)
22.1	gambol (v.)	34.1	jib (n.)	42.2	spotlight (n.)	51.7	yield (v.)
22.1	macaw	34.9	parable	42.6	sleigh (n.)	53.4	hunger (n.)
22.1	masochism	35.0	dissension	42.8	livelihood	54.3	blanket (n.)
22.1	putative	35.0	savor (v.)	43.1	pillar (n.)	54.6	ruler
25.1	virtuosity	35.4	monorail	43.4	advantageous	55.5	official (adj.)
25.5	amoral	35.5	edifice	43.8	bracelet	56.9	throw (n.)
26.9	point-blank (adj.)	35.5	middling (adj.)	44.9	choke (n)	57.1	refuse (v.)
27.7	crampons	36.0	prefect	46.4	toxic	58.7	settled
29.3	glower (v.)	36.5	forgo	46.4	weaken	62.4	wind (v.)
		36.6	unnerve	46.6	experimentation	65.3	second (adj.)
		37.2	instill	46.8	array (v.)	65.8	hard (adv.)
		37.4	inception	46.8	defensive	66.9	high (adv.)
		37.4	solstice	47.4	shuttle (n.)	70.3	back (v.)
		38.1	penance	48.6	residence	73.7	out
		38.4	cassava	49.7	studio		
		38.7	promontory				
		39.1	griddle (n.)				
		39.4	concede				
		39.6	coinage				

Note. Freq. = frequency of occurrence, v = verb, n = noun, adv. = adverb, adj. = adjective.

https://doi.org/10.1017/S0272263105050254 Published online by Cambridge University Press

APPENDIX B

INSTRUCTIONS TO THE TEST-TAKERS AND AN EXAMPLE OF A TEST ITEM

Instructions

The following activity will ask you to rate your degree of familiarity with a list of words. Please respond to each word in italics as accurately as you can. Do not leave any of the items unanswered. For option V, write down as many as *three* words that you associate with the item in italics. When the lexical category is specified next to the word in brackets (e.g., *noun* [n.], *verb* [v.], *adjective* [adj.], *adverb* [adv.]), please respond to the word as specified. *Note*. *If you choose option III or IV, you should do option V as well!*

Example of a Test Item

1. abattoir
   - I. I *have not seen* this word before ☐
   - II. I *have seen* this word before but I *don't remember* what it means ☐
   - III. I *think* this word means _____ (synonym, translation, or brief explanation)
   - IV. I *know* that this word means _____ (synonym, translation, or brief explanation)
   - V. I *associate* this word with _____, _____, _____

https://doi.org/10.1017/S0272263105050254 Published online by Cambridge University Press