Lexical measures for testing progress in Hebrew as Arab students' L2

Eihab Abu Rabiah [a] [1]

[a] Ben-Gurion University of the Negev, Beer-Sheva, Israel

Abstract

Since Ure (1971) and Halliday (1985), different measures have been proposed in the research literature as a proxy for investigating lexicon advances. They are called lexical richness measures and are an important tool in assessing advance of the lexicon (Johansson, 2009). This article evaluates three lexical richness measures of progress in the acquisition of Hebrew by Arab students in Israel: level of abstractness, lexical diversity, and density lexical. The corpus consisted of sixty essays written in Hebrew by 9th and 11th graders. The quantitative analysis reveals several findings. First, abstractness increases with subjects' age, as expected. The density test, however, did not show the expected increase in density with age. A likely reason for this is relatively high density in younger pupils' essays, indicating lack of cohesion rather than high linguistic competence. The national high school language curriculum for eleventh graders focuses on connectivity, which causes a decrease in density. Third, the diversity test also did not yield statistically significant findings, probably because this test is influenced by text length, which in the ninth grade still does not enable a reliable test. In conclusion, the lexical measures can be divided into two types: basic tests, that can be applied at any level of competence; and advanced tests, which require a high level of competence. Level of abstractness is of the first type: it is effective as a measure of progress in L2 acquisition even in the initial stages of acquisition, while diversity and density require a higher competence level.

Keywords: second language acquisition; lexicon development; assessing writing; level of abstractness; lexical diversity; lexical density

1. Introduction

Measures of lexical richness are essential tools in assessing the acquisition of the lexicon and are often used in L2 research (Johansson, 2009). Ever since (Ure, 1971; Halliday, 1985), different measures have been proposed in the literature as a proxy for investigating how L2 learners' lexicon advances. The assessment of lexical richness of texts is one of the major dilemmas in the field of quantitative linguistics (Tweedie & Baayen, 1998). Therefore, many measures were developed to evaluate the lexicon. Measures such as; lexical density, lexical diversity, and lexical sophistications (Webb, 2019).

---

[1] Corresponding author.
  E-mail address: aehab@post.bgu.ac.il

This paper focuses on three such measures to measure lexical acquisition of Hebrew by Arabic-speaking high school students: level of abstractness, lexical diversity and lexical density. All three measures were used to examine a corpus of Hebrew essays by 9th graders and 11th graders in a Negev Bedouin high school- southern Israel. The purpose of this study is twofold: firstly, to determine how performance on each measure correlates with length of acquisition. Next it is discussed whether they are equally usable at these two stages of acquiring Hebrew as L2 to Arabic.

The three lexical measures in focus have been successfully implemented in studying many pairs of languages - but so far not Hebrew as an L2 for Arabic speakers. As the lexical measures were developed for analytic languages such as English, this paper discusses ways of applying some adaptations in order to adjust them to more synthetic languages like Hebrew and Arabic. The results have broader implications for the use and interpretation of such measures, beyond the pairing of Hebrew and Arabic. These theoretical and methodological innovations will hopefully benefit research on acquisition of Hebrew, L2 acquisition, and language acquisition in general.

1.1. Literature review

1.1.1. Contact between Hebrew and Arabic in Israel

The Arab citizens of Israel, who account for roughly 21% of Israel's population (Central Bureau of Statistics, 2019), are in constant contact with Hebrew, the language of the Jewish majority. Although both Hebrew and Arabic were official until 2018, Hebrew is the dominant language in the country (Henkin, 2011). Most of the work of the governmental Israeli authorities is conducted in Hebrew, which is therefore perceived as more important and prestigious than Arabic, even among Arab speakers in Israel (Amara, 1999; Amara and Mar'i, 2002). Currently, Israel has two parallel education systems: one that is conducted in Hebrew, and another, which serves the non-Jewish population, in which the language of instruction is Arabic, and Hebrew is taught as a required foreign language (Mor-Sommerfeld et al., 2007).

Demographic and sociolinguistic variables determining the acquisition of Hebrew as second language (L2) of Arabic speakers in different locations across Israel include degree of exposure to Hebrew, and hence degree of Hebrew influence on Arabic, and the attitude of Arabic speakers to Hebrew. The degree of contact and effect of Hebrew on Arabic are most intensive in mixed Jewish-Arab cities and decrease in remote villages, such as those of the Galilee (Saiegh-Haddad, 2003, p. 432; Henkin, 2011, p. 78). The level of Hebrew in the ninth grade in the Negev is not necessarily the same as it is in other regions, hence the need to examine issues in acquiring Hebrew in different locations, subject to factors such as linguistic contact, exposure, and attitude to Hebrew.

1.1.2. Lexical acquisition in a second language

The importance of investigating L2 acquisition processes has long been recognized (MacWhinney, 2005; Zhang, 2005; Lefebvre et al., 2006; Dimroth, 2008; Flynn & O'Neil, 2015; Nassaji, 2020). This study focuses on a specific aspect of L2 acquisition – lexical acquisition. Numerous studies (e.g., Qian, 2004; Jiang, 2007; Mondria & Wiersma, 2007; Nation, 2007; David, 2008; Fitzpatrick et al., 2008; Hilton, 2008; Lu, 2012; Kojima & Yamashita, 2014; Vedder & Benigno, 2016; Ha, 2019) have attempted to quantify the development of vocabulary according to lexical measures. These studies were conducted mainly on pairs of English and another language, generally one that also uses Latin script. This study aims to expand this research by investigating the effectiveness of three measures for assessing progress in the acquisition of Hebrew as L2 for Arabic speakers in Israel.

Lexical richness, lexical skill, lexical depth, and lexical competence are all terms for the same ability – a command of the various expressions of the lexicon, such as the use of the correct word in a specific context. The term lexical competence will used herein. Lexical competence combines two types of

competence: inter-lexical competence, that is, the ability to choose the correct word from a group of similar words, and intra-lexical competence, that is, the ability to use one word in all of the correct contexts (Tanaka, 2012, p. 1).

The acquisition of a lexical value in L1 and L2 is more complicated than simply remembering and interpreting a word. It involves identifying a vocal pattern and matching it to the orthography (in the case of acquisition of written language), precise pronunciation, and semantic precision, as well as matching it to the conceptual and contextual content (Bogaards, 2001; Nation, 2005).

L2 speakers tend to transfer semantic relationships from L1 to L2 in a process of semantic interference (Ferré et al., 2006; Bürki et al., 2020) and, similarly, they transfer lexical properties from L1 to L2 in a process of lexical interference (Prior et al., 2017). Many words are similar in both sound and meaning in the closely related Semitic languages of Arabic and Hebrew. However, many words which may sound similar in the two languages have different meanings altogether, which constitutes a problem for learners. For example, Arabic speakers sometimes use the word leḥem 'bread' in place of 'meat' in Hebrew, because in Arabic laḥm means 'meat' (basar in Hebrew). A recent study has shown that lexical interference of this type is very common among students and, unlike syntactic interference, does not diminish with increased Hebrew proficiency (Prior et al., 2017).

The lexical development of L2 was found to be influenced by the lexical development in L1 (Crossley et al., 2009). In both, development depends, among other things, on the time and amount of exposure, and is measured in a variety of shared measures such as abstractness, lexical diversity, lexical density, lexical accuracy or sophistication (the most accurate word for a given context), and the number of errors. All of these are important for evaluating the development of the lexicons of L1 and L2 and progress in language acquisition. Moreover, lexical competence in L2 has been associated with academic achievement in general (Polio, 2001; Daller et al., 2003), so lexical development tests also contribute to predicting success in other areas and to monitoring general cognitive development.

This paper examines and attempts to assess the effectiveness of the three measures: level of abstractness, lexical diversity, and lexical density.

1.1.3. The measures examined for lexical development

Building solid vocabulary is an essential part of successful second language learning (Zheng, 2016), reading ability, communicative ability and academic achievement (August & Shanahan, 2006; Crossley & Skalicky, 2019). Assessing vocabulary knowledge is a difficult task, therefore different lexical measures have been developed and are used in many studies to evaluate vocabulary progress, for instance (Zhai, 2016; Juanggo, 2018; Gharibi & Boers, 2019; Wood et al., 2019; Borràs & Llanes, 2020; Ebedy, 2020). The assessment of the lexical richness of texts is one of the major dilemmas in the field of quantitative linguistics (Tweedie & Baayen, 1998). Therefore, many measures were developed to evaluate the lexicon. The following are the main lexical measures: lexical diversity (comparing the number of different words (types) with the total number of words (tokens) in a text), lexical variation (equal to the lexical diversity but focused only on lexical words), lexical density (the proportion of "content" or "lexical" words in a text), lexical sophistication (the proportion of "advanced words" in a text) and lexical individuality (the proportion of words used by only one participant in a group) (Šišková, 2012).

1.1.3.1. Level of abstractness

Put simply, a concrete noun is a noun that can be perceived by one of the five senses, whereas an abstract noun is one that cannot be perceived by any of the five senses but is, rather, perceived in thought (Paivio et al., 1968; Mayer et al., 2017). Unlike abstract words, concrete words do not have an

identifiable, edged and clearly perceptible referent (Borghi et al., 2017). Therefore, different types of abstract nouns should be distinguished (Borghi et al., 2018).

Abstract thought is one of the ratifications of human cognition (Borghi et al., 2018). Ravid (2006) found that children use mainly concrete, prototypical, and unambiguous nouns, and as they get older, at a later stage of language acquisition, they use more abstract, metaphorical and ambiguous nouns. She also notes that the older the child, the greater the use of nouns in general, and the greater the use of abstract nouns (Ravid 2006: 794) and the metaphorical use of concrete nouns (Ravid, 2005, p. 337).

L2 speakers also acquire common words, which are typically concrete, more easily than rare and abstract words in the target language. The longer the exposure to L2, the more frequent the use of abstract language (Tanaka, 2012). Thus, the abstractness level of L2, as well as of L1, in a given corpus serves as an important measure of lexical development. Therefore, the level of abstraction is the first measure tested in this study. The level of abstraction is defined here as the proportion of abstract nouns relative to all the nouns in the text tokens.

1.1.3.2. Lexical diversity

The second measure examined in this study is lexical diversity. A speaker displaying high lexical diversity is judged positively, as able to convey complex messages (Dillard and Pfau, 2002). Therefore, measuring this capacity as a measure of linguistic and communicative ability is highly motivated.

Wendell Johnson (1939, 1944) was the first to introduce the type-token ratio (TTR), a measure that examines the lexical diversity of a text by counting the types (distinct words) in relation to the overall number of tokens of words The smaller the number of lexical repetitions in a given text, the greater the text's lexical diversity (Gebril & Plakans, 2016; Jarvis, 2017). High lexical diversity indicates that the speaker or writer has a rich vocabulary, by virtue of which he does not need to use the same types again and again (Johansson, 2009, p. 62).

Nevertheless, the TTR index has a systematic disadvantage – it decreases as the length of the text increases (McCarthy & Jarvis, 2010; Jarvis, 2012, 2013, p. 91; Kyle, 2019). The explanation for this negative correlation is that our types are limited, while tokens are not, since there are a limited number of word types available on a given subject, but can write or say infinite word tokens on the same subject and may repeat certain types due to pragmatic needs such as emphasizing an idea. Moreover, since length of written texts increases with the age of the writer, TTR decreases accordingly (Johansson, 2009, p. 63). In sum, it is recommended not to use the TTR index as a single measure of lexical competence in texts of different lengths.

The "theoretical vocabulary approach" attempts to neutralize the effect of text length on TTR. This approach proposes randomly selecting several consecutive or non-consecutive segments from the text totaling a certain number of words (e.g., 100 tokens) and then sorting them into the various types and calculating the lexical diversity. The diversity measured in a given segment is considered to represent the rest of the text, regardless of its length. Thus, texts of different lengths can be measured without length being an influencing factor (Johansson, 2009, p. 65). Another solution to overcoming the effect of length (Jarvis, 2013) is to divide long texts into samples of fifty tokens and determine the TTRs of each sample and then calculate the mean TTR of all samples in all texts.

All the measures of lexical diversity were found to be affected in one way or another by the length of texts ranging from 50 to 200 tokens, but the effect is reduced when text length ranges from 100 to 200 tokens (Koizumi, 2012; Koizumi and In'nami, 2012). It is therefore recommended to use lexical diversity measures (including the TTR index) in texts of 100 to 200 tokens. Indeed, researchers have tended to use texts of approximately 100 tokens in L1 and L2 spoken and written texts for both children and adults (Malvern & Richards, 2002; Johansson, 2009). Although longer L1 and L2 texts have also

been examined (for example by Hess et al., 1986; Choi & Jeong, 2016). In accordance with these recommendations, the current study was conducted on texts of 100 tokens wherever possible.

1.1.3.3. Lexical density

The third measure examined is lexical density, which is the ratio between content words and function words in a given text. Content words are lexemes such as nouns, verbs and adjectives that contain semantic or lexical value (Demir-Vegter et al., 2014). This is an open class that constitutes most of the vocabulary in a language and expands in different ways (Abdalla, 2012), such as internal derivation and borrowing from a foreign language.

Function words, on the other hand, are a small closed class with just a grammatical value. This category includes prepositions, conjunctions, pronouns, auxiliary verbs, inflections and interjections (Juste et al., 2012).

A lexically dense text, with a relatively high ratio of content verbs, contains more lexical information than a text with lower density or more function words (Johansson, 2009, p. 65), and lexical density is widely considered a mark of lexical competence. But this is not a straightforward relation. In L1 acquisition, content words precede function words. Children often say and write sentences without any function words at all. As they get older, they learn function words and their importance in linking content words and begin to use them as well (Johansson, 2009, p. 66).

Not only children's language, but simplified registers in general, including foreigner talk and pidgins, are often lacking in function words, rendering them very dense. Therefore, a rise in lexical density is not necessarily a mark of linguistic competence - language first becomes less dense following the acquisition of a variety of function words, and only later does lexical density again increase with the development of literacy and competence in the use of academic registers. It is this later stage that Halliday (1989) characterizes, where spoken language is less lexically dense than written language. In other words, on average, spoken language has fewer content words relative to all the words in the text than written language does. Ure (1971) believes that the lexical density in most spoken texts is less than 40%, whereas in written texts it is greater than 40%. Halliday (1989, p. 62) also explicitly states that lexical density is the main difference between written language and spoken language.

1.2. Research question

The research question is: how effectively do the three lexical measures (level of abstractness, lexical diversity, and lexical density) measure the progress in the Hebrew interlanguage of Arabic speakers between the two groups: 9th and 11th graders?

2. Method

This study involved 60 high school students who were divided into two age groups (ninth grade and 11th grade). The students watched a relevant video for their community and subsequently wrote an expository essay on women's work topic. The essays were analysed according to the three lexical measures: level of abstractness, lexical diversity and lexical density.

2.1. Participants and corpus

This study was conducted on 60 compositions of 30 ninth-grade and 30 eleventh-grade Arab high school students in the Negev, in southern Israel. The youngest students were 14 years old and the older students were 16. The number of boys and girls was similar in each age group as well as between the

two age groups. The general level of achievement of the students is similar, because all the students were admitted to the school after passing the entrance exams. While not all the students live in the village where the school is located they do live with in the Negev region in southern Israel.

2.2. Data collection and analysis

The participants watched a video[2] that contrasted housewives with women who work full-time with respect to the question: which group is happier in their lives?. The students were then asked to write a composition expressing their views on the subject of women's work. There was no time limit, and no limit on the length of the composition. The teacher ensured that no help was provided. Ninth graders were initially expected to write nearly a hundred words, and eleventh graders to write more than a hundred words, which was indeed the case.

2.2.1. Calculations of lexical measures

The three tests were performed on each composition, beginning with the level of abstractness. Following Ravid (2005), The level of abstractness in a text was defined as the percentage of abstract noun tokens out of the total number of nouns. For example, given five abstract nouns in a text totaling ten nouns, the level of abstractness in the text is 50%. A noun is considered concrete if it could be perceived by one of the five senses, e.g. *kise* 'chair'; otherwise it was abstract in the binary classification.

Second, lexical diversity was determined using the TTR measure, that is, the percentage of word types out of the total of word tokens.

Third, lexical density was calculated using the most common method, namely dividing the number of content word tokens by the number of tokens in the entire text and calculating this ratio as a percentage (Johansson, 2009, p. 65).[3]

As seen above, given the effect of text length on the lexical diversity test, many studies recommend that this test be performed on texts of 100 to 200 tokens. Thus, it was decided to check all three on the first 100 tokens in each composition.

2.2.2. Considerations in counting words

A major consideration in counting words in the corpus relates to the orthographic uniqueness of the Hebrew language relative to the languages in which such studies have been conducted. In Hebrew (and in Arabic) only some of the function words are separate words in writing, while the rest are written attached to content words. Therefore, there are a large number of orthographic words that include both a function word and a content word. As a result, 100 orthographic words in Hebrew typically include more than 100 words in English. Consider the following sentence: *halaxnu le-betenu she-ba-kfar* 'We went to our house which is in the village'. This sentence contains 3 orthographic words in Hebrew, and 10 words in English.

Therefore, the question of what counts as a 'word' is crucial when adapting these lexical measures to Hebrew, but its importance may differ in the different tests. Different methods were chosen in accordance with the needs of the test. In a lexical density test that is based entirely on the difference between content and function, it was important to separate textual units composed of content and function words into their components and to count each element separately. For example, in the phrase *ve-ha-ne'arim halxu* 'and the boys went', which includes only two orthographic words, four elements

---

[2] https://www.youtube.com/watch?v=4C4qkDHUjFQ

[3] There are other methods of testing for density. First, "nominal density," which consists of the number of noun tokens divided by the total tokens in the text (Johansson, 2009, p. 65). Second, verb, adjective, and adverb tokens are divided by the total tokens in the text (Johansson, 2009, p. 65). Neither of these methods has been successful among lexical density researchers, hence the conventional method was used.

were counted: two content words *ne'arim* 'boys' and *halxu* 'went', the conjunction *ve* 'and' and the definite article prefix *ha-* 'the'. [4]

On the other hand, in the tests of the level of abstractness and lexical diversity an orthographic word composed of a content word and a function word as one word was retained. In the level of abstractness test, function words are simply relevant - the test addresses only nouns. In the lexical diversity test, the reason for disregarding certain function words is that they were expected to be distributed homogeneously, attaching to content words with the same average frequency, regardless of the writer's lexical competence. It is therefore possible to ignore them and focus on clear lexical elements such as content words and independent function words. An orthographic word composed of a content word preceded by a function word was counted as a single content word: *ba-batim* 'in the houses' was considered a token of the type *bayit* 'house'.

A major issue that has not yet been satisfactorily solved was the classification of certain word classes as either content words or function words. Nouns, verbs, and adjectives were classified, as is customary in Hebrew linguistics, as content words. The classification of adverbs is far less straightforward, and is an open debate in the linguistic literature. However, for the purpose of measuring lexical density, it is generally accepted that it not so important where the line between lexical and grammatical adverbs is drawn, as long as the classification is consistent (Johansson, 2009, p. 67). In Hebrew, the problem is known to be particularly challenging and there is no general agreement regarding the characterization of adverbs, therefore consistency was maintained without attempting to justify the boundary set. Adverbs were counted as content words or function words according to their semantic roles: *mamash* 'really' is a function word and *be-simcha* 'gladly' is a content word.

2.3. Hypotheses

At first glance, improvements in performance were expected in all the three categories in direct correlation with age and duration of exposure, as many studies on lexicon development, especially in second languages, have shown (Meira & Fitzpatrick, 2000; Laufer & Hulstijn, 2001; Vermeer, 2001; Qian, 2004; Jiang, 2007; Mondria & Wiersma, 2007; Nation, 2007; David, 2008; Fitzpatrick et al., 2008). In other words, there were expectations for the levels of abstractness, lexical diversity, and lexical density to be higher in the writing of eleventh graders than in the writing of ninth graders. Indeed, many studies have linked improvements in the three categories with progress in L2 acquisition and overall academic achievement (Kroll & Tokowicz, 2001; Polio, 2001; Daller et al., 2003; Ravid 2005).

In the case of lexical diversity, however, there are indications that the differences in the length of the texts may affect the results. Ninth-grade students wrote shorter texts than eleventh-grade students.

3. Results and Discussion

Table 1 summarize the results of the tests by age group:

---

[4] Inflectional suffixes were not separated, so *ne'arim* 'boys') is a content word, as is *halchu* 'went', despite the fact that in a morphological analysis *nearim* can be seen as a combination of *naar* and a masculine plural suffix, and the word *halchu* includes a masculine plural suffix of the past tense of the verb in the third person.

Table 1. Means, standard deviations and differences between the two age groups according to the three lexical measures

| | Grade | | | | Difference |
| Lexical measure | Ninth (n=25) | | Eleventh (n=25) | | |
	M.	S.D.	M.	S.D.	t
Level of Abstractness	35.76%	.1365	46.24%	.1215	.0133*
Lexical diversity	50.12%	.0818	49.96%	.0896	.9487
Lexical density	41.72%	.0580	39.60%	.0415	.1520

*p <.05

3.1. Level of abstractness

The data on level of abstractness in grades nine and eleven confirms my expectations. There is an increase in the average integration of abstract nouns relative to all nouns as the age and time of acquisition of Hebrew as L2 increases. As shown in Table 1, the language of the eleventh-graders is more abstract than that of the ninth-graders (46.24% and 35.76% respectively). According to a t-test, this gap is statistically significant (p< 0.05).

The research literature reviewed on abstractness in the texts of L2 acquirers indicates that there is greater integration of abstract language as more time is spent in acquiring L2 (Kroll & Tokowicz, 2001; Ravid, 2005), but it does not establish the degree of correlation. Lack of precision is justified at this stage because the acquisition process may depend on a number of factors beyond the amount of time of exposure to L2. These may include intensity of use and exposure to the acquired language and the learner's motivation to learn it, as typically determined by the relative status of L1 and L2 in the community, and personal and general attitudes to both. Therefore, only by isolating these factors and conducting separate quantitative tests will it be possible to determine unequivocally the dependence between, say, the degree of abstractness and the time spent on acquisition in comparison with other factors.

This study has used the abstractness level of a text as an index of L2 competence. According to this measure, it is possible to conclude that the level of proficiency in Hebrew as L2 in the ninth grade is lower than it is in the eleventh grade, as expected. It can be concluded that level of abstractness as a measure of the development of Hebrew as L2 among native Arabic speakers is an effective test in the age groups examined here. Moreover, there is no need for a uniform text length. As noted, the difference in abstractness between the two groups is very significant notwithstanding the clear difference between the two groups of students in terms of the length of the compositions. The conclusion is that this test, unlike other lexical tests, in particular tests of lexical diversity, is applicable in short texts up to 100 tokens.

3.2. Lexical diversity

Before this study began, lexical diversity was expected to increase with the acquisition time of L2 (Daller et al., 2003; Malvern et al., 2004). However, there was no evidence of a difference in lexical diversity between the two grades. Table 1 actually shows a slight decrease between ninth grade (50.12%) and eleventh grade (49.96%), but according to a t-test conducted, this difference is not statistically significant (p= 0.8344). What, then, is the explanation for the lack of increase in lexical diversity between grades nine and eleven? Is such a situation indeed possible after students have undoubtedly learned more lexemes over the course of two years?

A possible explanation for the lack of increase is the length of the compositions. The average length of the composition in the ninth grade is 76.28 tokens, while in the eleventh grade it is 98.88. That is, there is a difference of approximately 20 tokens between the two groups.

According to what is known in the research literature the length of the text undoubtedly affects its lexical diversity, because the number of types is usually limited, and the number of tokens is not. As the text grows longer but continues to address a limited subject, there is a repetition of previously mentioned types because the types are limited and as a result the lexical diversity decreases. It is believed that when a writer moves to a different topic within the same composition, the possibility for additional lexical diversity may rise, raising the lexical diversity index; but in the present case there was one topic, hence the tendency for writers to repeat a limited variety of lexeme types.

Many studies on the lexical diversity measure focused on texts consisting of 100 to 200 tokens, considered optimal for measuring lexical competence. A minimum of 100 tokens is considered necessary to indicate lexical competence, while a text longer than 200 words is affected by length in a way that skews the results. Accordingly, I, too, chose this length range and, given the writing abilities of high school L2 acquirers, it was decided to only count the first 100 tokens of each composition. But in the ninth-grade some compositions received did not even reach 70 tokens. Shortening the sample size to 53 tokens, in accordance with the shortest composition in the corpus, would have reduced the reliability of the lexical competence measures, so all the compositions were checked, and only the first one hundred of those that were longer than one hundred textual units.

This difference in the length of the compositions may be the cause of the lack of improvement in the results of the lexical diversity index from the ninth grade to the eleventh grade. Assuming that lexical diversity does indeed rise with exposure, this deviant result may be attributed to text length differences. This is consistent with the large number of studies that indicated that there was no increase and sometimes even a decrease in the lexical diversity index when the number of tokens in a text increased (Hess et al., 1986; Malvern et al., 2004; McCarthy & Jarvis, 2007, 2010). Therefore, this measure is not practical for testing progress in L2 Hebrew-L1 Arabic, at least at Grade 9 in the Negev, certainly not as a single measure, because most students of these ages do not write long enough essays on one subject. It is recommended to postpone tests of lexical diversity to high school or higher education, where essays may be expected to exceed 100-word tokens on one subject.

Another problem with this criterion is that apparent lexical diversity can sometimes mask linguistic deficits – such as incorrect lexical choices, semantic inaccuracies, and inappropriate register. A superficial examination of the writings of the ninth-graders reveals prominent lexical flaws even when the lexicon seems to be diverse.

3.3. Lexical density

The lexical density in the written compositions of the ninth-graders was 41.72% and in those of the eleventh-graders 39.6% (See Table 1). Prior to conducting the study, lexical density was expected to increase over time (see e.g. Daller et al., 2003; Malvern et al., 2004). A review of Table 1 shows that the average lexical density among ninth-graders is 41.72%, while among eleventh-graders it is 39.6%. In other words, there is a slight decrease between ninth grade and eleventh grade. Once again, according to the t-test, this difference is not statistically significant ($p > 0.05$). The question is how to explain the lack of an increase in lexical density between grades nine and eleven. The answer proposed is based on the curriculum of Hebrew as L2 for native speakers of Arabic and the matriculation exams in Hebrew as L2, which may explain the lack of increase in lexical density between Grades 9 and 11.

3.3.1. Considerations of cohesion and coherence in the context of lexical density

The curriculum of the Ministry of Education (2011) emphasizes the importance of understanding written texts and therefore promotes writing coherent and cohesive texts with high readability (i.e. relatively easy to read, process, and understand). Means for this end include the use of logical connectors, conjunctions, and anaphora, which are words that refer to words that appeared earlier in the text. These anaphoric references help create coherence and cohesion, as does proper use of conjunctions and logical connectors, helping the reader understand the logical connection between sentences or paragraphs within the text.

According to Harrison & Bakker (1998) and Khamahani (2015), the readability of a text depends, among other things, on its density: a relatively non-dense text, with a relatively high ratio of function words, will be more readable than a denser text. The logical connectors indicate the relationship between the sentences and therefore reduce the reader's need to perform complex cognitive analysis of the text. The awareness that they produce guides the reader to move forward and backward within the text correctly (Olshtain & Cohen, 2005, p. 317).

As students approach the matriculation exam in Hebrew as L2 (comprehension, expression, and language knowledge) they encounter more texts that focus on logical connectors, diverse conjunctions and anaphora. They are instructed to write coherent and cohesive compositions and receive explicit metalinguistic guidelines for the use of these measures. As a result, one can expect an increase in the proper use of these elements, leading to a decrease in lexical density.

This explanation was supported by selecting compositions from the two groups of students at random and comparing the use of function words whose purpose is to create coherence and cohesion in the texts. The first and last compositions of each group of students were chosen for comparison:

Table 2. Conjunctions and anaphora in the first composition of each age group

Specific function words	class	
	9	11
Conjunctions	*ve* (and) *4, *o* (or) *4, *im* (if), *basof* (finally, ultimately)	*ve* (and) *9, *o* (or) *2, *gam* (and)*2, *ki* (because), *aval* (but), *bishvil* (for)
Anaphora	*hu* (he)*5, *ze* (that), *le* (to)*5, *oto* (same)	*lo* (to him), *bo* (in him) *6, *hu* (he)*5, *shelo* (of him)*7
Total	22	35

Note. * indicates the number of tokens (the number of repetitions) of the same type.

The first student from the ninth grade used 22 function tokens that belong to the realm of cohesion and coherence, while the first eleventh-grade student used 35 function tokens. This is a significant difference relative to the length of the text.

Table 3. Conjunctions and anaphora in the last composition of each age group

Specific function words	class	
	9	11
Conjunctions	*ve* (and), *im* (with), *ki* (because)	*ve* (and) *8, *gam* (also)
Anaphora	*hu* (he)*4, *ze* (that), *lo* (to him)*5, *bo* (in him) *3, *shelo* (of him) *3, *oto* (him)*2	*lo* (to him)*2, *bo* (in him) *5, *hu* (he)*4, *shelo* (of him) *5, *oto (him), ze* (that)
Total	21	27

Note. * indicates the number of tokens (the number of repetitions) of the same type.

The last student from the ninth grade used 21 function tokens that belong to the realm of cohesion and coherence, while the last student from the eleventh grade used 27.

This analysis is further supported by Johansson's (2009) findings in tests conducted among 13-year-old Swedish speakers. In a comparison of narrative texts with expository texts in this age group, a decline in lexical density was found in the expository genre, in contrast to what was expected, since the expository genre is known in linguistic research as denser than the narrative genre. The explanation suggested by the researcher (Johansson, 2009, p. 74) is that at this stage of the development of literacy in the Swedish mother tongue, the children demonstrated sensitivity to the cohesion and coherence requirements of the expository genre and used many logical connectors and pronouns, that is, function words, which reduced the density relative to that of the narrative genre. In applying this explanation to the present population, this phenomenon could be expected at a later age, since it involves the acquisition of a second language. This explanation is in keeping with the findings regarding the eleventh grade, where great attention is paid to coherence.

It is proposed to distinguish between basic "primitive" connectors and "advanced" ones. The early ones are acquired very early on and include **and**, **but**, and so on, while the advanced ones are acquired in later stages of literacy, so that a quantitative examination of the connectors, without lexical distinction, will not necessarily reveal lexical progression. Ninth grade students generally used primary connectors such as **and**, while eleventh graders used many "advanced" connectors, such as **accordingly**, **despite this**, and so on.

The compositions of ninth graders contain many binomials, i.e. pairs of words that are related and tend to appear together as a collocation, such as English 'high and mighty'. This binomial style characterized Semitic languages, including Classical Hebrew, e.g. *yafe vena'im* 'pretty and pleasant', and is appreciated in Modern Arabic. But in Modern Hebrew, where it is no longer natural, it creates an expectation of additional information in the second word, which does not actually exist, so that the reader may be confused. Pairs such as *ta'ana vede'a sheli* 'my argument and opinion', *lo de'a o ta'ana* 'he has an opinion or argument', and *yesh harbeh de'ot ve-ta'anot* 'there are many opinions and arguments' (first essay, ninth grade) raise the apparent density (and at the same time the lexical diversity) of the texts. But they do not increase cohesion according to the norms of Modern Hebrew, but rather create clumsy texts.

The last two tables (Tables 2 and 3) show that eleventh grade students use more logical connectors – conjunctions – and various anaphora tokens (35 and 27 tokens, respectively), especially high-level ones, than ninth graders (22 and 21 tokens, respectively). For example, in one essay by an eleventh grader, the following sentence appears: *Mi-tzad echad, hi kim'at chasrat to'elet la-chevra shela, mi-tzad sheni hi menaka et ha-bayit osa ochel o ozeret li-yladeha.* 'On the one hand, she is almost useless to her society, on the other hand she cleans the house makes food or helps her children'. Here the use of high-

level conjunctions that indicate a logical connection of contrast is apparent. Ninth graders, even if they use high-level logical connectors, do not necessarily use them correctly, resulting in many unclear sentences. For example, a ninth grader wrote the following sentence *hi (ha-isha) rotza lihiyot im mora im rofa o akeret bayit.* 'She (the woman) wants to be if a teacher if a doctor or a housewife'. This use of the conditional particle *im* 'if' where there is no condition (but rather a choice) is the result of interference from Arabic, where the phonetically similar *imma* is the particle conveying choice. It is known that proficiency in logical connectors in Hebrew language is one of the most difficult tasks in the development of literacy even among native speakers (Olshtain & Cohen, 2005, p. 317).

As mentioned above, the lack of increase in the lexical density of eleventh-grade students may be due to the fact that they are approaching the matriculation exam in Hebrew and the emphasis placed in the curriculum on connectors and coherence by means of correct function words. One may therefore assume that the lack of increase in lexical density does not indicate a decrease in lexical competence, but rather is anchored in the scholastic context, with deliberate linguistic intervention that is formulated in the curriculum. In order to support this explanation, it will be shown that in recent years, in all the matriculation exams in Hebrew, Arab students had questions about logical connections and anaphora. In the summer of 2016 in questionnaire 014381 (the first three units in matriculation "Hebrew for Arabic speakers"), six points of the total of fifty points in questions about the text were related to questions about conjunctions and anaphora.

There is another reservation regarding the linkage between high lexical density and lexical competence in unequivocal and direct correlation with age. As mentioned above, the more one advances along the acquisition axis, the more content words are used relative to all the tokens in texts written under academic circumstances. But it is clear that this situation is true up to a certain, optimal point on the acquisition axis and there is no situation in which the proportion of content words out of all the tokens in the text is higher than the optimal point because the functional words play a very important role in understanding the text. Without them, the text would not be natural or coherent at all. Function words are an inherent part of natural language. Their removal creates a very specific kind of text, such as a text in the style of a telegram or newspaper headlines, for example, "Monday 4:00 a.m arrive Beer-Sheva" which probably means "On Monday at four a.m. I will arrive in Beer-Sheva." It is almost impossible to maintain this style for more than a line or two, hence the focus on cohesion in the eleventh grade.

To conclude, there may be one integrated explanation for the findings suggested in this section. First, in the ninth grade, lexical density testing is probably not yet effective as a measure of lexical competence. Among ninth graders there is a misuse of connectors, including the omission of essential functional words, which creates the impression of density as a sign of literacy. For example, a ninth grader writes: *Yesh anashim rotzim ha-nashim shelahem lihyot ovdot.* 'There are people want their women to be workers'. The correct Hebrew sentence is *yesh anashim **she**-rotzim **she**-ha-nashim shelahem ya'avdu* literally 'there are people **who** want **that** their women should work'. The subordinating conjunction *she* 'who, that', a necessary link in Hebrew in this context, is omitted twice via interference from Arabic. We can conclude that at this stage of linguistic competence, the lexical density measure in its present form is not yet relevant.

In the eleventh grade, after the students have had explicit instruction in cohesion, the lexical density measure seems more reliable as a marker of lexical and linguistic-academic competence, although this meta-linguistic training leads to an apparent reduction in lexical density compared to the ninth grade rather than the expected increase.

Another possible explanation for the absence of any significant difference in lexical density between the groups is that the age gap is relatively small and a larger gap is needed to test lexical density in L2.

Recall Johansson's study, in which she found that between the ages of 10 and 13, and even between the ages of 13 and 17, no difference in lexical diversity or even in density was apparent. The expected differences were found only in the choice of large ranges (Johansson, 2009, p. 77). One may assume that here, too, the increase in lexical density, and possibly lexical diversity, would be significant when comparing groups of greater age differences than those examined.

It is therefore recommended to conduct this examination in classes with older high school students or in institutions of higher education after the expected increase due to the additional exposure takes place. A similar conclusion was reached in the field of morphology by Saiegh-Haddad and Jayusy, who note that some linguistic tests (in this case, morphological ones) require a certain level of linguistic competence on the part of their subjects in order to yield representative and reliable results (Saiegh-Haddad & Jayusy, 2016, p. 432).

4. Conclusion

In this paper, lexical competence was examined in Hebrew as L2 among speakers whose first language is Arabic using three measures, taking into account the L2 curriculum in the field of Hebrew as the L2 of Arabic in the local school system. Not all lexical measures were found to be as useful for testing progress, contrary to what may have been expected based on the extant literature.

Initial expectations predicted an increase in the average achievements in the three lexical tests correlating with age. With regard to the level of abstraction of the texts, which reflects progress in second language acquisition, the expectations were fully realized. The increase in the use of abstract nouns relative to all nouns in the compositions correlated directly with age and exposure to Hebrew.

Second, with respect to lexical diversity, there was no clear increase with age, mainly because the comparison was between texts of different lengths, since most ninth graders did not write 100 tokens. Therefore, it may be concluded that reliable results cannot be expected until after the ninth grade, when students reach a level of proficiency that allows them to compose structured and readable texts that include 100 tokens.

Third, similarly to the findings for lexical diversity, no clear increase in lexical density was found to correlate with age. The explanations proposed for the lack of increase as opposed to the expectations are the emphasis in the eleventh-grade curriculum on the writing of coherent, cohesive texts through the relatively frequent use of conjunctions and anaphora, which are function words, which reduces lexical density. In addition, the use of function words in ninth graders is still not sufficiently consolidated (according to Hebrew norms) to serve as a measure of linguistic competence: there is a lack of cohesion on the one hand and a surplus of connectivity on the other, inter alia under the influence of Arabic. It is clear that there is not necessarily a direct univalent connection between lexical density and lexical competence. A lack of increase in density may indicate an improvement in connectivity. Among students in the eleventh grade, there is already a better and more correct use of conjunctions and anaphora due to the meta-linguistic intervention that takes place by means of the curriculum, and from this point on it is possible to see the lexical density measure as a marker for progress in linguistic-academic ability in L2.

In light of the findings of this study, the lexical measures should be divided into two groups: beginning (including level of abstraction) and advanced (including lexical diversity and lexical density). The first group should be performed in high level of competence. The second group is already reliable in low level of competence.

The results of this study show that there is still much room for further research on applying adaptations of the lexical measures in order to adjust them to synthetic languages like Hebrew and

Arabic. One should bear in mind that some measures are effective only from a certain level of competence, and that for younger learners we need to find or develop other measures and test their effectiveness in a controlled manner.

5. Ethics Committee Approval

The author(s) confirm(s) that the study does not need ethics committee approval according to the research integrity rules in their country (Date of Confirmation: November 19, 2019).

Acknowledgements

References

Abdalla, F., Robb, M. P., & Al-Shatti, T. (2010). Stuttering and lexical category in adult Arabic speakers. *Clinical Linguistics & Phonetics*, *24*(1), 70-81.

Amara, M. H. (1999). *Politics and sociolinguistic reflexes: Palestinian border villages* (Vol. 19). John Benjamins Publishing.

Amara, M., & Mar'i, A. A. R. (2002). *Language education policy: The Arab minority in Israel*. Kluwer Academic Publishers.

August, D. (Ed.), Shanahan, T. (Ed.). (2006). Developing Literacy in Second-Language Learners. New York: Routledge. doi: https://doi.org/10.4324/9781315094922.

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

Borghi, A. M., Barca, L., Binkofski, F., & Tummolini, L. (2018). *Varieties of abstract concepts: development, use and representation in the brain*. Philosophical Transactions of the Royal Society B: Biological Sciences.

Borghi, A. M., Binkofski, F., Castelfranchi, C., Cimatti, F., Scorolli, C., & Tummolini, L. (2017). The challenge of abstract concepts. *Psychological Bulletin*, *143*(3), 263.

Borràs, J., & Llanes, À. (2020). L2 reading and vocabulary development after a short Study Abroad experience. *Vigo International Journal of Applied Linguistics*, *17*, 35-55.

Bürki, A., Elbuy, S., Madec, S., & Vasishth, S. (2020). What did we learn from forty years of research on semantic interference? A Bayesian meta-analysis. *Journal of Memory and Language*, *114*, 104125.

Central Bureau of Statistics, 2019. Available online at: https://www.cbs.gov.il/en/mediarelease/Pages/2019/Israel-in-Figures-Selected-Annual-Data-2019.aspx.

Choi, W., & Jeong, H. (2016). Finding an appropriate lexical diversity measurement for a small-sized corpus and its application to a comparative study of L2 learners' writings. *Multimedia Tools and Applications, 75*(21), 13015-13022.

Crossley, S. A., & Skalicky, S. (2019). Examining lexical development in second language learners: An approximate replication of salsbury, crossley & McNamara (2011). *Language Teaching, 52*(3), 385-405.

Daller, H., Van Hout, R., & Treffers-Daller, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied linguistics, 24*(2), 197-222.

David, A. (2008). Vocabulary breadth in French L2 learners. *Language Learning Journal, 36*(2), 167-180.

Demir-Vegter, S., Aarts, R., & Kurvers, J. (2014). Lexical richness in maternal input and vocabulary development of Turkish preschoolers in the Netherlands. *Journal of psycholinguistic research, 43*(2), 149-165.

Dillard, J. P., & Pfau, M. (2002). *The persuasion handbook: Developments in theory and practice*. Sage Publications.

Dimroth, C. (2008). Age effects on the process of L2 acquisition? Evidence from the acquisition of negation and finiteness in L2 German. *Language learning, 58*(1), 117-150. doi: http://dx.doi.org/10.1017/S0261444817000362.

Ebedy, D. (2020). Developing Lexical Richness among EFL Students: Effect of Morphological Awareness Training on Writing Performance. *Journal of Research in Curriculum Instruction and Educational Technology*, 6(2), 45-68. doi: http://10.21608/jrciet.2020.80131.

Ferré, P., Sánchez-Casas, R., & Guasch, M. (2006). Can a horse be a donkey? Semantic and form interference effects in translation recognition in early and late proficient and nonproficient Spanish-Catalan bilinguals. *Language Learning*, *56*(4), 571-608.

Fitzpatrick, T., Al-Qarni, I., & Meara, P. (2008). Intensive vocabulary learning: A case study. *Language Learning Journal, 36*(2), 239-248.

Flynn, S., & O'Neil, W. (Eds.). (2012). *Linguistic theory in second language acquisition* (Vol. 8). Springer Science & Business Media.

Gebril, A., & Plakans, L. (2016). Source-based tasks in academic writing assessment: Lexical diversity, textual borrowing and proficiency. *Journal of English for academic purposes*, *24*, 78-88.

Gharibi, K., & Boers, F. (2019). Influential factors in lexical richness of young heritage speakers' family language: Iranians in New Zealand. *International Journal of Bilingualism*, *23*(2), 381-399.

Ha, H. S. (2019). Lexical Richness in EFL Undergraduate Students' Academic Writing. *English Teaching*, *74*(3), 3-28.

Halliday, M.A.K. (1985). *Spoken and written language*. Victoria: Deakin University Press.

Halliday, M.A.K. (1989). *Spoken and written language*. (2nd Edition). Oxford: Oxford University Press.

Harrison, S., & Bakker, P. (1998). Two new readability predictors for the professional writer: pilot trials. *Journal of Research in Reading*, *21*(2), 121-138.

Henkin, R. (2011). Hebrew and Arabic in asymmetric contact in Israel. *Lodz Papers in Pragmatics, 6*(2), 195-228.

Hess, C. W., Sefton, K. M., & Landry, R. G. (1986). Sample size and type-token ratios for oral language of preschool children. *Journal of Speech, Language, and Hearing Research, 29*(1), 129-134.

Hilton, H. (2008). The link between vocabulary knowledge and spoken L2 proficiency. *Language Learning Journal, 36*(2), 153-166.

Jarvis, S. (2012). Lexical challenges in the intersection of applied linguistics and ANLP. In C. Boonthum-Denecke, P.M. McCarthy, & T.A. Lamkin (Eds.), *Cross-disciplinary advances in applied natural language processing: Issues and approaches* (pp. 50-72). Hershey, PA: IGI Publishing.

Jarvis, S. (2013). Capturing the diversity in lexical diversity. *Language Learning, 63*, 87-106.

Jarvis, S. (2017). Grounding lexical diversity in human judgments. *Language Testing, 34*(4), 537-553.

Jiang, N. (2007). Semantic transfer and development in adult L2 vocabulary acquisition. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 101-126). Cambridge: Cambridge University Press.

Johansson, V. (2009). Lexical diversity and lexical density in speech and writing: A developmental perspective. *Lund Working Papers in Linguistics, 53*, 61-79.

Johnson, W. (1939). *Language and speech hygiene: An application of general semantics*. Chicago: Institute of General Semantics.

Johnson, W. (1944). Studies in language behavior: A program of research. *Psychological Monographs, 56*(2), 1-15.

Juanggo, W. (2018). Investigating lexical diversity and lexical sophistication of productive vocabulary in the written discourse of Indonesian EFL learners. *Indonesian Journal of Applied Linguistics*, *8*(1), 38-48.

Juste, F. S., Sassi, F. C., & de Andrade, C. R. F. (2012). Exchange of disfluency with age from function to content words in Brazilian Portuguese speakers who do and do not stutter. *Clinical linguistics & phonetics*, *26*(11-12), 946-961.

Khamahani, G. (2015). A corpus-based analysis of Tehran Times and Azeri News headlines: Focus on lexical density and readability. *International Journal of Humanities Social Sciences and Education (IJHSSE)*, *2*(1), 12-16.

Koizumi, R. (2012). Relationships between text length and lexical diversity measures: Can we use short texts of less than 100 tokens?. *Vocabulary Learning and Instruction, 1*(1), 60-69.

Koizumi, R., & In'nami, Y. (2012). Effects of text length on lexical diversity measures: Using short texts with less than 200 tokens. *System, 40*(4), 522-532.

Kojima, M., & Yamashita, J. (2014). Reliability of lexical richness measures based on word lists in short second language productions. *System*, *42*, 23-33.

Kroll, J., & Tokowicz, N. (2001). The development of conceptual representation for words in a second language. In J. Nicol (Eds.), *One Mind, Two Languages: Bilingual Language Processing* (pp. 49-71). Malden, MA: Blackwell Publishers.

Kyle, K. (2019). Measuring Lexical Richness. In S. Webb (Eds.) *The Routledge Handbook of Vocabulary Studies* (pp. 454-475). New York: Routledge.

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied linguistics, 22*(1), 1-26.

Lefebvre, C., White, L., & Jourdan, C. (Eds.). (2006). *L2 acquisition and Creole genesis: dialogues* (Vol. 42). John Benjamins Publishing.

Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal*, *96*(2), 190-208.

MacWhinney, B. (2005). A unified model of language acquisition. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 49-67). New York, NY: Oxford University Press.

Malvern, D., & Richards, B. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing, 19*(1), 85-104.

Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004*). Lexical diversity and language development*. New York: Palgrave Macmillan.

Mayer, K. M., Macedonia, M., & von Kriegstein, K. (2017). Recently learned foreign abstract and concrete nouns are represented in distinct cortical networks similar to the native language. *Human Brain Mapping*, *38*(9), 4398-4412.

McCarthy, P. M., & Jarvis, S. (2007). Vocd: A theoretical and empirical evaluation. *Language Testing, 24*(4), 459-488.

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods, 42*(2), 381-392.

Meara, P., & Fitzpatrick, T. (2000). Lex30: An improved method of assessing productive vocabulary in an L2. *System, 28*(1), 19-30.

Ministry of Education (2011). Hebrew as 2nd language in Arabic schools in Israel, 9-12th grade. Ministry of Education. Jerusalem. Available online at: http://cms.education.gov.il/NR/rdonlyres/202A56F5-EC9E-43CD-9592-0EAE76451DF2/212803/29916.pdf.

Mondria, J., & Wiersma, B. (2007). Receptive, productive, and receptive + productive L2 vocabulary learning: What difference does it make?. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and Assessing Vocabulary Knowledge* (pp. 79-100). Cambridge: Cambridge University Press.

Mor-Sommerfeld, A., Azaiza, F., & Hertz-Lazarowitz, R. (2007). Into the future: Towards bilingual education in Israel. *Education, Citizenship and Social Justice*, *2*(1), 5-22.

Nassaji, H. (2020). Assessing the effectiveness of interactional feedback for L2 acquisition: Issues and challenges. *Language Teaching*, *53*(1), 3-28.

Nation, I.S.P. (2005). Teaching and learning vocabulary. In E. Hinkel (Eds.), *Handbook of Research in Second Language Teaching and Learning* (pp. 581-595). Mahwah, NJ: Lawrence Erlbaum.

Nation, I.S.P. (2007). Fundamental issues in modelling and assessing vocabulary knowledge. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and Assessing Vocabulary Knowledge* (pp. 33-43). Cambridge: Cambridge University Press.

Olshtain E., Cohen E. (2005) Logical connectors in Hebrew: How well do eighth-graders master them?. In: Ravid D.D., Shyldkrot H.BZ. (Eds.), *Perspectives on Language and Language Development* (pp. 315-326). Boston, MA: Springer.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of experimental psychology, 76*, 1–25.

Polio, C. (2001). Research methodology in second language writing research: The case of text-based studies. In T. Silva, & P.K. Matsuda (Eds.), *On Second Language Writing* (pp. 91-115). Mahwah, NJ: Lawrence Erlbaum.

Prior, A., Degani, T., Awawdy, S., Yassin, R., & Korem, N. (2017). Is susceptibility to cross-language interference domain specific?. *Cognition*, *165*, 10-25.

Qian, D. (2004). Second language lexical inferencing: Preferences, perceptions, and practices. In P. Bogaards, & B. Laufer (Eds.), *Vocabulary in a Second Language: Selection, Acquisition, and Testing* (pp. 155-169). Amsterdam; Philadelphia: John Benjamins.

Ravid, D. (2005). Emergence of linguistic complexity in later language development: Evidence from expository text construction. In D. Ravid, & H. Bat-Zeev Shyldkrot (Eds.), *Perspectives on Language and Language Development* (pp. 337-355). Dordrecht: Kluwer.

Ravid, D. (2006). Semantic development in textual contexts during the school years: Noun scale analyses. *Journal of Child Language, 33*(4), 791-821.

Saiegh-Haddad, E. (2003). Linguistic distance and initial reading acquisition: The case of Arabic diglossia. *Applied Psycholinguistic, 24*(3), 431-451.

Saiegh-Haddad, E., & Jayusy, A. (2016). Metalinguistic awareness in reading Hebrew L2. *Acquisition and Development of Hebrew: From Infancy to Adolescence, 19*, 353-386.

Šišková, Z. (2012). Lexical richness in EFL students' narratives. *Language Studies Working Papers*, *4*, 26-36.

Tanaka, S. (2012). New directions in L2 lexical development. *Vocabulary Learning and Instruction, 1*(1), 1-9.

Tweedie, F. J., & Baayen, R. H. (1998). How variable may a constant be? Measures of lexical richness in perspective. *Computers and the Humanities*, *32*(5), 323-352.

Ure, J. (1971). Lexical density and register differentiation. In G.E. Perren, & J .M. Trim (Eds.), *Applications of Linguistics* (pp. 443-452). London: Cambridge University Press.

Vedder, I., & Benigno, V. (2016). Lexical richness and collocational competence in second-language writing. *International Review of Applied Linguistics in Language Teaching*, *54*(1), 23-42.

Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied psycholinguistics, 22*(2), 217-234.

Webb, S. (Ed.). (2019). *The Routledge Handbook of Vocabulary Studies*. London: Routledge. https://doi.org/10.4324/9780429291586.

Wood, C. L., Bustamante, K. N., Schatschneider, C., & Hart, S. (2019). Relationship between children's lexical diversity in written narratives and performance on a standardized reading vocabulary measure. *Assessment for Effective Intervention*, *44*(3), 173-183.

Zhai, L. (2016). A study on Chinese EFL learners' vocabulary usage in writing. *Journal of Language Teaching and Research*, *7*(4), 752-759.

Zhang, Y (2005) Processing and formal instruction in the L2 acquisition of five Chinese grammatical morphemes. In: Pienemann, M (Eds.), *Cross-linguistic aspects of Processability Theory* (pp. 155-177). Philadelphia, PA: John Benjamins.

Zheng, Y. (2016). The complex, dynamic development of L2 lexical use: A longitudinal study on Chinese learners of English. *System*, *56*, 40-53.

Arap öğrencilerin ikinci yabancı dil olarak İbranice'deki ilerlemeyi test etmek için sözcüksel ölçüler

Öz

Ure (1971) ve Halliday (1985) 'ten beri, araştırma literatüründe, sözlükteki gelişmeleri araştırmak için bir vekil olarak farklı önlemler önerilmiştir. Bunlar sözcüksel zenginlik ölçüleri olarak adlandırılırlar ve sözlüğün ilerlemesini değerlendirmede önemli bir araçtır (Johansson, 2009). Bu makale, İsrail'deki Arap öğrencilerin İbranice edinmesindeki ilerlemenin üç sözcüksel zenginlik ölçüsünü değerlendiriyor: soyutluk düzeyi, sözcük çeşitliliği ve sözcük yoğunluğu. Kitap, 9. ve 11. sınıflar tarafından İbranice yazılmış altmış denemeden oluşuyor. Nicel analiz birkaç bulguyu ortaya çıkarmıştır. Birincisi, beklendiği gibi, öznenin yaşı ile birlikte soyutluk artar. Bununla birlikte yoğunluk testi, yaşla birlikte beklenen yoğunluk artışını göstermedi. Bunun olası bir nedeni, daha genç öğrencilerin denemelerindeki nispeten yüksek yoğunluktur, bu da yüksek dil yeterliliğinden çok uyum eksikliğini gösterir. On birinci sınıflar için ulusal lise dili müfredatı, yoğunluğun azalmasına neden olan bağlantıya odaklanır. Üçüncüsü, çeşitlilik testi de istatistiksel olarak anlamlı bulgular vermedi, çünkü bu test muhtemelen dokuzuncu sınıfta hala güvenilir bir test sağlamayan metin uzunluğundan etkilendi. Sonuç olarak, sözcük ölçütleri iki türe ayrılabilir: herhangi bir yeterlilik düzeyinde uygulanabilen temel testler; ve yüksek düzeyde yeterlilik gerektiren gelişmiş testler. Soyutluk seviyesi birinci tiptedir: edinimin ilk aşamalarında bile L2 edinimindeki ilerlemenin bir ölçüsü olarak etkilidir, çeşitlilik ve yoğunluk ise daha yüksek bir yeterlilik seviyesi gerektirir.

Anahtar sözcükler: ikinci dil edinimi; sözlük geliştirme; yazıyı değerlendirmek; soyutluk seviyesi; sözcük çeşitliliği; sözcük yoğunluğu

AUTHOR BIODATA

Eihab Abu Rabiah is a Ph.D. student at the department of Hebrew language at Ben-Gurion University of the Negev, Israel. He holds a Masters Degree in Hebrew language. He has participated in many national and international conferences and seminars. He teaches in the Pre-academic Studies Center at Ben-Gurion University of the Negev and at the Department of Literature, Language and the Arts at the Open University of Israel. His research areas include second language acquisition and vocabulary development. ORCID: https://orcid.org/0000-0002-8837-1089