


Eihab Abu-Rabiah

Hebrew Language Department,
Kaye Academic College of Education, Israel

 <https://orcid.org/0000-0002-8837-1089>

Exploring Lexical Sophistication in Second Language: An Analysis of Vocabulary Using a Word-Rating Method

Abstract

Previous research has established that determining lexical sophistication (i.e., the percentage of sophisticated words in a text) through the judgment of teachers on a corpus of words is a more accurate method than relying on word frequency-based lists. However, this approach can be time-consuming. To overcome this drawback, a new method is proposed in this study, which involves rating specific words out of context.

A list of 68 words that appeared in approved high-school textbooks of teaching Hebrew to Arabic speakers was given to six experienced Hebrew teachers, who then categorized the words into four levels of lexical sophistication: (1) very basic words to (4) very advanced words. From this, a list of 28 words was created, with seven words from each level, and the lexical sophistication level was agreed upon by two-thirds of the teachers. Nineteen Arabic-speaking learners of Hebrew were asked to define the chosen words (passive vocabulary) and compose a sentence including each (controlled-active vocabulary) in a test-retest study at two time-points: the 11th and 12th grade.

The results indicated that although there was no significant increase in lexical sophistication over time, significant differences emerged between the four levels of lexical sophistication, with students' accuracy decreasing as the level of lexical sophistication increased. Additionally, only in the 11th grade was passive vocabulary found to be significantly larger than controlled-active vocabulary. However, as acquisition time increased, the gap between these two vocabulary types narrowed, due to improved performance in the controlled-active task. Furthermore, a significant correlation was found between passive and controlled-active vocabulary, which became stronger with more acquisition time.

Keywords: language assessment, second language acquisition, passive vocabulary, controlled-active vocabulary, advanced/sophisticated words

State of the Art

The various lexical indices in use today assume that there are two main ways to evaluate a learner's vocabulary: by evaluating the size of the vocabulary (vocabulary breadth knowledge) and by evaluating the quality of the vocabulary (vocabulary depth knowledge) (Nation, 1990; Milton, 2009; Ellis, 2015). The index of lexical diversity applies only to the breadth of knowledge, as it evaluates the variety of words in a sample of speech or writing (Read, 2000). This is typically done by calculating the ratio of unique words (types) to total words (tokens) (Jarvis, 2017; Abu-Rabiah, 2023). However, it is important to also examine the depth of knowledge, as high lexical diversity does not always point to high lexical proficiency—the extent of diversity can be expressed in very simple words which do not testify to lexical sophistication. The index of lexical sophistication differs from the index of lexical diversity, as it attributes a different level of importance to different words, as opposed to simply counting the number of distinct types (Daller et al., 2003). Consider the following two sentences:

- Pupils answered questions.
- Looters smashed panes.

Each one includes three tokens and three types, and so their lexical diversity stands at 100%. Both sentences possess the highest possible level of diversity, but with an essential difference: the first is comprised of simple words which do not testify to a high level of proficiency, while the second sentence is based on advanced vocabulary which testify to the learner's high lexical proficiency. This means that the quality of word use in the two sentences is not identical. As a result of this difference between words, researchers proposed an index of lexical sophistication that enables a distinction between simple and advanced words.

Lexical sophistication is defined as the percentage of sophisticated words or advanced words in a text (Lindqvist et al., 2011). This index also allows the quantification of the appropriate use of low frequency vocabulary items (Malvern et al., 2004). It also enables testing the use of the technical terms and jargon which allow the author to express meaning in a more precise and acceptable manner, given the discipline about which he or she is communicating (Read, 2000). There is no agreement regarding the exact meaning of *sophisticated words* or *advanced words*, leading different researchers to define these terms differently (Lindqvist et al., 2011). One approach is to view them as rare words (Vermeer, 2000; Kyle & Crossley, 2016), where the first approach to testing lexical sophistication developed on the basis of this definition.

Testing Lexical Sophistication According to Frequency Tables

The first method of testing lexical sophistication is based on frequency lists. Frequency is an important factor in acquiring a vocabulary, and is therefore important when testing vocabulary (Ellis, 1997, 2002; Nation, 2001, 2006; Cobb & Horst, 2004; Milton, 2007; Kojima & Yamashita, 2014). Common words are easier to acquire than rare words (Tidball & Treffers-Daller, 2008). The most common words are accessible to most learners, but rare words are only accessible to the advanced learner. The learners' lack of acquaintance with the most frequent words points to their relatively meager vocabulary, as it does not include common words which are frequently used. The use of advanced words in written texts is an indicator of high language proficiency (Linnarud, 1986) and is a sign of a rich vocabulary and of the learner's academic success (Laufer & Nation, 1995).

Knowing the 2,000 most common words in a target language (English as a second language, or L2) is a basic threshold for verbal communication, especially at the beginning of the process of language acquisition (Laufer, 2005; Azodi et al., 2014). Furthermore, the 2,000 most common words make up 87% of written texts and 80% of typical academic texts in English, while the 1,000 most common words make up 75% of official written texts in English (Laufer & Nation, 1999).

Scholars proposed various indices for testing sophistication based on frequency tables, such as the Advanced Guiraud, P-LEX, LFP, S, Advanced TTR, and others. These are used to calculate the ratio between the number of advanced/sophisticated words and the total number of words in a given corpus (Kojima & Yamashita, 2014). One of the first tools using this method is termed the Lexical Frequency Profile (LFP). This tool enables the evaluation of the learner's productive vocabulary size by testing the words he wrote or said according to the frequency in which the various words appear (Laufer & Nation, 1995; Laufer, 2005).

A second tool for testing lexical sophistication based on frequency tables is the P-LEX (Meara & Bell, 2001). This tool divides the text into segments of ten words and calculates the number of *difficult* words in each segment. This tool uses the term *difficult words* to refer to words that are not on the short list of the 1,000 most common words (Meara, 2001). Higher calculated values point to a larger vocabulary. Beyond 2,000 is yet another tool that was proposed for testing lexical sophistication, also based on the division into levels of frequency. This tool distinguishes between two frequency groups: the 2,000 most common words, which make up a basic vocabulary, and words which indicate a more advanced vocabulary (Laufer, 1995).

The aforementioned indices, which test lexical sophistication, calculate the ratio between advanced words and the total number of words (Kojima & Yamashita, 2014). Researchers noted the disadvantages of each of these indices,

for example, beyond 2,000 is inefficient when testing low levels of English proficiency, as learners on these levels do not know many words beyond the 2,000 most common ones (Meara & Bell, 2001). This claim is also valid for the S index, which does not distinguish between basic words and advanced words, but rather focuses only on the advanced words which are classified into different levels of frequency. This tool assumes that advanced words which are close in frequency to the common words should not be considered advanced and rare, since as the learner's vocabulary grows, so does the frequency range of the words he uses (Kojima & Yamashita, 2014).

There are studies which tested lexical sophistication through the combined use of several different tools. In a study of English as L2 to Japanese, as reflected in essays (the average length of each essay is about 400 words) which students wrote in 60 minutes, lexical sophistication was tested by using Advanced TTR, Beyond 2000, S, and P-LEX (Kojima & Yamashita, 2014). In this study grammatical errors were corrected, words that were not in English were not included, and if the usage of a word was incorrect, then it was not counted, since it was inadequately acquired by the learner. Results of this study showed that the P-LEX and S indices are not influenced by text length, while Beyond 2,000 and Advanced TTR are more sensitive to text length, and therefore it is inadvisable to use them for texts of different lengths. The researchers noted that there is an optimal length of texts which these indices test: The S index is suitable for texts of around 200 words, the P-LEX for around 300 words, and Beyond 2,000 and Advanced TTR are best used for texts over 1,000 words in length (Kojima & Yamashita, 2014).

Other studies found testing lexical sophistication using exclusively frequency-based tools to be insufficient, and that other methods based on different considerations (such as teachers' evaluations) are more efficient for testing lexical sophistication. This is true because a word's level of lexical sophistication results from many factors, of which frequency is only one (Horst & Collins, 2006; Tidball & Treffers-Daller, 2008).

Testing Lexical Sophistication According to Teacher Evaluation

The second method for testing lexical sophistication relies upon teachers' evaluations of sophisticated or advanced words. This method recognizes the fact that the level of a word's sophistication is not only a function of frequency, since there are rare words that are easily acquired. For example, rare words in the mother tongue (L1) which have a cognate in L2 will be acquired more easily than other words within the same level of frequency which are not cognates (Tidball & Treffers-Daller, 2008). Despite their rarity, these words are

easily acquired thanks to their cognate in L1, and for those learners are not to be considered sophisticated or advanced.

A longitudinal study tested the writing skills of French speakers who were studying basic English before and after 400 hours of instruction, gathering a corpus of 80,000 words. Results showed that the learners preferred the use of rare words which are cognates to words in their mother tongue to common words which are not cognates (Horst & Collins, 2006). The researchers noted an increase in the learners' lexical sophistication over time: they used fewer cognates as time went by, interspersed more of the common words, and even used morphologically complex forms. This means that over time there was an increase in the level of lexical sophistication, but relying exclusively on word frequency tables (like the LFP) would not have enabled the researchers to detect this lexical progress.

The approach based on teachers' evaluations developed as a result of the various factors which can influence the level of word sophistication. Using this approach, the researchers give teachers words to rate according to a scale ranging from basic/unsophisticated/common to advanced/sophisticated/rare. The number of levels on the scale is not fixed, but varies according to the evaluations of both the researcher and the experienced teachers. Some researchers give the teachers a list of all the word types in a particular corpus and ask them to rate the words on a basic-advanced scale, after which they measure the ratio of advanced words to all the words in the text. For example, one study using this method (Daller et al., 2003) had seven teachers of Turkish as L2 rate 693 types written by students learning Turkish at a basic level. The teachers were asked to classify the types into three groups characteristic of those who learn Turkish as L2: one group of words for learning on a basic level, one for learning on an intermediate level, and one for learning on an advanced level. The teachers sorted the words into these three groups based on their teaching experience. The researchers found that the classification of advanced and basic words based on the evaluations of experienced teachers was an efficient method for testing progress in vocabulary acquisition. This finding can be explained in two primary ways: first, frequency tables sometimes include words closer to those in the higher level as well as closer to those in the lower level within the same list. This means that the level of frequency is a graduated property that does not clearly divide into separate groups. Second, the teachers' operationalization for what is considered a basic vocabulary is more precise than relying upon frequency lists. This means that teachers define basic vocabulary more precisely than they do advanced vocabulary, and even distinguish better between different levels of frequency (Tidball & Treffers-Daller, 2008).

A similar study of lexical sophistication in French as L2 for English speakers had three experienced teachers rate 932 word types in a corpus written by the students, according to a scale comprised of seven levels of lexical sophis-

tication: Level 1 marked the most basic words, and Level 7 marked the most advanced words. The researchers found that determining lexical sophistication based on teachers' evaluations is more precise than tests which are based on word frequencies (Tidball & Treffers-Daller, 2008).

The disadvantage of testing lexical sophistication using teachers' evaluations of types in a particular corpus is that not all the declensions of any given type are equal in their level of sophistication. Some of the declensions are more basic, while others are more advanced. The declensions within the different levels of difficulty exist in basic types, such as *child*, *children*, and *childhood*, and also in the advanced words, such as *chamber* and *chambers*. In addition, this method requires much of the teachers' time, as rating hundreds of words is not an easy task, especially when rating on a multi-level scale (Abu-Rabiah, 2022).

Studies of Lexical Sophistication Using Both Methods

One study tested the relationship between passive vocabulary knowledge and lexical sophistication in writing, as well as general lexical richness, using four frequency groups: Up to 2,000 words, 2,000–3,000, 3,000–5,000, and 5,000–10,000 (Henriksen & Danelund, 2015). Speakers of English as L2 were asked to write argumentative essays in 90 minutes with no assistance. Results show that students with a broader active vocabulary tend to have higher levels of lexical sophistication. A slight correlation was found between the use of sophisticated words and the general writing score as well (Linnarud, 1986).

A number of lexical richness indices, including lexical sophistication and lexical diversity, were tested on essays written by Hebrew- and Arabic-speaking first-year English Literature students at an Israeli university. These students wrote two essays in English as L2, one before the beginning of their studies and the second after their first or second semester. Both tasks were part of a university exam—ensuring the students' high motivation for success—without assistance and without prior knowledge of the essay topic. Results demonstrate that the percentage of frequent words declined over time, while the percentage of rare words increased. The lexical profile of the students had changed; however, it was still twice as low as that of native English speakers (Laufer, 1991 b).

Similarly, Astridya (2018) tested a number of lexical richness indices, including lexical sophistication and lexical diversity, in argumentative essays written in English as L2 by native Indonesian high-school students in the 10th, 11th, and 12th grades. Both lexical sophistication and lexical diversity increased with age, where the 12th-graders had the highest level of lexical richness. In 10th grade the students' vocabulary was still limited, with repetitive use of the same types; the 11th graders started using vocabulary appropriate to the given

topic; while the 12th graders clearly and carefully chose the words they wrote. It seems that these students improved their writing skills by choosing the words best-suited for argumentative writing, with skills increasing according to age and exposure to the language.

Lindqvist et al. (2011) also measured different indices of lexical richness in French and Italian as L2 for native Swedish speakers by using word frequency tables in all three languages. Lexical richness indices between speakers of different levels of proficiency within each language were also compared. They found that there are differences in the various lexical richness indices between the different groups, according to the different levels of proficiency: In French, the lexical profile of high-proficiency students was identical to that of native French speakers, while in Italian, the lexical profile of high-proficiency speakers was nowhere near that of native Italian speakers. This is explained by the fact that some of the highest-proficiency French speakers had spent a few years in the target language country, while the Italians did not.

In a different study on lexical sophistication and lexical diversity, Waldvogel (2014) found a significant increase in the two indices between middle- and high-level students of Spanish as L2. He noted that the increase in lexical richness of the high-proficiency group was slower than that of the less proficient group, because the latter felt that they still had much to learn, and therefore put in a greater effort. Contrastingly, the former felt that they had learned a lot, with their previous learning sufficing, and therefore did not put in much effort. This phenomenon, the Active Vocabulary Threshold Hypothesis, posits that a learner's vocabulary increases up to the average level of the group they are in, after which they no longer make an effort to broaden their lexical repertoire and invest less in the learning process. Similar findings were reported in a study on the lexical sophistication L1 and L2 English writers (Kwon, 2009).

To the best of my knowledge, this is the first study which tested lexical proficiency in Hebrew in general, and in Hebrew as L2 for Arabic-speakers in particular.

Limitations of Testing Lexical Sophistication

In general, word frequency changes according to the written corpus upon which the frequency tables were created. A word can appear very frequently in an academic corpus, but very rarely in a journalistic corpus. General corpora collected from a variety of internet sites are not identical, and the same word can appear in different levels of frequency. For example, in the corpus of the 10,000 most common words in Hebrew, the indefinite type *baayit* [house], without a preposition or subordinating conjunction, is listed as one of the 50

most common words. As opposed to this, this word is only one of the 300 most common words in sketchengine's heTenTen's (Hebrew) corpus.

Furthermore, frequency tables generally do not distinguish between acquiring a language as L1 or L2, though there is clearly no symmetry between acquiring a language as L1 and acquiring it as L2. A word acquired in an intermediate level in L1 may be acquired at an earlier stage in L2 due to its necessity given the learning environment and the average age of the learners.

Testing lexical sophistication in Hebrew based on frequency tables has several disadvantages. First, the distribution list of words in Hebrew is based on orthographic types. This means that there are no frequency tables of lemmas. In this case, we cannot derive exact conclusions about the frequency level of the lemma. This difference is especially notable when the lemma is accompanied by a definite tag, or by a preposition or subordinating conjunction, as opposed to not being thus accompanied. For example, in sketchengine's heTenTen frequency list: The type *ha-baayit* [the house] is listed as one of the 200 most common words, but the same type without the definite article and without preposition or subordinating conjunction, *baayit* [house], is rarer (among the 300 most common words), and when accompanied by the preposition 'in' [*babayit*, in the house] it is even rarer (among the 900 most common words).

Second, Hebrew frequency tables do not distinguish between different types written in a similar manner (homographs). For example, in Wiktionary's Hebrew frequency list, the words *em/im* [mother/if] and *ha-em/ha-im* [the mother/is it] are among the 100 most common words, the word *ve-em/ve-im* [and the mother/and if] is among the 500 most common words, and the word *she-em/she-im* [that the mother/that if] is among the 800 most common words. In all these words there is no way of knowing whether they refer to a noun (*em*, mother) or a conditional (*im*, if). The prevalence of homographs in Hebrew, due to its consonantal writing system (Cook, 2016), poses a significant challenge in language testing contexts. Given their occurrence in approximately 23% of isolated words within the language (Shimron & Sivan, 1994), reliance solely on frequency lists for language testing purposes becomes less than optimal.

Testing lexical sophistication in Hebrew using teachers' evaluations, as in the previous studies, also has limitations. First, similar to testing lexical sophistication through frequency tables, different tokens of the same type do not always reflect uniform levels of difficulty or distribution. Second, looking for experienced teachers who teach Hebrew as a L2 to native Arabic speakers who are willing to voluntarily rate all the word types in a given corpus according to levels of lexical sophistication is unrealistic, as this type of classification requires time and effort.

Given all the disadvantages and limitations of the two methods of testing lexical sophistication, I used a practical new method which enabled me to overcome the aforementioned disadvantages and limitations while adapting to the characteristics of the tested language.

Methodology

Consolidating a List of Words and Determining Levels of Lexical Sophistication

I conducted a survey of Israeli Ministry of Education approved Hebrew L2 textbooks specifically designed for Arabic speakers. These textbooks were used in Arabic-speaking schools in the Negev in southern Israel over the past five years. I put together a list of 68 words, characteristic of Hebrew as L2 for high-school-aged Arabic speakers, which appeared in many of the texts within the different textbooks. This list was given to six Hebrew teachers who are native Arabic speakers themselves, each of whom had at least two years of teaching experience. The teachers were asked to rate the 68 words according to 4 levels of lexical sophistication:

- 1 = Very basic/common words
- 2 = Basic/common words
- 3 = Sophisticated/advanced/rare words
- 4 = Very sophisticated/advanced/rare words

The choice of four levels of lexical sophistication, and not three as in Daller et al. (2003) or seven as in Tidball & Treffers-Daller (2008), stemmed from my personal experience of rating the words. I found that classifying the words into three levels of lexical sophistication did not adequately express the gaps in lexical sophistication between words of average frequency, as some are closer to the common words and others are closer to the rare words. Therefore, I added a fourth level of lexical sophistication. Trying to rate the words according to five or even seven levels of lexical sophistication was not practical in this case, as the borders between adjacent levels of lexical sophistication on both ends of the scale were amorphic, akin to the difference between the two first levels on a five-level scale or the difference between the two lowest levels on a seven-level scale.

The teachers' ratings were carried out according to their evaluations, based on their knowledge which also included the curricula for Hebrew as L2 to Arabic speakers, the levels of texts that they teach in their various high-school classes, the texts on which the students are tested in their matriculation exams, and more. The teachers' ratings led to the creation of a list of 28 words which are equally distributed between the four levels of lexical sophistication, with seven words in each level. Only words which at least two-thirds of the teachers (four teachers out of six) agreed belonged to the same level of lexical sophistication were included in the task which was later given to the students. Note that there were words not included in the above list which, though at least four out of six teachers agreed upon, were not included in the task in order to keep

the number of words in each level of lexical sophistication equal (seven in each level), a number also dictated by the time limitation given for the students' task.

Tasks

As opposed to testing lexical sophistication through teachers' evaluations of unstructured essays (knowledge of active/productive vocabulary words), in this study the lexical sophistication tasks were constructed so as to test not only the level of word lexical sophistication but also the distinction between passive knowledge and active/productive knowledge, or between the students' vocabulary size/breadth and the depth of their vocabulary. In these tasks the participants were asked to define each of the 28 chosen words, in either Hebrew or Arabic—testing the breadth of vocabulary and passive knowledge—and to also insert each word within a sentence of their own (in L2), a task which tests vocabulary depth and controlled active knowledge. The definition of each word, as well as its insertion within a sentence, enables an examination of the relationship between these two types of vocabulary as well as of the relationship between them throughout the acquisition process.

In many cases, it is possible to understand a new word given its context, and therefore in this task, the words were given on their own, without context. This enables an examination of the breadth of the learners' vocabulary as it stands, with no additional factors related to the learner's ability to connect pieces of new information and understand new words from context.

Participants

The participants in the current study were Arab high school students in northern Negev, Israel. They started learning Hebrew in the second grade and studied about three to five hours a week as part of their formal education (Abu-Rabiah et al., 2023). The 19 research participants were students enrolled in the same science class, exhibiting comparable levels of academic achievement. During the study, they were taught Hebrew by an Arabic-speaking teacher as L1 (This population was described in detail in Abu-Rabiah, 2020). They performed the task at two-time points: the beginning of the year in the 11th grade and the beginning of the year in the 12th grade.

Research Questions and Hypotheses

RQ1-a: Do the differences in difficulty between the four levels of lexical sophistication match those given by the teachers?

Differences in difficulty between the four levels of lexical sophistication are expected to match the teachers' evaluations, as the teachers have much experience teaching Hebrew as L2, including preparing the students for their matriculation exams. The students' performance is expected to be lower as the level of lexical sophistication, according to the teachers' evaluations, rises.

RQ1-b: Are there differences in the students' improvement within the different levels of lexical sophistication?

I hypothesize an improvement. This improvement is expected to be relatively greater in the higher levels of lexical sophistication (3 and 4) than in the lower levels (1 and 2), as the lower levels of lexical sophistication include words which are considered to be easy, and as the students mature, they are more likely to be exposed to the more difficult words.

RQ2: Does the students' general level of lexical sophistication rise over time?

I expect an increase in the students' lexical sophistication over time, as they had many Hebrew classes during this period in which they worked on different types of exercises. In addition, they also heard Hebrew spoken outside of the classroom, both in formal settings (such as outings) and in informal settings (such as the media, visiting cities of Hebrew-speakers, and more).

RQ3-a: Are there differences between the size/breadth of the passive vocabulary (the word-defining task) and the controlled active vocabulary (the word-insertion task), and in the degree of improvement, over time?

A difference between depths of the two types of vocabulary is expected, as found in many studies, the passive vocabulary is generally larger than the controlled active vocabulary. In order to express one's passive knowledge, meaning to use it actively, the learner must control other linguistic skills beyond lexical ability, including a knowledge of syntax and grammar. As to the degree of improvement—no difference is expected as the result of instruction. This is because during the process of instruction, which is based on the curriculum of teaching Hebrew to Arabic speakers, no emphasis is placed on either of the two types of vocabulary.

RQ3-b: Is there a correlation between the two types of vocabulary?

The hypothesis of this study is that there will be a correlation between the passive vocabulary and the controlled active vocabulary, since the process of instruction includes not only the meaning of the words, but practice using these words within sentences as well. Thus, both passive and active knowledge are combined in the learning process. In addition, I expect this correlation to be stronger during the second time period (12th grade) as the instruction process continues, as does the expected increase in the students' linguistic

abilities, since the learners have more opportunities to use both their active and passive vocabularies.

Results and Discussion

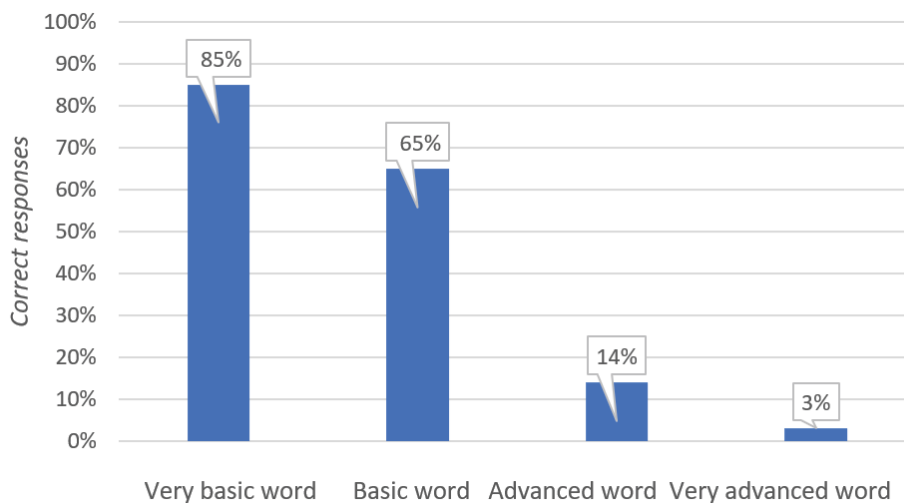
The Relationship Between the Level of Lexical Sophistication and the Degree of Improvement Over Time, and Teachers' Evaluations

RQ1

The following figure presents the percentage of correct responses within the four levels of lexical sophistication.

Figure 1

Percentage of Correct Responses within the Four Levels of Lexical Sophistication



The first part of the first research question looked for differences between students' achievements in the different word groups, which were based upon the words' level of lexical sophistication. This question ascertains whether or not the teachers' evaluations constitute a valid measure of the words' level of lexical sophistication (difficulty level). I expected differences between the different word groups according to levels of difficulty to match the teachers'

evaluations of the words. This is because the teachers have much experience teaching Hebrew as L2 for Arabic. I also expected the students' performance to be lower as the level of lexical sophistication rises, with the best performance in Lexical Sophistication Level 1 (very basic words) and the lowest performance in Lexical Sophistication Level 4 (very advanced words).

The results supported this hypothesis (see Figure 1), and show that generally—in both classes together—the differences between the two word groups are highly significant (chi-square test value: $p < .00$). In addition, as the level of lexical sophistication rises, the students' performance decreases (chi-square test value: $p < .00$) (see Table 1).

Table 1

Differences Between the Four Levels of Lexical Sophistication

Answer	Lexical Sophistication Level				Total	Chi-Square Test
	Very basic word	Basic word	Advanced word	Very advanced word		Test value
Correct	453	346	77	15	891	1024.133
Incorrect	79	186	455	517	1237	P-value
Total	532	532	532	532	2128	0.000

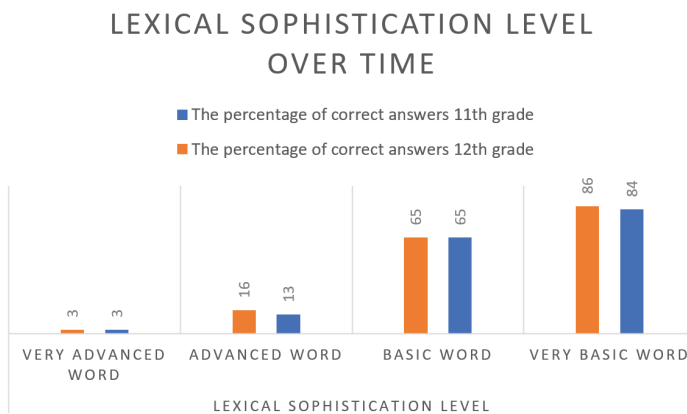
The easiest group is that of the most basic words (Lexical Sophistication Level 1): 453 correct responses and 79 incorrect responses; 85% of the responses were correct. The second easiest group is that of the basic words (Lexical Sophistication Level 2): 346 correct responses and 186 incorrect responses; 65% of the responses were correct. In these two groups almost two-thirds of the given responses were correct. The students apparently had mastery of these words. The next two groups of words were much more difficult. The third easiest group is that of the advanced words (Lexical Sophistication Level 3): 77 correct responses and 455 incorrect responses; 14% of the responses were correct. The fourth, and most difficult group, is that of the very advanced words (Lexical Sophistication Level 4): 15 correct responses and 517 incorrect responses; only 3% of the responses were correct. The students' great difficulty with the two groups of advanced words is apparent. Success in these groups is very limited, where less than one-fifth of all responses were correct. As noted above, there are significant differences between all groups of words, but the differences between the two basic groups and the two advanced groups are especially prominent. There is a 20% difference of correct answers between the two basic groups and an 11% difference between the two advanced groups. As

opposed to this, there is a 51% difference between the group of basic words (Lexical Sophistication Level 2) and the group of advanced words (Lexical Sophistication Level 3).

Similar to the general findings in the two age groups, results show that each of the two age groups shows statistically significant differences between the different groups of words, where the order of difficulty is identical: The most difficult level is the Very advanced words, then the Advanced words, followed by the Basic words and the Very basic words (chi-square test value: $p < .00$).

Figure 2

Lexical Sophistication Level over Time



The differences between success in the groups of basic words and the groups of advanced words can testify to the students' higher level of exposure to advanced words than to basic words. The expectation was that given the rise in instruction time, there will be greater exposure to the advanced words, which would be expressed in the students' increased mastery of the words in the advanced groups. However, results show that these high-school students have only a most basic grasp of the advanced Hebrew words (see Figure 2). These include words which are essential to their next stage in life—academic studies and the job market, which are difficult to navigate with only a basic vocabulary. Words which are considered advanced in relation to the study population are basic in many academic and work-related contexts, such as *zakai* (entitled), *hithayvut* (commitment), *tekes* (ceremony), and others. In the highest level of lexical sophistication, there was no improvement at all. This group included words such as *higyenah* (hygiene), *karukh* (encompass), *hikhriya* (determined), *poreh* (fruitful), *hasagah* (achievement), and others. Some of these words are considered basic in various situations and different places, and completing high

school without knowing them will make it difficult for the learner to succeed in the stages beyond high school, especially in academic settings.

Table 2

Multivariate Tests (Lexical Sophistication Level and Timepoints)

Multivariate Tests							
	Effect	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Lexical sophistication level	Pillai's Trace	.833	876.923 ^b	3.000	528.000	.000	.833
	Wilks' Lambda	.167	876.923 ^b	3.000	528.000	.000	.833
	Hotelling's Trace	4.983	876.923 ^b	3.000	528.000	.000	.833
	Roy's Largest Root	4.983	876.923 ^b	3.000	528.000	.000	.833
Lexical sophistication level * timepoint	Pillai's Trace	.003	.544 ^b	3.000	528.000	.652	.003
	Wilks' Lambda	.997	.544 ^b	3.000	528.000	.652	.003
	Hotelling's Trace	.003	.544 ^b	3.000	528.000	.652	.003
	Roy's Largest Root	.003	.544 ^b	3.000	528.000	.652	.003

As can be seen in Table 2, the effect of the level of lexical sophistication on the percentage of correct answers is statistically very significant ($p < .01$; .000), in contrast to the effect of the time point ($p > .05$; .652). That is to say, the main factor that affected the students' success in the task was the level of lexical sophistication of the words they were asked to translate and put into sentences and not what grade they were in when they performed this task.

These findings show that the teachers' vast experience teaching Hebrew as L2 to Arabic speakers enables them to differentiate between words belonging to different levels of difficulty, and to rate them on a scale according to the difficulty levels as relevant to Arab students studying Hebrew as L2. These ratings were not a basic dichotomic division between basic words and advanced words, but a more complex rating that included four levels of difficulty. The teachers who rated the words for this study were not trained in classifying words according to the level of difficulty, and had never done so in the past. Despite this, statistically significant differences were found in the students' achievement levels with the different levels of lexical sophistication which matched the teachers' evaluations. This shows that using teachers' evaluations for determining the difficulty or lexical sophistication level of words is a valid tool. This proof, the first of its kind within studies on Hebrew vocabulary, was found in previous studies as well, such as the one conducted on Turkish as L2 (Daller et al., 2003) and on French as L2 (Tidball & Treffers-Daller, 2008).

The Relationship Between Lexical Sophistication and Increase in Acquisition Level

RQ2

This question tested whether there is a statistically significant difference between the students' level of achievement on the two tasks between the beginning of the research (11th grade) and its end (12th grade). Since the students learned many hours of Hebrew throughout the year, and also heard and experienced it in both formal and informal settings, I expected an increase in lexical sophistication over time. This expectation was refuted, as results show no statistically significant differences between the two points in time based on the number of correct responses given by students in both grades (chi-square test value: $p > .05$; Mann Whitney test: $p > .05$) (see Table 3). There is a slight increase in the number of correct responses: In 11th grade 438 out of the 1064 responses were correct, whereas in 12th grade 453 out of 1064 responses were correct, but this increase is not statistically significant.

Table 3
Lexical Sophistication Level over Time

Answer	Grade		Total	Chi-Square Test
	11th grade	12th grade		Test value
Correct	438	453	891	0.434
Incorrect	626	611	1237	P-value
Total	1064	1064	2128	0.510
Answer	Grade	N	Mean Rank	Mann-Whitney test value
	11th grade	1064	1072.00	558068
	12th grade	1064	1057.00	P-value
	Total	2352	2128	0.510

The lack of statistically significant increase in the students' level of lexical sophistication can be explained by the fact that one year of instruction is likely not long enough to significantly develop the vocabulary of the Arabic-speaking learners of Hebrew. A study which examined English as L2 for Kannada speakers found a significant difference in passive vocabulary—as opposed to the controlled active vocabulary—between students in the 8th and 12th grades (Nemati, 2010). This difference reached a level of significance only after five

years of instruction, and not after one year as in the present study. A different study evaluated, among others, the development of controlled active vocabulary among Canadian students of English as L2. Even after two years of instruction in an L2 learning environment, no statistically significant improvement in controlled active vocabulary was found. The conclusion is that sometimes it takes more than two years to achieve such an increase (Laufer & Paribakht, 1998).

The limited number of words used in the study may have played a role in the lack of noticeable significant development. It included a small part of the learners' vocabulary and an even smaller fraction of all the words in the language. The current study used 28 words. However, even studies which included a higher number of words did not find a statistically significant increase in some lexical aspects. For instance, one study on passive vocabulary which examined 156 words and a controlled active vocabulary of 90 words, among 100 8th–12th grade English as L2 students in India, did not find any statistically significant difference between each class and the class one year above it (including 11th and 12th grades), neither in passive vocabulary nor in controlled active vocabulary (Nemati, 2010).

The present study employed a limited number of words for technical reasons over which I, as a researcher, had no control. Reasons included the busy schedule of the class tested, given the preparations for the matriculation exams, as well as delays and changes in school scheduling which, of course, influenced the times when the tasks were given to the students as well as the amount of time allotted for work on these tasks.

Differences and Correlations Between Passive Vocabulary and Active Vocabulary

Differences Between Passive and Active Vocabulary

RQ3-a

The first part of the third research question looked for differences between the two tasks, each of which evaluates a different type of vocabulary: The first task, writing a definition for each word, evaluated the respondents' passive vocabulary, while the second task evaluated their controlled active vocabulary by having them insert the words into sentences. The differences between the two types of vocabulary were examined from two aspects: level of breadth and the extent of improvement over time.

Table 4
Differences Between Passive and Active Vocabulary

Answer	Task		Total	Chi-Square Test
	Writing a definition (Passive vocabulary)	Inserting into a new sen- tence (Active vocabulary)		Test value
Correct	472	419	891	5.423
Incorrect	592	645	1237	P-value
Total	1064	1064	2128	0.020

The findings do not fully support the expectation for a statistically significant difference between the two types of lexicons, given the level of difficulty. The passive vocabulary was found to be significantly broader than the controlled active vocabulary only in the lower (11th) grade (chi-square test value: $p < .05$) (see Table 4). In this class 239 correct responses were given on the word-definition task and only 199 correct responses on the word-insertion task. The highest level of controlled active vocabulary was anticipated, as its expression necessitates the learner's mastery of many linguistic skills, including lexicality, grammar, syntax, and more, and not just lexicality as in the passive vocabulary.

Table 5
Differences Between Passive and Active Vocabulary

Grade			Task		Total	Chi-Square Tests
	Answer	Correct	Writing a defi- nition (Passive vocabulary)	Inserting into a new sentence (Active vocabulary)		Test value
11th grade	Answer	Correct	239	199	438	6.209
		Incorrect	293	333	626	P-value
	Total		532	532	1064	0.013
12th grade	Answer	Correct	233	220	453	Chi-Square Tests
		Incorrect	299	312	611	.650
	Total		532	532	1064	P-value 0.420

Conventional wisdom in the field of vocabulary research is that those who learn a language know more words than they can use (Fan, 2000). This thesis was supported by many previous studies (Laufer & Paribakht, 1998; Nemati, 2010; Hsu, 2014) as well as by the present study. According to this finding, the number of words which learners of Hebrew as L2 to Arabic identify when they read (or hear) them is significantly larger than the number of words available for their active use in a writing task in 11th grade (see Table 5).

In 12th grade the gap between the passive vocabulary and the controlled active vocabulary was greatly reduced. In this class 233 correct responses were given on the word-definition task and 220 correct responses on the word-insertion task. In addition, a correlation exists between the two types of vocabulary in 11th grade, whereas no such correlation exists in the 12th grade subjects. The gap between the number of correct responses between the two types of vocabulary was 40 in the 11th grade, but only 13 in the 12th grade. The decrease in this gap stemmed from the learners' improved performance only on the word-insertion task (from 199 correct responses in 11th grade to 220 in 12th grade), whereas there was no improvement in the word-definition task, and the difference between the two classes was insignificant. This finding does not contradict the previous finding, which proves the basic assumption in lexicon research that the passive vocabulary is greater than the active one. This finding only limits this assumption, and shows that it is not fixed within all levels of proficiency and may change with the increase in acquisition time.

There was an expectation that there would be no gap in the level of improvement between the two types of vocabulary, since in the instruction process, which is based on the Hebrew as L2 to Arabic learning curriculum, there is no focus on either type of vocabulary. This expectation was refuted, as findings revealed a prominent improvement in controlled passive vocabulary as opposed to the improvement in the passive vocabulary. The reason for this is probably that students throughout the school year had many opportunities for actively using their vocabulary, not only during Hebrew lessons but also in various formal and informal activities, including field trips and visits to Hebrew-speaking schools. Actively practicing the use of words they knew made it easier for them to transfer words from passive to active use. Therefore, no improvement was recorded in their passive vocabulary, which in any case was considered easy and basic, but there was considerable improvement in their active vocabulary.

Identical findings were documented in Nemati's (2010) research of 8th—12th grade Indian students of English as L2. The gap between passive and controlled active vocabularies in the higher grades was found to be smaller than in the lower grades, probably because the older students used English for communication purposes more than the younger ones. The study was carried out in a school where the language of instruction was English, the students' second language. In the present study, although Hebrew was not the language

of instruction in all subjects in the school, except for Hebrew, Hebrew is a second language for the students and is present in a variety of formal and informal contexts. This creates many opportunities for the students to actively express their passive knowledge. In a different study of L2 acquisition within a foreign environment, the gap between the breadth of the two types of vocabulary was high, as the lack of exposure to L2 in the foreign environment and the limited number of opportunities to actively use it delay the transfer of words that the learner knows (passive vocabulary) to active use (Hsu, 2014). The increase in acquisition does not always lead to a narrowing of the gap between the two types of vocabulary, as the relationships between the two types of lexicons are more complex than they appear (Fan, 2000).

Correlation between Passive and Active Vocabulary

RQ3-b

This research question examines the relationship between the two types of vocabulary within each of the time periods separately, and in both together. I expected a correlation between the two tasks, as each of them reflects one type of vocabulary. Results show a very significant correlation between the word-definition task and the word-insertion task (Pearson correlations = .744, $p < 0.01$). On the assumption that the first task reflects a passive vocabulary whereas the second a controlled active vocabulary, there a statistically significant correlation exists between the passive and controlled active vocabularies. This means that when the learner's performance increases in one of the two tasks (word-definition or word-insertion), then it increases in the other task as well. This relationship was recorded in many studies (Laufer, 1998; Nemati, 2010).

Furthermore, I expected an increase in this correlation within the second time period (12th grade) as the instruction process continues, together with the anticipated increase in the students' linguistic abilities resulting from the many opportunities for active use of their passive vocabulary. In addition, this expectation of an increase in correlation is supported by the finding of the previous research question, wherein the gap between passive vocabulary and controlled active vocabulary was greatly decreased.

Conclusion and Practical Implications

This research was the first to examine lexical sophistication in Hebrew as L2 to Arabic speakers (or in Hebrew in general). Given the many limitations imposed by testing lexical sophistication using word frequency lists, this study employed a new and more suitable method, which was based on teachers' evaluations of the level of lexical sophistication of words, as is acceptable by studies of this lexical index to date.

Results showed that among Arabic-speaking high-school students learning Hebrew as L2, only the lower grade (11th) had a broader passive vocabulary than controlled active vocabulary, but with acquisition time, the gap between these two vocabularies diminished. Decreasing the gap was a result of an increase in the students' performance on the controlled active vocabulary task. The increase in the students' ability to actively use their passive vocabulary is probably the result of the many opportunities they had, throughout the school year, to actively use the vocabulary they acquired within formal and informal settings. In addition, a highly significant correlation was found between the passive vocabulary and the controlled active vocabulary, which is somewhat strengthened as acquisition time increases.

The results did not yield any statistically significant increase in the level of lexical sophistication after one year of Hebrew instruction in high school. Two explanations were offered for this finding. First, that one year of instruction is not sufficient for producing an increase in lexical sophistication. The second explanation has to do with the limited number of words used in studies of passive vocabulary.

Results showed very significant differences between the different levels of lexical sophistication. These differences were even greater between the first two levels of lexical sophistication (basic words) and the second two levels (advanced words), which were based on teachers' evaluations of the words' lexical sophistication. Furthermore, the findings indicated that as the level of lexical sophistication increased, the performance of the students decreased. These findings prove that the teachers' evaluations serve as a valid index for determining the level of sophistication of words, as found in studies on other languages.

Instruction of vocabulary should be planned according to the levels of lexical sophistication, or the difficulty level of the words. Skilled teachers with many years of teaching experience can help construct lists of words classified by difficulty. Additionally, words from each of the difficulty levels can be incorporated with each age group of instruction. As age increases, more words from the advanced groups of words should be incorporated.

Ultimately, educators ought to select texts that align with the proficiency levels of their students. These instructional materials can include words from different levels of difficulty: On the basic level—more basic than advanced words; and on the higher levels—more advanced than basic words. A gradual increase in the level of lexical sophistication makes the process of vocabulary development more efficient, and ensures progress on the well-established axis of language acquisition—from the simple to the complex.

References

- Abu-Rabiah, E. (2020). Lexical measures for testing progress in Hebrew as Arab students' L2. *Journal of Language and Linguistic Studies*, 16(3), 1096–1114. <https://doi.org/10.17263/jlls.803551>
- Abu-Rabiah, E. (2022). *Acquisition of Hebrew lexicon among Negev Bedouin high school students: An empirical account* [Unpublished Doctoral dissertation]. Ben-Gurion University of the Negev. [in Hebrew]. <http://dx.doi.org/10.13140/RG.2.2.17502.51528>
- Abu-Rabiah, E. (2023). Evaluating L2 vocabulary development features using lexical density and lexical diversity measures. *LLT Journal: A Journal on Language and Language Learning*, 26(1), 168–182. <https://doi.org/10.24071/llt.v26i1.5841>
- Abu-Rabiah, E., Gafter, R., & Henkin, R. (2023). Where syntactic interference persists: The case of Hebrew written by native Arabic speakers. *L1-Educational Studies in Language and Literature*, 23, 1–19. <https://doi.org/10.21248/l1esl.2023.23.1.399>
- Astridya, F. W. (2018). Lexical richness of the expository writing in Indonesian by senior high school students. *Lingual: Journal of Language and Culture*, 5(1), 23–29. <https://doi.org/10.24843/LJLC.2018.v05.i01.p04>
- Azodi, N., Karimi, F., & Vaezi, R. (2014). Measuring the lexical richness of productive vocabulary in Iranian EFL university students' writing performance. *Theory & Practice in Language Studies*, 4(9), 1837–1849. <https://doi.org/10.4304/tpls.4.9.1837-1849>
- Cobb, T., & Horst, M. (2004). Is there room for an academic word list in French? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 15–38). John Benjamins.
- Cook, V. (2016). *Second language learning and language teaching*. Routledge.
- Crossley, S. A., Cobb, T., & McNamara, D. S. (2013). Comparing count-based and band-based indices of word frequency: Implications for active vocabulary research and pedagogical applications. *System*, 41(4), 965–981. <https://doi.org/10.1016/j.system.2013.08.002>
- Daller, H., Van Hout, R., & Treffers-Daller, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics*, 24(2), 197–222. <https://doi.org/10.1093/applin/24.2.197>
- Ellis, N. C. (1997). Vocabulary acquisition: Word structure, collocation, word-class, and meaning. In M. McCarthy & N. Schmidt (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 122–139). Cambridge University Press.
- Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2), 143–188. <https://doi.org/10.1017/S0272263102002024>
- Ellis, R. (2015). *Understanding second language acquisition*. Oxford University Press.

- Fan, M. (2000). How big is the gap and how to narrow it? An investigation into the active and passive vocabulary knowledge of L2 learners. *Relc Journal*, 31(2), 105–119.
- Henriksen, B., & Danelund, L. (2015). Studies of Danish L2 learners' vocabulary knowledge and the lexical richness of their written production in English. In P. Pietilä, K. Doró, & R. Pípalová (Eds.), *Lexical issues in L2 writing* (pp. 29–56). Cambridge Scholars Publishing.
- Horst, M., & Collins, L. (2006). From *faible* to strong: How does their vocabulary grow?. *Canadian Modern Language Review*, 63(1), 83–106. <https://doi.org/10.3138/cmlr.63.1.83>
- Hsu, W. (2014). The effects of audiovisual support on EFL learners' productive vocabulary. *ReCALL*, 26(1), 62–79. <https://doi.org/10.1017/S0958344013000220>
- Jarvis, S. (2017). Grounding lexical diversity in human judgments. *Language Testing*, 34(4), 537–553. <https://doi.org/10.1177/0265532217710632>
- Kojima, M., & Yamashita, J. (2014). Reliability of lexical richness measures based on word lists in short second language productions. *System*, 42, 23–33. <https://doi.org/10.1016/j.system.2013.10.019>
- Kwon, S.-H. (2009). Lexical richness in L2 writing: How much vocabulary do L2 learners need to use? *English Teaching*, 64(3), 155–174.
- Kyle, K., & Crossley, S. (2016). The relationship between lexical sophistication and independent and source-based writing. *Journal of Second Language Writing*, 34, 12–24. <https://doi.org/10.1016/j.jslw.2016.10.003>
- Laufer, B. (1995). Beyond 2000: A measure of productive lexicon in a second language. In L. Eubank, L. Selinker, & M. Sharwood Smith (Eds.), *The current state of interlanguage: Studies in honor of William E. Rutherford* (pp. 265–272). John Benjamins.
- Laufer, B. (1998). The development of passive and active vocabulary in a second language: Same or different?. *Applied Linguistics*, 19(2), 255–271. <https://doi.org/10.1093/applin/19.2.255>
- Laufer, B. (2005). Lexical frequency profiles: From Monte Carlo to the real world: A response to Meara. *Applied Linguistics*, 26(4), 582–588. <https://doi.org/10.1093/applin/ami029>
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307–322. <https://doi.org/10.1093/applin/16.3.307>
- Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, 16(1), 33–51. <https://doi.org/10.1177/026553229901600103>
- Laufer, B., & Paribakht, T. S. (1998). The relationship between passive and active vocabularies: Effects of language learning context. *Language Learning*, 48(3), 365–391. <https://doi.org/10.1111/0023-8333.00046>
- Lindqvist, C., Bardel, C., & Gudmundson, A. (2011). Lexical richness in the advanced learner's oral production of French and Italian L2. *International Review of Applied Linguistics in Teaching*, 49, 221–240. <https://doi.org/10.1515/iral.2011.013>
- Linnarud, M. (1986). *Lexis in composition: A performance analysis of Swedish learners' written English*. Liber Forlag.
- Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Palgrave Macmillan. <https://doi.org/10.1057/9780230511804>
- Meara, P. & Bell, H. (2001). P-Lex: A simple and effective way of describing the lexical characteristics of short L2 texts. *Prospect*, 16(3), 323–337.
- Meara, P. (2001). *PLEX v1.1*. University of Wales Swansea. <http://www.swan.ac.uk/cals/calsres>.
- Milton, J. (2007). Lexical profiles, learning styles and the construct validity of lexical size tests. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 47–58). Cambridge University Press.
- Milton, J. (2009). *Measuring second language vocabulary acquisition*. Multilingual Matters. <https://doi.org/10.21832/9781847692092>

- Nation P. (1990). *Learning and teaching vocabulary*. Newbury House.
- Nation, I. (2006). How large a vocabulary is needed for reading and listening?. *Canadian Modern Language Review*, 63(1), 59–82. <https://doi.org/10.3138/cmlr.63.1.59>
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press. <https://doi.org/10.1017/9781009093873>
- Nemati, A. (2010). Active and passive vocabulary knowledge: The effect of years of instruction. *The Asian EFL Journal Quarterly*, 12(1), 30–46.
- Read, J. (2000). *Assessing vocabulary*. Cambridge University Press.
- Shimron, J., & Sivan, T. (1994). Reading proficiency and orthography evidence from Hebrew and English. *Language Learning*, 44(1), 5–27.
- Tidball, F. and Treffers-Daller, J. (2008). Analysing lexical richness in French learner language: What frequency lists and teacher judgement can tell us about basic and advanced words. *Journal of French Language Studies*, 18(3), 299–313. <https://doi.org/10.1017/S0959269508003463>
- Vermeer, A. (2000). Coming to grips with lexical richness in spontaneous speech data. *Language Testing*, 17(1), 65–83. <https://psycnet.apa.org/doi/10.1191/026553200676636328>
- Waldvogel, D. A. (2014). An analysis of Spanish L2 lexical richness. *Academic Exchange Quarterly*, 18(2), 1–8.
- Xue, G., & Nation, I. S. P. (1984). A university word list. *Language Learning and Communication*, 3(2), 215–229.