LLT Journal, e-ISSN 2579-9533, p-ISSN 1410-7201, Vol. 26, No. 1, April 2023, pp. 168-182

LLT Journal: A Journal on Language and Language Teaching
http://e-journal.usd.ac.id/index.php/LLT
Sanata Dharma University, Yogyakarta, Indonesia

EVALUATING L2 VOCABULARY DEVELOPMENT FEATURES USING LEXICAL DENSITY AND LEXICAL DIVERSITY MEASURES

Eihab Abu-Rabiah
The University of Utah, United States
correspondence: aburabiah.e@gmail.com
https://doi.org/10.24071/llt.v26i1.5841
received 1 February 2023; accepted 24 March 2023

Abstract
Most of the research on lexical measures was conducted in English and conducted similarly in other languages without accurate adaptations to the language being tested. The first objective of this study is to identify differences between applications of lexical density and lexical diversity when tested as in English and when adapted to the language being tested. The second objective is to inspect the effect of acquisition time on intermediate-level second-language learners' lexical density and diversity. In a test-retest study over one year, Arabic Hebrew bilinguals (n=23) wrote argumentative essays, in which lexical density and diversity were analyzed according to two approaches; a conventional approach as in English and an adapted approach more aligned with Hebrew linguistic features. In both approaches, lexical density was calculated as the number of lexical words in proportion to the total number of words and lexical diversity using the type-token ratio (TTR) and Corrected TTR (CTTR). Findings reveal statistically significant differences between the two approaches and the adapted approach allowed for cross-language and interlanguage comparisons. Although no significant increase in lexical density and diversity over the year was found, various patterns for lexical development were observed; at the second time point, students with the lowest level of lexical diversity substantially improved and more essays fell within the lexical density range of a typically written language.

Keywords: argumentative essays, Arabic-Hebrew bilinguals, language assessment, lexical measures, second language acquisition

Introduction

Most people acquire a second language (L2) at some point in their lives (Klein et al., 1986). Second Language Acquisition (SLA) is the field dedicated to understanding the process of L2 acquisition, in both naturalistic and instructed contexts (Ortega, 2014). Within SLA, the article focuses on the acquisition of vocabulary (lexicon), which is one of the most important aspects of this process (Cameron, 2001).

In general, vocabulary is defined as the words used by a particular person or all the words that exist in a particular language or subject (Cambridge Dictionary). Words are the building blocks of language, from which larger structures are created, such as sentences, paragraphs, and texts. Without them, there is no language

(Milton, 2009). Vocabulary knowledge makes a great contribution to the four language skills: reading, writing, speaking, and listening, and also can predict them (Dakhi & Fitria, 2019). It is central to communicative competence and either leads to successful communication or can prevent it (Alqahtani, 2015; Schmitt, 2000).

Furthermore, the acquisition of vocabulary plays an important role in the formation of complete spoken and written texts (Read, 2000; Nation, 2001). This knowledge of vocabulary is used as an indicator of academic skill and academic success (Roche & Harrington, 2013). Since a lack of vocabulary knowledge is an obstacle to learning (Alqahtani, 2015), learners see vocabulary acquisition as the main task in L2 acquisition and devote most of their time to expanding it (Read, 2000).

To assess the size of a learner's vocabulary, several quantitative lexical measures have been devised and used for over a hundred years (Jarvis, 2013b). Two of these measures are central to the research: lexical density and lexical diversity. The study's primary aim is to assess the applicability of these commonly used measures in testing vocabulary development in a language other than English. The study intends to investigate if the results obtained from these measures align with the unique linguistic characteristics of the language under analysis, or if the measures require modification to be effective. In other words, the objective is to determine the validity of applying the same methods used in English language research to another language with and without considering their distinct linguistic features, as well as testing the effect of acquisition time on lexical density and lexical diversity of intermediate-level second-language learners.

Lexical density

Lexical density is based on the distinction between content words and function words, which is a central distinction in studies on syntactic categories of natural languages (Cann, 2000; Corver & Van Riemsdijk, 2013). The main distinction is that content words provide the main semantic content of the sentence while function words connect content words (Stubbs, 2002; Ure, 1971). Content words are also called lexical words since they include a lexical value (Chanturidze et al., 2019; Shi et al., 2006), and open-class words since they include most of the vocabulary in languages, and expand in different ways (Abdalla, 2010; Gordon & Caramazza, 1982). Function words are also called grammatical words since they include a grammatical value (Richels et al., 2010; Shi et al., 2006), and closed class words since they include a small number of all the words in languages and are very rarely expanded (Gordon & Caramazza, 1982; Vejdemo & Hörber, 2016).

There is, however, no complete agreement between researchers on the components of the two categories. For many, content words include nouns; verbs; adjectives; and adverbs. The other parts of speech, including prepositions (above), conjunctions (therefore), quantifiers (most), pronouns (those), numbers (five), and determiners (the) are included in function words (Demir-Vegter et al., 2014; Gkalitsiou et al., 2017; Halliday, 1989; Krejtz et al., 2016; Lyons, 1968; Perfetti, 1969).

Lexical density is "the density with which the information is presented" (Halliday, 1989, p. 62) and is calculated as the percentage of content words out of all words featured in spoken or written texts (Berman, 2008; Halliday, 1989; Malvern et al., 2004; Ure, 1971). For example, in the sentence: "The door and the

window are closed", there are three content words (door, window, and closed) out of seven words. Therefore, the lexical density is 42%. Lexical density quantifies the knowledge and ideas of a text: a denser text with a high percentage of content words contains more information than a text with a relatively high percentage of function words (Johansson, 2008; Michel et al., 2007). Since content words convey the bulk of the semantic content of a text (Berman et al., 2011), their use is an important indicator of textual richness and "information packaging" (Johansson 2008). This measure is thus used to compare the linguistic development of speakers and writers in both L1 and L2 (Laurén, 2002; Linnarud, 1976). Specifically, the development of the academic language of L1 and L2 writers can be measured (Snow & Uccelli, 2009).

Lexical diversity

Lexical diversity is "the range and variety of vocabulary deployed in a text by either a speaker or a writer" (McCarthy & Jarvis, 2007, p. 459). It is indicated by the number of different words in a sample of speech or writing of a set length and how far repetition is avoided (Malvern et al., 2004). The greater the range, the higher the lexical diversity (Jarvis, 2013a, 2013b; McCarthy & Jarvis, 2010). Furthermore, lexical diversity is the most common measure in vocabulary testing. This is probably why it is also called lexical richness by some researchers (such as Torruella & Capsada, 2013). Whereas by others, lexical richness is used as a general concept for different lexical measures, such as lexical diversity, lexical density, lexical sophistication, lexical originality, and more (Johansson, 2008). This measure is also known as the lexical range, lexical variability, lexical variation, and lexical variety (Crystal, 1982; Engber, 1995; Jarvis, 2002; Yu, 2010).

High lexical diversity indicates that the speaker or writer has a rich vocabulary and large repertoire, so he does not need to repeat the same types (Berman & Verhoeven, 2002; Johansson, 2008). Lexical diversity is found to be an indicator of the learner's linguistic competence and also of the stage in language acquisition (Jarvis, 2013a; McCarthy & Jarvis, 2010; Nasseri & Thompson, 2021). It also predicts the writing quality (Yu, 2010). A speaker displaying high lexical diversity is judged positively, as he can convey complex messages (Dillard & Pfau, 2002). This capacity is therefore an important measure of linguistic and communicative ability.

One of the simplest indices of lexical diversity is the Type-Token-Ratio (TTR), which is calculated as the number of unique words (types) in a text divided by the total number of words (tokens) in the text (Johnson, 1939). Namely, tokens are all the words in a sample, including repetitions, while each distinct/different word is a type (McKee et al., 2000; Youmans, 1990). For example, in the previous example: "The door and the window are closed", there are seven tokens and six types since the word appears twice. So, the lexical diversity is 85%. Lexical diversity, therefore, increases as the number of types increases or as the number of repetitions decreases (Lieven, 1978; McCarthy & Jarvis, 2007).

However, the main issue of TTR is that it is affected by the length of the text: as the text becomes longer, the lexical diversity decreases (Hess et al., 1986; Jarvis, 2002, 2012; Koizumi, 2012; Linnarud, 1986; Malvern et al., 2004; McCarthy & Jarvis, 2007). This is because as a text becomes longer, there will be a linear and constant increase in the number of tokens (since by definition every additional word

equals a new token), while the increase in types becomes slower as the length increases (McCarthy & Jarvis, 2010). An explanation for this negative correlation is that our types are limited, while our tokens are not - we have a limited number of word types available on a given subject, but can write or say infinite word tokens on the same subject (Schlesinger, 2000). In addition, no text includes more than a handful of words that can be meaningful without repetition (McCarthy & Jarvis, 2010). Many researchers have proposed new indices to solve this issue, including Carroll, 1992; Johnson, 1939; McKee et al., 2000, McCarthy & Jarvis, 2007; Vidal & Jarvis, 2020, but TTR can be used in texts of the same length.

Challenges in applying indices in English and Hebrew and similar languages

The definition of lexical density as the percentage of lexical words out of all words in a text (Ure, 1971) and the definition of lexical diversity as the number of unique words in a text divided by the total number of words raises an initial consideration of what counts as a word. Halliday (1985) preferred to refer to items rather than words. His doctrine sought to address situations where two or more orthographic words are needed to express one meaning or one idea, for example in expressions or verbal phrases such as switch off. This is correct in many languages, including English and Hebrew.

English, like some other languages, is relatively amenable to automatic lexical density and diversity testing of written corpora, since, broadly speaking, each word is orthographically distinct. Hebrew, on the contrary, is a synthetic language, with orthographic properties that make automatic analysis difficult. A primary difficulty is that several lexical or functional items can be combined in one orthographic word: for example, the orthographic Hebrew word שבכפרו (she-be-kfar-o) means "which is in his village" and contains 3 function words and one content word. A second difficulty is that in Hebrew several types can be combined in the same orthographic word. For instance, in the previous example, one orthographic word includes four types (which_is, in, his, village). The third difficulty is the fact that written Modern Hebrew is mainly non-vowelized. Therefore, 25% of the words in Hebrew are ambiguous if written separately. Only context disambiguates many words in any written text. For example, the orthographic word ספר (sfr) has four definitions: two nouns: a book, a barber, and two verbs: to count and to tell a story.

Limitations and gaps of previous research

So far, the vast majority of analyses of lexical density in English and all analyses in Hebrew as L1 (Berman et al., 2011; Sarel, 1999; Schlesinger, 2000) have counted only orthographic units, as an automated analysis was used. So, an orthographic unit that included a function and a content word or two function words was counted as one item only. We shall call this method the conventional approach of lexical density. In contrast, in the proposed adapted approach of lexical density that I suggest, each item is counted separately, even if orthographically incorporated in another orthographic word.

This is especially essential since function words have an important role in the text: they carry grammatical information about the content words. This role is reflected in them whether they are written separately or attached to the word following them. In the conventional approach, the same function word that may

appear in two different ways is counted differently. This language difference has no semantic justification because if two words play the same role, whether semantic or grammatical, then why count them in one case and not in the other? In the adapted approach, the same function word is counted both when it is separated and when it is attached to other words.

Regarding lexical diversity, previous research on Hebrew as L1 counts only one-word type in a single orthographic word, as is the case in English. In both languages, in most cases, types that consist of more than one orthographic word were not included. We shall call this method the conventional approach to lexical diversity. In my proposed adapted approach of lexical diversity, all the different types included in one orthographic word are counted. This is essential since types of function words are the most frequent types in languages, and in Hebrew, many of them are written attached to other types in a single orthographic word.

To conclude, the differences between the two approaches tested are: the conventional approaches count one function, content, or word type in any orthographic word, whereas the adapted approaches count all of the function, content, or word types in any orthographic word or that consists of more than one orthographic word. This study aims to fill these gaps of not fully adapting the two lexical measures for the language tested and evaluating the results' differences between the conventional and adapted approaches.

Methods

Research questions

The research questions are:

1. How will the adapted approaches affect the results of the tested lexical measures?
2. What are the effects of acquisition time length on lexical density and lexical diversity?

Participants

Participants in the current study were Arab high school students in northern Negev, Israel. The 23 participants were all native speakers of Negev Arabic; they all studied in the same advanced-level science-track class, at the same public high school. They all began instructing Hebrew as L2 at the same age and had received the same amount of formal exposure to Hebrew.

Procedure and corpus

Since longitudinal changes in L2 lexical usage can be tested by capturing changes in L2 of the learners (Kim, 2021), the procedure of this study was a test-retest of written essays performed by students in 11th grade and a year later in 12th grade. The participants watched an animated video called "Bridge". The video, which features no spoken dialogue, shows two situations of two animals approaching each other from separate sides of a narrow bridge as they try to cross it. In the first situation, they are not willing to negotiate a compromise, so they both fall over the bridge. In the second, they do negotiate and therefore succeed in crossing. The video was given to help the students start the complex process of writing without, however, providing any linguistic help. It lasted less than three minutes, during which no discussion was conducted.

Afterward, the students were asked to write an argumentative essay dealing with the question: "Are you willing to negotiate and reach a compromise to promote a common goal?". The task was performed under supervision, with the teacher ensuring that no one received any help. All students completed their writing in less than 90 minutes. The essays were not returned to the writers, and no feedback was given. In the final stage of this process, a corpus of 46 essays of approximately 7,000 running words was collected.

Analysis

Since there is no automatic analysis software in Hebrew that allows for an accurate analysis of lexical density and lexical diversity, as I propose, I used one of the digital tools developed in the digital/computational humanities field. The tool is called CATMA (Computer Assisted Text Markup and Analysis) (Gius et al., 2022). Lexical diversity and lexical density were calculated after manual annotation of the words in the essays using the CATMA tool.

The main difference between measuring lexical density and lexical diversity is that the density is based on the word form or the word token. In contrast, the diversity is based on word types and lemmas. Lemmas are made up of a headword and its inflections, all of which have the same part of speech. Most other research on lexical diversity does not indicate which lexical units were used (Treffers-Daller et al., 2018). But following Treffers-Daller et al. 2018; Jarvis & Hashimoto 2021 and others, I indicate that I use lemmas. This measure was tested not only through the TTR formula but also through CTTR, Herdan's index, and Uber's index. All of them were tested at the same length (first 80 tokens of every essay). But the lexical density analysis was conducted over the entire length of each essay since previous research does not indicate any effect of length on lexical density. The analysis excluded non-Hebrew words, written either in Arabic script or (rarely) in English. I follow Zareva (2019) in excluding any lexicon external to the language studied. The proper names of people, countries, companies, and more were also excluded, following, for example, Milton (2009).

Results and Discussion

Lexical density

As one would expect based on the different ways in which they are calculated, at both time points the results of both approaches differ considerably from the conventional approaches ($p<0.0001$) (see Table 1).

Table 1. Means, ranges, and standard deviations of lexical density by approach and time

	Time 1		**Time 2**	
N=23	Conventional Approach	Adapted Approach	Conventional Approach	Adapted Approach
Mean	0.52	0.40	0.52	0.41
Range	0.20	0.15	0.24	0.15
Standard Deviation	0.06	0.04	0.05	0.03
Paired Sample T-Test	**0.0001*** **		**0.0001*** **	
***p <.001**				

It has been found that the most common words in the text are function words not only in Hebrew but also in English and many other languages (Gordon & Caramazza, 1982; Shi et al., 2006; Milton, 2009; Vejdemo & Hörber, 2016). In Hebrew, attached function words are as common as separate function words (Berman & Verhoeven, 2002; Bolozky & Berman, 2020). Therefore, it is appropriate to deal with both groups similarly. The adapted approach ensures the counting of the most common word groups in Hebrew texts.

Since this approach is given a more representative place for the function words compared to the conventional approach, the lexical density level in it is lower than the conventional approach, as shown in Figure 1:

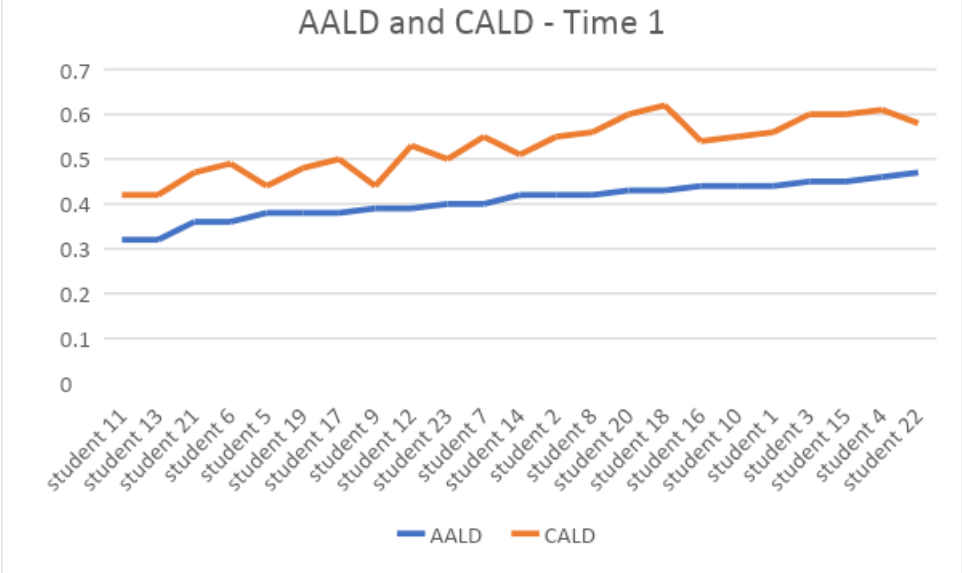


Figure 1. The adapted- and conventional approaches of lexical density at time 1

It has been found in previous studies on English that a lexical density of over 50% is usually produced in writing by native learners and not L2 learners (Read, 2000). According to this, the adapted approach indicates values more similar to findings in English than the conventional approach. That is, the use of the adapted approach allows for cross-linguistic comparison, unlike the conventional approach which limits the comparison to one language only.

Regarding the acquisition time length effect, the mean of the values for the adapted approach at the two-time points is 40% and 41% respectively, showing a slight increase between periods. It is interesting to note, however, that the mean of the conventional approach values at both time points is 52%, with no increase whatsoever. The differences are not statistically significant (Paired Sample T-Test ($p > .05$; .2002)) with a small effect size, Cohen's D = 5.0 <.2002> 2.0)).

While we generally might expect an increase in lexical density over time, studies often show it to be stable. Therefore, this result is not particularly surprising, nor is it necessarily evidence that the students didn't improve their lexical competence. It should be noted that the lack of increase in lexical density over time can be attributed to a sharp decrease in only three students' essays. These three experienced a relatively dramatic decrease against our expectations of vocabulary

development. According to the Paired Sample T-Test, if these essays were not included in the analysis, there would be a statistically significant increase in the lexical density level of the rest of the class between the 11th and 12th grades.

But if we examine the data from a different perspective, we may discover interesting findings. An important feature of lexical density is the ability to distinguish between different language styles, especially between spoken language and written language (Halliday, 1985; Laurén, 2002). According to Ure (1971), the average lexical density in oral texts is less than 40% and ranges from 24% to 43%, while in written texts the average lexical density is higher than 40% and ranges from 36% to 57% (O'Loughlin, 1995).

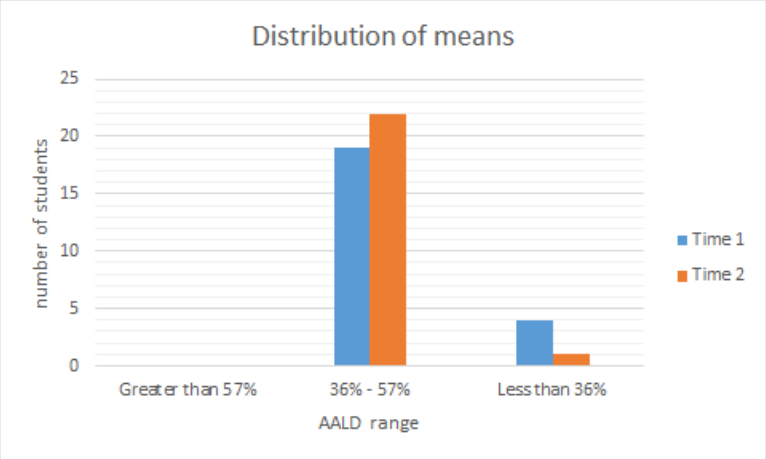


Figure 2. Distribution of means of the adapted approach of lexical density

As can be seen in Figure 1, almost all the essays fall within that range at both time points (according to the adapted approach). However, it is notable that while at the beginning of the year, four out of the 23 essays were below 36%, by the second time point only a single student fell below that range. In other words, over time, more of the class falls within the range of lexical density that has been claimed to be typical of written language. This figure shows the general progress of the whole class in terms of lexical density, which we cannot see if we examine each student individually.

If we examine the results of the lexical density according to the conventional approach at the two-time points, we see that six essays indicate values higher than 67%. Whereas zero essays, according to the adapted approach. The tendency is clear: written Hebrew essays tested according to the conventional approach indicate very high levels of lexical density, higher than average written texts in L1. The fact that these texts were written by L2 learners who have not reached an advanced level of proficiency weakens the validity of this approach since very high values are expected of L2 earners with a very high level of competence, as well as of native speakers.

Lexical diversity

According to the Paired Sample T-Test, the difference between the results of the two approaches of calculating lexical diversity according to the four indices:

TTR; Herdan's index; CTTR; Uber index is very statistically significant ($p < .01$; .0000) (see Table 1 for TTR and CTTR).

Table 2. Means, ranges and standard deviations of lexical diversity by approach and time

Time 2				**Time 1**				
Adapted Approach		Conventional Approach		Adapted Approach		Conventional Approach		N=23
CTTR	TTR	CTTR	TTR	CTTR	TTR	CTTR	TTR	
3.58	0.50	3.65	0.58	3.56	0.50	3.64	0.58	Mean
1.39	0.20	1.71	0.27	1.68	0.24	1.91	0.32	Range
0.41	0.06	0.48	0.08	0.42	0.06	0.47	0.08	Standard Deviation
of CTTR		of TTR:		of CTTR:		of TTR:		Paired Sample T-Test
0.0001* **		**0.0001*** **		**0.0001*** **		**0.0001*** **		
***p <.001								

This finding is expected because, in the adapted approach, attached function types are counted, whereas, in the conventional approach, they are not. Although the number of these types is small in any language (Gordon & Caramazza, 1982; Vejdemo & Hörber, 2016), their frequency is very high not only in English (Shi et al., 2006) but also in Hebrew (Bolozky & Berman, 2020) and many other languages. Of course, frequency is a fundamental issue in examining lexical diversity. Therefore, counting them allows for a more accurate examination of the diversity of the learner's vocabulary.

Regarding the acquisition time length effect, in both approaches of calculating the lexical diversity - two indices (TTR; Herdan's index) do not indicate any progress in lexical diversity even after a year of learning Hebrew as L2 to Arabic speakers. Whereas the other two indices (CTTR; Uber index) indicate a small and slow increase that is not statistically significant. It is worth mentioning that in all indices there was a decrease in the range. At time point 2, the averages became more concentrated. Meaning, in the 12th grade, the lexical diversity averages are similar to each other, more than in the 11th grade. Similar findings have been reported in various studies, including Kalantari & Gholami, 2017, Laufer, 1991, and Vidal & Jarvis, 2020.

This lack of significant increase has two possible explanations: the first one is that an essential part of the acquisition of words in a language occurs during the initial stages of learning. Therefore, at fundamental levels, a gap of one year is enough to assess a change in the lexicon, while a more extended period is needed at advanced levels. It has been found that lexicon development patterns are not always prominent if one does not look at long-term development (Abu-Rabiah, 2017, 2020a, 2020b, 2022). Second, in some studies, older students use their writing within a limited but very familiar vocabulary or "playing it safe", which lowers the lexical diversity in their writing (Laufer, 1991).

Although the general results show no significant increase in the lexical diversity of all students, a more focused examination of the data reveals an important finding. There is a statistically significant increase in lexical diversity in the essays of the five students with the lowest level of diversity in 11th grade. These students significantly improved their lexical diversity throughout the year.

Table 3. Means, ranges, and standard deviations of lexical diversity by approach and time

Time 2				**Time 1**				
Adapted Approach				Adapted Approach				N=5 (st.2; st.5; st.7; st.12; st.19)
Uber index	Herdan's index	CTTR	TTR	Uber index	Herdan's index	CTTR	TTR	
13.07	0.84	3.49	0.49	10.45	0.80	2.93	0.40	Mean
2.44	0.03	0.46	0.08	3.19	0.05	0.83	0.1	Range
1.12	0.01	0.19	0.03	1.21	0.01	0.31	0.03	Standard Deviation
of Uber index: **0.01***		of Herdan's index: **0.00****		of CTTR: **0.02***		of TTR: **0.01***		Paired Sample T-Test

***p <.05**
****p <.01**

They tried harder than the rest of the class to enrich their vocabulary. They realized that their vocabulary was scant compared to the other students from the teachers' feedback on the essays they were required to write during the year. The same finding of individual progress was documented in previous studies (Laufer, 1991).

Regression toward the mean - the common phenomenon in statistical measurements - can also explain this increase. This is because when testing the same index (lexical diversity in this case) of the same population at different time intervals, there is a probability that the participants with high levels of lexical diversity at the first time point will have lower levels of lexical diversity the second time. In contrast, the students with low levels of lexical diversity at the first time point will have higher levels of lexical diversity at the second time point (Abu-Rabiah, 2022). It is a regression of the two extremes toward the mean.

A third explanation for this increase is that there is a broader possibility of improvement in a measure/skill/ability at a low-performance level than at a high-performance level. That is, students with a low level of lexical diversity have a greater scope for enriching their lexical diversity. In contrast, this possibility is more limited for those with a high level of lexical diversity already (Abu-Rabiah, 2022).

Conclusion

Although testing the vocabulary development reveals no significant increase in lexical density and lexical diversity, over one year of learning Hebrew as L2 to Arabic, various patterns for lexical development have been observed. First, the students with the lowest level of lexical diversity at time point 1 significantly improved their lexical diversity at time point 2. Second, more student essays fell within the lexical density range of a typically written language at the second time point. This examination of Hebrew vocabulary according to two approaches of lexical density and lexical diversity, one that is similar to English, and another which is adapted to Hebrew linguistic features, found statistically significant differences between them. The conventional approaches show very high values that are mainly expected in the written language of skilled native speakers, while the

adapted approach shows lower values that are more typical of second language learners. Moreover, the adapted approach even allows for the comparison of research between Hebrew and other languages and does not limit the comparison to studies conducted in the Hebrew language only.

This study comprised a relatively small number of participants (n=23) because all students who did not complete the task at both time points were excluded. The two major implications are: first before any test of a lexical index or a linguistic phenomenon in a new language, the principles on which it is based must be thoroughly understood and adapted accordingly to the language being tested. By doing this, there is a greater chance of comparing different languages since we are adequately testing the same index or phenomenon. It is not the case that if the technical implementation is similar between the languages, it is more accurate and better reflects the principles of the index. Second, a period greater than one year is necessary to measure lexical development in a second language, especially in the intermediate stages, and an extended period of two or more years is recommended for the examination of the learners' writing.

This study leads to two possible future directions. First, a comparison that assesses vocabulary development in heavily inflected languages like Hebrew concerning analytical languages, with a focus on differences between lexical units: word forms, lemmas, flemmas, and word families (Jarvis & Hashimoto, 2021). Second, distinguishing between proficiency levels based on the lexical units in such heavily inflected languages.

Acknowledgements

This work is based on the author's Ph.D. dissertation: Acquisition of Hebrew lexicon among Negev Bedouin high school students: An empirical account at Ben-Gurion University of the Negev, under the supervision of Roni Henkin-Roitfarb and Roey J. Gafter, to whom he is very grateful. This research was supported by the Ministry of Science and Technology of Israel [scholarship no. 3-16617], and by Negev Southern Wind scholarship, The Kreitman School of Advanced Graduate Studies, Ben-Gurion University of the Negev.

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.

Abu-Rabiah, E. (2017). *Lexical diversity, lexical density and abstract language as measures of acquisition by Negev Bedouin students (15-25) of Hebrew as L2 to Arabic* (Master's thesis). Ben Gurion University of the Negev, Israel. https://doi.org/10.13140/RG.2.2.25444.12167

Abu-Rabiah, E. (2020a). 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. (2020b). The level of abstractness, lexical diversity and lexical density as measures of progress in the acquisition of Hebrew as a second language for Arabic speakers. *Oryanut ve-safa [Literacy and Language], 8*, 71–93.

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, Israel.

Alqahtani, M. (2015). The importance of vocabulary in language learning and how to be taught. *International Journal of Teaching and Education, 3*(3), 21–24.

Berman, R. A. (2008). The psycholinguistics of developing text construction. *Journal of child language, 35*(4), 735–771.

Berman, R., & Verhoeven, L. (2002). Cross-linguistic perspectives on the development of text-production abilities: Speech and writing. *Written Language & Literacy, 5*(1), 1–43.

Berman, R., Nayditz, R., & Ravid, D. (2011). Linguistic diagnostics of written texts in two school-age populations. *Written Language & Literacy, 14*(2), 161–187.

Bolozky, S., & Berman, R. A. (2020). Parts of speech categories in the lexicon of Modern Hebrew. In R. A. Berman (Ed.), *Usage-based studies in modern Hebrew: Background, morpho-lexicon, and Syntax* (pp. 265–330). Amsterdam: John Benjamins Publishing Company.

Cameron, L. (2001). *Teaching languages to young learners*. Cambridge: Cambridge University Press.

Cann, R. (2000). Functional versus lexical: A cognitive dichotomy. In R. D. Borsley (Ed.), *The nature and function of syntactic categories* (pp. 37–78). San Diego, CA: Academic Press.

Carroll, S.E. (1992). On cognates. *Second Language Research, 8*(2), 93–119.

Chanturidze, M., Carroll, R., & Ruigendijk, E. (2019). Prepositions as a hybrid between lexical and functional category: Evidence from an ERP study on German sentence processing. *Journal of Neurolinguistics, 52*(4). https://doi.org/10.1016/j.jneuroling.2019.100857

Corver, N., & Van Riemsdijk, H. (Eds.). (2013). *Semi-lexical categories: The function of content words and the content of function words*. Berlin: Walter de Gruyter.

Crystal, D. (1982). *Profiling linguistic disability*. London: Edward Arnold.

Dakhi, S., & Fitria, T. N. (2019). The principles and the teaching of English vocabulary: A review. *Journal of English teaching, 5*(1), 15–25.

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*. Thousand Oaks: Sage Publications. https://doi.org/10.4135/9781412976046

Engber, C. A. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of second language writing, 4*(2), 139–155.

Gius, E., Meister, J. C., Meister, M., Petris, M., Bruck, C., Jacke, J., Schumacher, M., Gerstorfer, D., Flüh, M., & Horstmann, J. (2022, January). *Catma*. Zenodo. https://doi.org/10.5281/zenodo.6046763

Gkalitsiou, Z., Byrd, C. T., Bedore, L. M., & Taliancich-Klinger, C. L. (2017). Stuttering on function words in bilingual children who stutter: A preliminary study. *Clinical Linguistics & Phonetics, 31*(10), 791–805.

Gordon, B., & Caramazza, A. (1982). Lexical decision for open-and closed-class words: Failure to replicate differential frequency sensitivity. *Brain and Language, 15*(1), 143–160.

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

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

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.

Jarvis, S. (2002). Short texts, best-fitting curves and new measures of lexical diversity. *Language Testing, 19*(1), 57–84.

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. (2013a). Capturing the diversity in lexical diversity. *Language Learning, 63*(1), 87–106.

Jarvis, S. (2013b). Defining and measuring lexical diversity. In S. Jarvis & M. Daller (Eds.), *Vocabulary knowledge: Human ratings and automated measures* (pp. 13–34). Amsterdam: John Benjamins.

Jarvis, S., & Hashimoto, B. J. (2021). How operationalizations of word types affect measures of lexical diversity. *International Journal of Learner Corpus Research*, *7*(1), 163–194.

Johansson, V. (2008). 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*. New York: Institute of General Semantics.

Kalantari, R., & Gholami, J. (2017). Lexical complexity development from dynamic systems theory perspective: Lexical density, diversity, and sophistication. *International Journal of Instruction, 10*(4), 1–18.

Kim, M. (2021). Considerations and challenges in longitudinal studies of lexical features in L2 writing. *Vocabulary Learning and Instruction, 10*(2), 82–90. https://doi.org/10.7820/vli.v10.2.kim

Klein, W., Klein, W. W., & Wolfgang, K. (1986). *Second language acquisition*. Cambridge: Cambridge University Press.

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. http://dx.doi.org/10.7820/vli.v01.1.koizumi

Krejtz, I., Szarkowska, A., & Łogińska, M. (2016). Reading function and content words in subtitled videos. *Journal of Deaf Studies and Deaf Education, 21*(2), 222–232.

Laufer, B. (1991). The development of L2 lexis in the expression of the advanced learner. *The Modern Language Journal, 75*(4), 440–448.

Laurén, U. (2002). Some lexical features of immersion pupils' oral and written narration. *Working Papers in Linguistics, 50*, 63–78.

Lieven, E. V. M. (1978). Conversations between mothers and young children: Individual differences and their possible implication for the study of child language learning. In N. Waterson & C. E. Snow (Eds.), *The development of communication* (pp. 173–187). Chichester: Wiley.

Linnarud, M. (1976). Lexical density and lexical variation: An analysis of the lexical texture of Swedish students' written work. *Studia Anglica Posnaniensia: An International Review of English Studies, 7*, 45–52.

Linnarud, M. (1986). *Lexis in composition: A performance analysis of Swedish leaners' written English*. Malmo: Liber Forlag.

Malvern, D., Richards, B., Chipere, N. and Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Basingstoke: Palgrave Macmillan.

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.

McKee, G., Malvern, D., & Richards, B. (2000). Measuring vocabulary diversity using dedicated software. *Literary and Linguistic Computing*, 15, 323–337.

Michel, M. C., Kuiken, F., & Vedder, I. (2007). The influence of complexity in monologic versus dialogic tasks in Dutch L2. *International Review of Applied Linguistics in Language Teaching*, *45*(3), 241–259.

Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters.

Nasseri, M., & Thompson, P. (2021). Lexical density and diversity in dissertation abstracts: Revisiting English L1 vs. L2 text differences. *Assessing Writing*, *47*(2), 1-11. https://doi.org/10.1016/j.asw.2020.100511

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

O'Loughlin, K. (1995). Lexical density in candidate output on direct and semi-direct versions of an oral proficiency test. *Language testing*, *12*(2), 217–237.

Ortega, L. (2014). *Understanding second language acquisition*. New York: Routledge.

Perfetti, C. A. (1969). Lexical density and phrase structure depth as variables in sentence retention. *Journal of Verbal Learning and Verbal Behavior*, *8*(6), 719–724.

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

Richels, C., Buhr, A., Conture, E., & Ntourou, K. (2010). Utterance complexity and stuttering on function words in preschool-age children who stutter. *Journal of fluency disorders*, *35*(3), 314–331.

Roche, T., & Harrington, M. (2013). Recognition vocabulary skill as a predictor of academic English performance and academic achievement in English. *Language Testing in Asia*, *3*(12), 133–144.

Sarel, C. (1999). Lexical density in spoken and written discourse. In S. Sharvit (Ed.), *Studies in ancient and modern Hebrew language* (pp. 401–411). Ramat-Gan: Bar-Ilan University.

Schlesinger, J. (2000). *The language of the newspaper: Stylistic characteristics in various sections of the written language of daily newspapers*. Beer-Sheva: Ben-Gurion University of the Negev Press.

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

Shi, R., Werker, J. F., & Cutler, A. (2006). Recognition and representation of function words in English-learning infants. *Infancy, 10*(2), 187–198.

Snow, C. E., & Uccelli, P. (2009). The challenge of academic language. In D. Olson & N. Torrance (Eds.), *The Cambridge handbook of literacy* (pp. 112–133). Cambridge: Cambridge University Press.

Stubbs, M. (2002). *Words and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell.

Torruella, J., & Capsada, R. (2013). Lexical statistics and typological structures: A measure of lexical richness. *Procedia-Social and Behavioral Sciences, 95*, 447–454.

Treffers-Daller, J., Parslow, P., & Williams, S. (2018). Back to basics: How measures of lexical diversity can help discriminate between CEFR levels. *Applied Linguistics, 39*(3), 302–327.

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

Vejdemo, S., & Hörberg, T. (2016). Semantic factors predict the rate of lexical replacement of content words. *PloS one, 11*(1). https://doi.org/10.1371/journal.pone.0147924

Vidal, K., & Jarvis, S. (2020). Effects of English-medium instruction on Spanish students' proficiency and lexical diversity in English. *Language Teaching Research, 24*(5), 568–587.

Youmans, G. (1990). Measuring lexical style and competence: The type-token vocabulary curve. *Style, 24*(4), 584–599.

Yu, G. (2010). Lexical diversity in writing and speaking task performances. *Applied linguistics, 31*(2), 236–259.

Zareva, A. (2019). Lexical complexity of academic presentations: Similarities despite situational differences. *Journal of Second Language Studies, 2*(1), 71–92.