



How many words do you need to speak Arabic? An Arabic vocabulary size test

Journal:	<i>The Language Learning Journal</i>
Manuscript ID	RLLJ-2016-0073
Manuscript Type:	Original Paper
Keywords:	first language, vocabulary size, Arabic speakers, test validity, language proficiency

How many words do you need to speak Arabic? An Arabic vocabulary size test

This study describes the scores which emerge when a vocabulary size test in Arabic is used with 339 native speaking learners at school and university in Saudi Arabia. It is thought that native speaker vocabulary size scores of this kind should provide targets for attainment for learners of Arabic, should inform the writers of course books and teaching materials, and the test itself should allow learners to monitor their progress towards the goals of native-like knowledge and fluency. Educated native speakers of Arabic know about 25,000 words, a total which is large compared with equivalent test scores of native speakers of English. The results also suggest that acquisition increases in speed with age and this is tentatively explained by the highly regular system of morphological derivation which Arabic uses and which, it is thought, is acquired in adolescence. This, again appears different from English where the rate of acquisition appears to decline with age. While the test appears reliable and valid, there are issues surrounding the definition of a word in Arabic and further research into how words are stored, retrieved and processed in Arabic is needed to inform the construction of further tests which might, it is thought, profitably use a more encompassing definition of the lemma as the basis for testing.

Keywords: first language; vocabulary size; Arabic speakers; test validity; language proficiency

Introduction

Vocabulary knowledge is essential for overall language proficiency and underpins our ability to communicate (e.g. Clark 1993; Laufer 1989; Milton 2009; Nation 2001). Recent research has shown that relatively large vocabulary sizes are indispensable to perform successfully in a language, be it a first language (L1) or a second language (L2). For example, Nation (2006) suggests that a vocabulary size of around 8,000- 9000 word families is necessary for L2 learners to comprehend written English texts. Similarly, Milton and Treffers-Daller (2013) argue that monolingual English speakers might need a vocabulary size larger than 10,000 word families for easy comprehension

of university level texts. In language acquisition research, vocabulary size is often used as a proxy for general proficiency, since vocabulary size scores were found to correlate highly with scores on general proficiency tests (Alderson 2005). This ought to imply that both learners and teachers will want to assess vocabulary knowledge in order to understand the progress that learners are making towards their learning goals. Read (1990) notes, a first step to understanding the nature of the task facing language learners, is often to estimate the size of a native speaker's vocabulary as an ideal towards which these learners can aspire.

In English there is now a considerable body of research on the vocabulary size of both native speakers and EFL learners which allow us to set goals for learners who are working towards comprehension and communicative competence. There are also a number of widely used tests of English vocabulary knowledge (e.g. VST, Nation and Beglar 2007; VLT, Nation 1990; X-Lex, Meara and Milton 2003), which allow learners and teachers to chart progress. However, despite the fact Arabic is spoken by millions of users as both a first and a foreign language, there appears to be no standard test of Arabic vocabulary size and, perhaps because of this, an absence too in the literature of the scale of vocabulary knowledge needed for learners if they aspire to have the language competence of native Arabic speakers. The purpose of this paper is therefore to address this need and to present a test of written receptive vocabulary knowledge that can assess the size of the Arabic speaker's lexicon, explain some of the performance characteristics of this test through its application with native speakers, and derive the scale of learning needed for fluency in Arabic.

Vocabulary size estimates

Researchers, over more than 100 years, have reported a number of studies examining learners' written receptive vocabulary knowledge in a variety of native languages.

However, most prevalent are those studies that examine the vocabulary size of native English speakers. A feature of these studies is how widely disparate these estimates are. For example, two early studies, Seashore and Eckerson (1940) and Hartmann (1946), have reported that native English speakers know approximately 155,000 words and 200,000 words, respectively. More recent studies, on the other hand, have suggested that native speakers of English know approximately 60,000-80,000 words (Nagy and Herman 1987). More recently still there are estimates of 17,200 words (Goulden, Nation and Read 1990) and 16,785 words (D'Anna, Zechmeister and Hall 1991). At the heart of this disparity is the absence of consistency in what to count as a word, what to count as knowing a word, and how to construct a good methodology for measuring these factors so a reliable estimate of size can be made. It has taken a century of research in English to resolve these issues and they likewise present real challenges in measuring vocabulary size in Arabic.

One reason for the large disparity in size estimates reported above is lack of clarity over what to count as a word. The earlier, and largest, estimates reported above counted every different form of a word as a different word. Thus, in English *work*, *worked* and *works* would be counted as three different words. These produce an estimate of size which challenges our understanding about how a lexicon so big can be acquired. The most recent estimates have argued that words are not stored and retrieved as separate forms in this way and that it is more appropriate to count some kind of word family: a base form and some or many of its derived and inflected forms. By this method, *work*, *worked* and *works* would be counted as a single word. The larger unit of count this produces, of course, results in a much smaller estimate of size and this goes a long way to explain why the most recent estimates of lexical size are only a fraction of the earlier estimates. There appears to be some consensus in English that counting

1
2
3 vocabulary as some kind of word family is most appropriate and there is research on the
4
5 frequency and regularity of inflections and derivations in English (e.g. Bauer and
6
7 Nation 1993) which allows tools to be systematically constructed and made available
8
9 for testing (e.g. Nation and Beglar 2007).
10
11

12 While there is something like consensus in English, this discussion does raise
13
14 issues as to how best to count words in Arabic. The conclusions from English cannot be
15
16 easily applied to Arabic where word formation is significantly different (Masrai 2016;
17
18 Masrai and Milton 2015). In Arabic, word formation is primarily based on a system
19
20 which consists of consonant roots that interlock with patterns of vowels and certain
21
22 other consonants. There are at least two views which account for how complex forms
23
24 are represented in the Arabic mental lexicon and processed on-line. The first is a
25
26 morpheme-based approach, which proposes that Arabic surface forms comprise a root
27
28 and a word pattern (Cohen 1961; McCarthy 1982). This root and pattern approach
29
30 comprises at least three distinct versions which differ either in terms of the number of
31
32 morphemic units they put forward or in terms of the way surface word forms are
33
34 believed to be created. However, although some variations do exist between the three
35
36 versions of the root and pattern approach, there is an agreed upon unity underlying their
37
38 apparent diversity. Specifically, they all ascribe a morphemic status to the root and the
39
40 pattern.
41
42
43
44

45 The second is the stem-based approach, which is also not a homogenous
46
47 approach, but includes a number of different versions. For example, Heath (2003)
48
49 argues that when lexical representation and morphological derivation are concerned, the
50
51 consonantal root is best 'consigned to oblivion'. However, where lexical processing is
52
53 concerned, "root-like strings are extracted from input representations ... but these
54
55 extracted consonantal sequences do not correspond exactly to the traditionally
56
57
58
59
60

1
2
3 recognised roots, particularly where vowels and semi-vowels are concerned” (Heath
4
5 2003: 126-128). Generally, this overview of the Arabic word formation processes
6
7 provides some insights into the way rules, which are very regular in Arabic, are
8
9 executed to create a large number of new words from a given root. If a learner is good at
10
11 rule formation, and individuals must vary in this, then they can develop a sizeable
12
13 Arabic lexicon. This is because mental lexicon size in Arabic is less a function of the
14
15 number of base words, which is small in Arabic (Habash 2010), and more a function of
16
17 how many rules a learner knows and can apply to the base words in order to generate a
18
19 larger vocabulary (Masrai and Milton 2015).
20
21
22

23 It would seem therefore, that a count involving a base form and some kind of
24
25 word family might, as in English, make best sense in estimating the size of the lexicon
26
27 in Arabic. However, it is not entirely clear yet how this Arabic word family might most
28
29 appropriately be constructed. But given the regularity and productivity of the rules for
30
31 morphemic derivation in Arabic it seems quite likely that the size of the lexicon may be
32
33 large compared to English when measured this way. The construction of the size test in
34
35 Arabic will have to accommodate this.
36
37

38 A second reason for the disparity in the historical estimates of vocabulary size in
39
40 English is what to count as knowing a word. One of the earliest researchers, Kirkpatrick
41
42 (1891) explains that in his estimate he counted words that he recognised and could
43
44 explain and give a definition of, words he just recognised even if he did not know their
45
46 meaning, and words he had never encountered before but thought he could work out a
47
48 meaning for from their etymology. He included no check on the accuracy of his
49
50 answers, but Milton and Treffers-Daller (2013) suggest that in the absence of this and
51
52 among native speakers, estimates might be inflated by as much as 20%. The most recent
53
54 assessments of vocabulary size (e.g. Golden et al. 1990; D’Anna et al. 1991) are
55
56
57
58
59
60

consistent in requiring a more demanding criterion of knowledge than Kirkpatrick used, and require that speakers should demonstrate not just recognition of form but also knowledge of a meaning for that form.

To achieve this, tests can be delivered in multiple choice format (as in Nation and Beglar 2007) where the testees link a word to a correct definition from a choice of four, or as a checklist test (as in Goulden et al. 1990) where testees mark the words they know. Both formats are potentially susceptible to guesswork and in checklist tests this can be controlled for in a number of ways. Goulden et al. add a requirement that a definition should be produced as a check on knowledge. Some checklist tests aimed at foreign language learners particularly include false words and the responses to these is used to adjust the estimate (as in Meara and Milton 2003). Nation and Beglar's multiple choice VST does not adjust for guesswork, although there is evidence this does occur (Gyllstad, Vilkaite and Schmitt 2015), and argues that testees do not guess but can activate subconscious elements of word knowledge. The result is that the multiple choice format must include an element of over-estimation as is recognised by D'Anna et al. in their estimates. Bearing this in mind, checklist tests which can test many items, have a less complicated scoring system, a relatively short administration and scoring time, and can include a systematic adjustment for guesswork appear the more effective format. In this at least, the practices of testing in English can be transferred direct to testing in Arabic. Checklist formats have been used to devise the Arabic test presented here.

A final issue to resolve in creating a test which will usefully cast light on the scale of vocabulary knowledge needed to speak Arabic is how such a test should be constructed. Historically, and before the development of large corpora and frequency lists, tests were based on dictionary sampling techniques. In these types of tests, a

systematic selection of words (e.g. choosing the first word from every tenth page) from a large dictionary was used. From the responses to these words, and using often a rough and ready calculation of the number of words per page, an estimate of the total number of words a person knows can be made. However, this method has several problems which can influence the accuracy of the estimates made. The main issue leading to inconsistency in such estimates is that dictionaries of different sizes have been used (see Goulden et al. 1990 for further review) and the scale of the estimate is dependent on the size of the dictionary. There are also issues related to polysemy and the way this is likely to influence the test sample. Words derived from dictionary sample, it is argued (Anglin 1993), are likely to oversample the more frequent words which are likely to be polysemous and have multiple entries in a dictionary. A sample which favours frequent words, if this is not controlled, is likely to lead to over-estimation.

More recent estimates (e.g. Goulden et al. 1990; Nation and Beglar 2007) use frequency information and make a principled sample of words across frequency bands. The main part of Goulden et al.'s test, for example, samples 50 words across the most frequent 25,000 words in Thorndike and Lorge's (1944) frequency lists. The outcome is an estimate of knowledge of the most frequent 25,000 words but they reflect that knowledge beyond these frequency ranges is so slight that the 25,000 words test is likely to be a good estimate of overall vocabulary size. Again, the lessons from test construction in English can be carried over into Arabic tests and the test reported here is based on frequency information.

The study

The purpose of this paper is twofold. Firstly, it is to explain how a test of vocabulary size in Arabic has been constructed, and make this test available to people who could use it. Secondly, it is to report scores from the test when it is applied to a range of native

speakers. These scores ought to be useful to learners and teachers in setting benchmarks for attainment if learners aspire to be communicative in Arabic. These results also have the capacity to inform us whether the test appears to be valid and working reliably.

This double purpose can be refined into a number of more specific objectives:

- (1) To describe the construction of Arabic vocabulary size test.
- (2) To administer the test to a range of native speakers from 13 years-old learners in intermediate school in Saudi Arabia, to Saudi Arabian undergraduate and postgraduate students.
- (3) To report the scores obtained on the Arabic vocabulary size test to provide benchmarks for learners of Arabic as a foreign language.
- (4) To check the test is performing reliably by calculating:
 - a. Cronbach's Alpha scores between the parallel forms of the test, and
 - b. testing for internal consistency by calculating Cronbach's Alpha scores within one form of the test.
- (5) To check the performance of the test appears valid which should display
 - a. discrimination of vocabulary size among the different age groups,
 - b. frequency profiles for learners' vocabulary knowledge, and
 - c. a single construct.

The test can be argued to have good validity if the results can show vocabulary growth across levels. We should expect to see incremental growth of vocabulary over the course of learning and as age increases. If the test is working properly then we should also expect to see a frequency profile in the results (as demonstrated in Milton and Treffer-Daller 2013) where there is a tendency for learner to demonstrate greatest knowledge in the most frequent lexical bands and less knowledge in the less frequent bands. If the test is testing lexical knowledge and nothing else, then we should expect

factor analysis to reveal a single factor where results across the frequency bands are compared. There is a possibility, because of the way words in Arabic are constructed with extensive rule formation from a small number of base forms, that knowledge of the less frequent ranges of words might be tapping into something like syntactic rather than lexical knowledge and factor analysis will help examine this.

The source of test words in Arabic

To construct a vocabulary size test using frequency data and similar in construction to the most recent tests in English (e.g. VST, Nation and Beglar 2007) a reliable source of frequency data in Arabic must be used. The frequency lists must be derived from a well-constructed corpus and must include a clear definition of its unit of count: how it is defining a word in Arabic. Two sources were found which adhere to these requirements. One is the frequency dictionary of Arabic core vocabulary for learners (Buckwalter and Parkinson 2011), and the other is a web-based corpus of Arabic (Sharoff 2006).

Buckwalter and Parkinson’s dictionary of Arabic was developed using a corpus of around 30 million tokens, and includes the most frequent 5,000 lemmas. Although this dictionary can be a very useful source for the learners of Arabic, it includes words list that is too small to fulfil the need to create a test to measure the vocabulary size of the native speakers of Arabic, and potentially of Arabic second language learners too. Therefore, the Arabic web-based corpus (Sharoff 2006) was used as the source of the target words included in the Arabic-Lex.

This corpus includes around 180 million tokens from different domains, including natural sciences, applied sciences, social sciences, politics and arts, and ought to be representative of the words used in Arabic. From this corpus, a list of around 100,000 most frequent words, has been generated by Kilgariff et al. (2011), and is electronically available. This list is thought to be large enough to perform a systematic

word sampling for the intended test. The words included in the frequency list were clearly defined. The concept ‘lemma’ was used in the process of generating the list. Lemma was defined, similar to English, to include a headword and its most frequent inflection, where this process should not involve changing the part of speech from that of the headword, but one or more clitics can be attached to the beginning and the end of the headword (i.e., the conjunction letter *wa* ‘and’; the definite article *al* ‘the’) (Sawalha and Atwell 2011).

The Arabic vocabulary size test

The test is designed to measure the written receptive vocabulary knowledge of Arabic speakers from the first 50, 1000-word frequency levels and is designed as a checklist test. In the literature of vocabulary testing there is no agreement of the sampling rate used to create vocabulary size tests. For example, Goulden et al. (1990) used the sampling rate of 1:500 to test English native speakers. Nation and Beglar (2007), on the other hand, used the sampling rate of 1:100 in the creation of the VST. According to Gyllstad et al. (2015), a sampling rate of higher than 1:100 is probably needed in order to better represent the underlying population of words in the corresponding frequency bands in L2 tests. They are, however, referring to words in English and words in Arabic are rather different because it is such a highly inflected language. A sampling rate greater than 1:100 seems likely to create a test that is too long to be practical; over 500 test items. Thus, for practical reasons, the sampling rate of the Arabic test was 1:500, as in Goulden et al. (1990).

Implementing this sampling rate, a checklist test of Arabic was constructed drawn from the first 50,000 items in Sharoff’s (2006) lemmatised frequency list and comprises 120 test items. The test was divided into five equal 10,000-word frequency bands, each included a sample of 20 words. The 20 words were a principled sample

with two words selected from each 1000 frequency level. The test, further, included 20 pseudowords (non-words) which were intermixed in the test to adjust for guesswork when calculating the final score as in Meara and Milton’s X-Lex (2003). The testees are requested to make a decision whether they know each of these test words. They respond *Yes* to words they do know and *No* to words they do not. The test was designed to take no longer than 10-15 minutes to complete. This short duration encourages the participants to respond to all items in the test without being influenced by the test length. Two forms of the Arabic-Lex test, A and B, were created (copies of the test are provided in appendices 1 and 2). In these two forms, the frequency bands were organised from the most frequent band to the least frequent band (right to left).

In order to make a calculation out of 50,000 of the testee’s vocabulary size, each *Yes* response to a real word in the test is given a score of 500 to form an unadjusted vocabulary size score, and each false *Yes* response to a false word deducts 2,500 points from the unadjusted score to give an adjusted vocabulary size score. This final, adjusted, score is expected to represent a learner’s total vocabulary knowledge.

Methods

Participants

A total of 339 participants (185 male and 154 female), representing eight different levels, took part in the study. Each level is a different age group, ranging from school learners to university students. Information on the levels and groups is summarised in Table 1.

<Table 1 about here>

The first, second and third groups (labelled as years 7, 8 and 9) comprise the participants from the intermediate level in Saudi schools. The number of the participants

in these age groups were 43, 42 and 37, respectively. The fourth, fifth and sixth groups (labelled as years 10, 11 and 12) represent participants from the high school level. The number of the participants in these age groups were 37, 47 and 48, respectively. The 53 participants in the seventh group were at the undergraduate level. Finally, the 32 participants in group eight were Arabic students from different Arab countries (Saudi Arabia, Kuwait, Egypt, Sudan, Libya and Jordan) pursuing their MA and PhD degrees at Swansea University.

Procedure

Form A of the test was administered to the 339 participants during the usual school and university classes in pencil and paper format. Participants were given unique codes to use instead of their real identity and were asked to note their age on the test forms. The participants were told that there was no time limit to finish the test, although each form of the test was expected to take approximately 10-15 minutes to complete. They were also advised that the test includes non-words and false responses to these words are penalised. Written instructions in Arabic were also provided on the test forms.

Test B was also administered to the 32 of the postgraduate students so that parallel forms reliability testing could be carried out. The presence of the Postgraduate students was solely to allow a parallel forms test of reliability to be carried out.

Results

Arabic vocabulary size scores

The mean vocabulary size scores for schoolchildren and undergraduate students are summarised in Table 2.

<Table 2 about here >

Reliability calculations

The scores from the parallel forms tests, and the two split-halves internal consistency test conducted on test A, are summarised in Table 3.

<Table 3 about here>

The reliability of the Arabic-Lex test was tested by comparing the scores obtained from the two parallel forms of the test, A and B. The results indicate a strong and statistically significant correlation ($r = .91, p < .001$). The two forms of the test produce mean scores that are very similar in size.

Internal consistency reliability was further investigated within test A. A split-halves method (e.g. Bachman 2004) was utilised. Thus, form A was divided into odds-evens and top-bottom pairs. The results show high levels of correlations between odds-evens ($r = .89, p < .001$) and top-bottom ($r = .80, p < .001$). The mean scores in these two tests are, again, very similar.

The differences between mean scores were calculated for the three pairs and the differences were not found significant (see Table 4).

<Table 4 about here>

The final reliability measure for the Arabic-Lex test was performing Cronbach's Alpha analysis. The result shows high reliability indices for all the test pairs (all were above .80). The Cronbach's Alpha for the parallel forms, odds-evens and top-bottom halves are shown in Table 4. According to DeVellis (2003), the Alpha score is very good when it falls between .80 and .90 and excellent when it is above .90. These findings provide support for the test's reliability to estimates test-takers' performances in a consistent manner.

Does the test distinguish test-takers' vocabulary size at different age levels?

It is suggested that if the test is valid and performing well then vocabulary size should increase with age. Table 2 shows the descriptive statistics for the performance of the seven groups on the test and these indicate that the mean vocabulary sizes for each of the seven groups increase with age. There is incremental growth in the test-takers' written receptive vocabulary knowledge, from lower age and level groups to higher groups.

To examine the data more closely, a one-way ANOVA was performed to assess the differences between groups. The assumption that the higher groups would perform better than the lower groups on the test was supported except that the variances among the groups were unequal, Levene's test ($6, 300$) = 26.834 , $p < .001$; thus, the Welch and Brown Forsythe tests were utilised. As both tests were statistically significant, only the results of the Welch test are reported. The ANOVA was significant, $F(6, 130.033) = 93.033$, $p < .001$, and Dunnett's T3 post-hoc indicated that all pairwise comparisons were significant at $p < .05$, except between two groups (years 10 and 11), which did not differ significantly. Thus, it may be concluded that the test is capable of distinguishing the vocabulary sizes of test-takers of different ages.

Does the test produce the expected frequency profiles?

In the development of vocabulary size tests, it is generally assumed that items at successively decreasing frequency levels will form a continuum (e.g. Beglar 2010; Milton 2009, Read 2000). Thus, the assumption for the item hierarchies is that higher-frequency words will be learned than lower-frequency words and that scores will diminish as frequency diminishes. This forms the frequency profile described in L2 language learners by Meara (1992) and among L1 speakers by Milton and Treffers-

Daller (2013). The frequency profile is observable at each level examined and this information is summarised in Figure 1. The frequency profile moves upwards, indicating more knowledge in each of the frequency bands, with each successive year of age.

<Figure 1 about here>

A Freidman analysis confirms that the observable trend in the summary data reflects the trends that can be seen in the individual pieces of data ($\chi^2=1019.45, p < .001$). Mean rank scores for each frequency band are presented in Table 5. This result, generally, suggests that the test is sensitive in demonstrating the vocabulary knowledge profiles of test-takers, based on word frequency difficulty order.

<Table 5 about here>

Does the test measure a single construct?

Construct validity is concerned with the dimensionality of the measured construct (Messick 1989). The purpose of a vocabulary size test is to measure vocabulary and not other qualities of language knowledge and performance. It was thought that, potentially, two dimensions of knowledge, lexical and syntactic, might be visible in the results because of the elaborate rule or pattern-based word formation processes that Arabic displays. It was thought that if these two factors were at play then the syntactic knowledge would be observable in the less frequent bands when the results are analysed using a Principle Axis Factoring (PAF).

The appropriateness of the data for factor analysis was assessed prior to performing PAF. An examination of the correlation matrix specified the presence of minimum coefficients of .55 and above. The Kaiser-Meyer-Oklin (KMO) measure of sampling adequacy value was .86, which is above the recommended value of .6 (Pallant

2007). Additionally, Bartlett's Test of Sphericity was statistically significant ($p < .001$). Thus, the factorability of the correlation matrix is supported by these indices.

The PAF analysis indicates that there is only one eigenvalue greater than 1, which explained some 78.53 percent of the total variance, and this suggests that the test is measuring only one factor which, in line with other tests of this type, is believed to be passive receptive vocabulary size in Arabic. The factor loadings and Scree plot summarising this analysis are shown in Table 6 and Figure 2, respectively. The Scree plot clearly shows that there is a sharp drop between the first and the other four factors, which with the amount of variance explained by the first factor provide evidence of a single construct underlying the test. The PAF extraction method of only one factor also reveals that all items show a high level of loading on the factor. The results show that all factor loadings were greater than .7, which is above the minimum baseline loading proposed by Brown (2006), suggesting that only one factor underlies the scores and that the uniformity of loadings on this factor is high.

<Table 6 about here>

<Figure 2 about here>

Discussion

Native Arabic speaker's vocabulary size scores

A vocabulary size test, which works well and for which there are normalised scores for attainment, ought to be pedagogically useful. This information can be used to set and monitor progress towards learning goals in a crucial area of language knowledge. It can and should also provide essential information to syllabus and course book writers in helping to define the content of these materials, and help identify where such materials may be deficient (e.g. Alsaif and Milton 2012).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

In terms of goal setting, the results for the undergraduate students can be said to set a target for learners who aspire to have the lexical resource available to native speakers. Learners of who aspire to higher study through a foreign language, as indicated in Golden et al. (1990), will probably want and need a vocabulary something like a native speaker's in scale. As measured by the Arabic-Lex test in this study, it appears that learners with this aspiration would need something like 25,000 Arabic words. The figures for the ages below 20 suggest the degree of competence which learners might have attained when they hit other lexical milestones. Knowledge of 10,000 Arabic words suggests learners would have the vocabulary knowledge, and maybe the language competence, of a child of about 13 years old. Knowledge of 20,000 words suggests a learner would have the knowledge of a 17 or 18 year-old.

These numbers suggest that Arabic is rather different from English both in the volumes of vocabulary which adult native speakers acquire, and in the pattern of growth which emerges as these numbers are reached.

The numbers need to be treated with caution, of course, because Arabic and English are different languages with different construction of words. However, the count used in Arabic-Lex works on a unit of count that is very similar to English and other European language and gives very similar coverage figures (Masrai and Milton 2016). Nonetheless, it is notable that a twenty year-old speaking Arabic appears to have a lexicon which may be double the size of a twenty year-old speaking English. Milton and Treffers-Daller estimate undergraduate English speakers to know about 10,000 to 11,000 English words on average. D'Anna et al. (1991) estimate 14,000 to 16,000 words among their undergraduates. This ought to have significant learning implications since it implies that learning Arabic as a foreign language is a much bigger task than learning English as a foreign language. Lexicons of this scale are thought to challenge

our understanding of how these volumes of learning can be explained (e.g. Milton and Treffers-Daller 2013). This need not be the case with Arabic and the pattern of growth suggests why this might be the case.

Learners of English appear to learn 600 to 700 words a year during school education (e.g. Anglin, 1993; Biemiller and Slonim 2001;) and this slows to about 500 words a year by the time speakers are at university. After university the lexicon may stabilise or grow more modestly at 100 to 200 words per year (Wilson et al. forthcoming). It is thought the lexicon develops in relation to Zipf effects. There is a tendency for learners to learn the most frequent words first and these words contribute most to coverage. The more of these words that are attained, therefore, and the more coverage of normal language that they provide, the harder it is to encounter and learn the remaining, less frequent, words. Lexical growth will tail off with age as a consequence. In adolescence, learners of English are known to acquire the less frequent morphological derivations (Milton 2009) but, crucially, because these tend to be infrequent and irregular, these are learned as separate and less frequent items within the lexicon and are counted as separate word forms. The task at this age in English, therefore, is to carry on learning more and more, progressively less frequent, word forms.

Among the Arabic learners in this study a different process is suggested. Arabic learners up to the age of about 12 or 13 appear to learn words at about the same rate of English learners at this age; about 600 to 700 words a year. After this point, far from diminishing, the rate of lexical growth increases to something like 2,500 words a year at university. Our best explanation for this lies in the way Arabic has a smaller resource of base words than English and we suspect that learners of both languages, up to about adolescence, are engaged in learning these base words. In adolescence, once a basis of

word forms is known, Arabic speakers can also turn their attention to the acquisition of morphological derivations but the effect of this learning is quite different. In Arabic these are far more regular and frequent than in English and by learning a comparatively small number of derivational rules a whole class of new words can be generated more widely across the lexicon (Atwell 2008; Bar and Dershowitz 2012; Habash 2010). Learning new words this way leads to exponential rather than incremental growth. This means, in effect, more words are added to the lexicon and this might explain the increasing rate of lexical growth.

This interpretation suggests that the size of the lexicon need not be a barrier to learning Arabic as a foreign language, rather there are different pedagogical issues to be addressed in teaching so learners can acquire these highly productive word formation processes in addition to the base forms that are needed for communication. Teaching word formation rules is advocated in English (e.g. Bauer and Nation 1993; Coxhead 2000) although it is not clear it is really useful given the way words in English seem to be acquired. In our interpretation of the acquisition of Arabic, however, teaching and learning these word formation processes would appear indispensable. It is the word formation and the regularity of derivation which, to a large degree, explains and justifies the high vocabulary size scores of Arabic native speakers, particularly undergraduate students. There is some evidence (Marslen-Wilson 1997) that Arabic speakers process derivational morphemes in the same way it handles inflections and that the lemmatised system used here, appropriate to European languages, may not be so useful in Arabic. While this suggests that the estimates of Arabic speakers' mental lexicon size provided using the Arabic-Lex test are reasonable figures, it also suggests that reconsidering the unit of count might be useful and it draws attention to how little we understand about the storage, retrieval and processing of words in the Arabic mental lexicon. A version of

Arabic-Lex constructed using a bigger, Arabic, lemma and including the derivational affixes which are currently excluded, might give a different insight into the way a lexicon is acquired in Arabic.

Test validity and reliability

These conclusions, and the usefulness of the test to others, depend on being able to argue that the tests demonstrated are working properly and are reliable and valid. It is possible to make a good case that the tests are working well although, because these are first steps in the process of vocabulary testing in Arabic and systematically investigating the way the Arabic mental lexicon works, there are issues which are raised and further questions to be addressed.

Arabic-Lex appears to be a very reliable test. Parallel forms of the test, and the odds and evens internal consistency measure, both produce excellent Alpha scores. The top and bottom half measure is slightly lower, probably explained by the way the bottom half of the test is qualitatively different from the top half in that the items are less frequent and less likely to be known, but at 0.89 the Alpha score is still very, very good. The excellent correlation and Alpha scores for the two parallel forms of the test has important pedagogical implications since this suggests that, provided the right principles are applied, it is easy to generate multiple forms of the test which will behave equivalently. In addition to be reliable, the test has proved itself to be, as Mochida and Harrington (2006) contend, time- and resource-efficient. The test is potentially very useful, therefore, in both formal and informal testing.

As far as we can tell Arabic-Lex also appears valid and appears to be testing what it should be testing. The test results display a frequency effect with the more frequent vocabulary tending to be better known than the less frequent vocabulary and this is in line with our understanding of the way word frequency and learning interact

(e.g. Milton 2009; Nation 2001; Palmer 1917). Arabic-Lex is also able to discriminate between speakers of different age, and therefore we would argue, different ability groups. The cline of scores from youngest to oldest fit with our understanding of vocabulary growth, that it will increase with age up to adulthood. While the statistics support this conclusion it must be borne in mind that there is considerable individual variation and overlap between the age groups. However, variation among individuals is probably one trait of vocabulary measures (Milton 2009).

The results of factor analysis and an examination of the scree plot support the idea that a single primary construct underlies the test, accounting for about 79% of the total variance. All the five frequency levels had a high loading on this factor. Since the test items are focused on vocabulary assessment, it could be concluded that the single underlying construct is written receptive vocabulary knowledge. In a test designed to assess vocabulary size and no other factors, this is as it should be. However, we have argued that the growth of the lexicon beyond about 10,000 words in Arabic is syntactically driven, that the lexicon grows through the application of rules rather than the incremental learning of lexical items. Our factor analysis has not picked this out and we have no clear understanding why this might be the case. Perhaps we are seeing some evidence here of the minimalist hypothesis (McKoon and Ratcliff 1992) that lexical size drives other aspects of language such as syntax, in which case we might expect lexical size and syntactic knowledge development to be collinear. This result may reflect also that we have not yet arrived at a unit of count in Arabic, a definition of the word, which might permit for this kind of insight to be gained.

Conclusions and suggestions for future research

This study has presented figures for the size of the Arabic native speaker’s lexicon and these ought to be pedagogically useful. The figure of 25,000 words sets a standard for

1
2
3 learners to achieve if they aspire to become native like. It is a figure that should
4
5 concentrate the minds of course book and materials creators who will want to develop a
6
7 strategy for the presentation of these words so learners can make progress. The Arabic-
8
9 Lex test presented in this paper fills a gap in the field of vocabulary testing by providing
10
11 a systematic measure of the written receptive vocabulary knowledge of Arabic speakers
12
13 and learners which appears both reliable and valid. Future research may usefully
14
15 consider using the Arabic-Lex test with university students to examine whether it is
16
17 capable of predicting students' academic success and whether the test scores will
18
19 correlate with their achievements.
20
21

22
23 The test has also suggested other insights into the learning process of Arabic. It
24
25 is suggested that a lexicon of 25,000 words is probably driven, in its later stages, by the
26
27 application of word formation rules in adolescence leading to exponential rather than
28
29 incremental growth. The Arabic lexicon appears large compared to English and
30
31 achieving this size appears best explained by these processes rather than by the
32
33 incremental acquisition of words as it is understood in English. The teaching of these
34
35 word formation rules to learners of Arabic as a foreign language would appear
36
37 indispensable to the acquisition of a large Arabic lexicon. Further research into when
38
39 these rules are learned and how they might best be taught to foreign language learners
40
41 of Arabic should usefully be pursued.
42
43

44
45 Arabic-Lex has allowed some insight to be gained, therefore, into the way a
46
47 lexicon in Arabic appears to be acquired. However, much depends on the way the word
48
49 in Arabic has been defined for purposes of counting and in this paper we have used a
50
51 definition of word which fits well with counts made in European languages. While this
52
53 approach has produced a test which is clearly workable and informative, it is not at all
54
55 clear that this is the best or most appropriate way to handle details of the lexicon in
56
57
58
59
60

Arabic. Much of the work which now exists in English, for example on the frequency and regularity of derivational affixes and the way words are stored and processed in the lexicon, still remains to be done in Arabic. Our approaches to testing the lexicon are likely to change in the light of this research and different testing approaches, using different definitions of the word, seem likely to refine our understanding of the learning and teaching process for words in Arabic.

References

Alderson, J.C. 2005. *Diagnosing Foreign Language Proficiency: The Interface between Learning and Assessment*. London: Continuum.

Alsaif, A. and J. Milton. 2012. Vocabulary input from school textbooks as a potential contributor to the small vocabulary uptake gained by English as a foreign language learners in Saudi Arabia. *The Language Learning Journal* 40, no. 1: 21-33.

Anderson, A. and P. Freebody. 1983. Reading comprehension and the assessment and acquisition of word knowledge. *Advances in Reading/Language Research*: 231-256.

Anglin, J. 1993. *Vocabulary Development: A Morphological Analysis*: Monographs of The Society for Research in Child Development. (Serial No. 238). Chicago: University of Chicago Press.

Atwell, E. 2008. Development of tag sets for part-of-speech tagging. In *Corpus Linguistics: An International Handbook*, eds. A. Ludeling and M. Kyto, 501-526. Berlin, Boston: De Gruyter Mouton.

Bachman, L. F. 1990. *Fundamental Considerations in Language Testing*. Shanghai: Shanghai University Press.

Bachman, L. F. 2004. *Statistical Analyses for Language Assessment*. Cambridge: Cambridge University Press.

Bar, K. and N. Dershowitz. 2012. Deriving paraphrases for highly inflected languages from comparable documents. *Proceedings of the COLING*, Mumbai: 185-200.

Bauer, L. and P. Nation. 1993. Word families. *International Journal of Lexicography* 6, no. 4: 253-279.

Beglar, D. 2010. A Rasch-based validation of the vocabulary size test. *Language Testing* 27, no. 1: 101-18.

- 1
2
3 Biemiller, A. and N. Slonim. 2001. Estimating root word vocabulary growth in
4 normative and advantaged populations: Evidence for a common sequence of
5 vocabulary acquisition. *Journal of Educational Psychology* 93: 498-520.
6
7
8 Brown T. A. 2006. *Confirmatory Factor Analysis for Applied Research*. New York:
9 Guilford Press.
10
11 Buckwalter, T. and D. Parkinson. 2011. *A frequency Dictionary of Arabic Core*
12 *Vocabulary for Learners*. London and New York: Routledge.
13
14 Burnett, J. D. 1974. Parallel measurements and the Spearman-Brown formula.
15 *Educational and Psychological Measurement* 34, no. 4: 785-788.
16
17
18 Clark, E. 1993. *The lexicon in Acquisition*. Cambridge: Cambridge University Press.
19
20 Coady, J. 1997. L2 vocabulary acquisition: A synthesis of research. In *Second*
21 *Language Vocabulary Acquisition*, eds. J. Coady and T. Huckin, 225-237. New
22 York: Cambridge University Press.
23
24 Cohen, D. 1961. Le vocabulaire de base sémitique et le classement des dialectes du sud.
25 *Semitica* 11: 55-84.
26
27
28 Coxhead, A. 2000. A new academic word list. *TESOL Quarterly* 34: 213-238.
29
30 D'Anna, C. A., E. B. Zechmeister and J. W. Hall. 1991. Toward a meaningful definition
31 of vocabulary size. *Journal of Literacy Research* 23, no. 1: 109-122.
32
33 DeVellis, R. F. 2003. *Scale Development: Theory and Applications* (Second ed.).
34 Thousand Oaks, California: Sage Publications, Inc.
35
36
37 Gyllstad, H., L. Vilkaite and N. Schmitt. 2015. Assessing vocabulary size through
38 multiple-choice formats: Issues with guessing and sampling rates. *ITL International*
39 *Journal of Applied Linguistics* 166: 276-303.
40
41 Goulden, R., P. Nation and J. Read. 1990. How large can a receptive vocabulary be?
42 *Applied Linguistics* 11, no. 4: 341-363.
43
44 Habash, N. Y. 2010. Introduction to Arabic natural language processing. *Synthesis*
45 *Lectures on Human Language Technologies* 3, no. 1: 1-187.
46
47
48 Hartmann, G. W. 1946. Further evidence on the unexpected large size of recognition
49 vocabularies among college students. *Journal of Educational Psychology* 37, no. 7:
50 436-439.
51
52 Hazenberg, S. and J. Hulstijn. 1996. Defining a minimal receptive second-language
53 vocabulary for non-native university students: An empirical investigation. *Applied*
54 *Linguistics* 17, no. 2: 145-163.
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Heath, J. 2003. Arabic derivational ablaut, processing strategies, and consonantal “roots”. In *Language Processing and Acquisition in Languages of Semitic Root-Based, Morphology*, ed. J. Shimron, 100-115. Amsterdam: John Benjamins.

Holes, C. 1995. *Modern Arabic*. London: Longman.

Jaafar, Y. and K. Bouzoubaa. 2014. Benchmark of Arabic morphological analyzers challenges and solutions. In *Proceedings of the Intelligent Systems: Theories and Applications (SITA-14), 9th International Conference*: 1-6.

Kilgarrieff, A., F. Charalabopoulou, M. Gavrilidou, J. Bondi, S. Khalil, S. Johansson and E. Volodina. 2011. Corpus-based vocabulary lists for language learners for nine languages. *LREJ special issue*.

Kirkpatrick, E. A. 1891. Number of words in an ordinary vocabulary. *Science* 18, no. 446: 107-108.

Laufer, B. 1989. What percentage of text is essential for comprehension? In *Special Language; from Humans Thinking to Thinking Machines*, eds. C. Lauren and M. Nordman, 316-323. Clevedon: Multilingual Matters.

Laufer, B. and P. Nation. 1999. A vocabulary-size test of controlled productive ability. *Language Testing* 16, no. 1: 33-51.

Levelt, W. 1989. *Speaking: from Intention to Articulation*. Cambridge, MA: MIT Press.

Marslen-Wilson, W. D. and L. K. Tyler, L. K. 1997. Dissociating types of mental computation. *Nature Reviews Neuroscience* 387: 592–594.

Masrai, A. 2016. The influence of morphological knowledge on lexical processing and acquisition: The case of Arab EFL learners. *Ampersand* 3: 52-60.

Masrai, A. and J. Milton. 2016. How different is Arabic from other languages? The relationship between word frequency and lexical coverage. *Journal of Applied Linguistics and Language Research* 3, no. 1: 15-35.

Masrai, A. and J. Milton. 2015. The impact of L1 lexical organisation on L2 vocabulary acquisition. *Language in Focus* 1, no. 1: 15-34.

McCarthy, J. 1982. Prosodic templates, morphemic templates, and morphemic tiers. In *The Structure of Phonological Representations*, eds. H. Van der Hulst and N. Smith, 191-223. Dordrecht: Foris Publications.

McKoon, G. and R. Ratcliff. 1992. Inference during reading. *Psychological Review* 99, no. 3: 440-466.

Meara, P. 1992. *EFL vocabulary tests*. Swansea: University College Swansea: Centre for Applied Language Studies.

- 1
2
3 Meara, P. and B. Buxton. 1987. An alternative to multiple choice vocabulary tests.
4 *Language Testing* 4, no. 2: 142-154.
5
6 Meara, P. and J. Milton. 2003. *The Swansea levels test*. Newbury: Express.
7
8 Messick, S. 1989. Validity. In *Educational measurement*. 3rd edition, ed. R.L. Linn, 13-
9 103. New York: Macmillan.
10
11 Messick, S. 1995. Validity of psychological assessment: Validation of inferences from
12 persons' responses and performances as scientific inquiry into score meaning.
13 *American Psychologist* 50: 741-749.
14
15 Milton, J. 2009. *Measuring Second Language Vocabulary Acquisition*. Bristol:
16 Multilingual Matters.
17
18 Milton, J. and J. Treffers-Daller. 2013. Vocabulary size revisited: the link between
19 vocabulary size and academic achievement. *Applied Linguistics Review* 4, no. 1:
20 151-172.
21
22 Mochida, K. and M. Harrington. 2006. The yes/no test as a measure of receptive
23 vocabulary knowledge. *Language Testing* 23, no. 1: 73-98.
24
25 Nagy, W. and P. Herman. 1987. Breadth and depth vocabulary knowledge: implications
26 for acquisition and instruction. In *The Nature of Vocabulary Acquisition*, eds. M.
27 McKeown and M. Curtis, 19-36). Hillsdale: NJ; Lawrence Erlbaum Associates.
28
29 Nation, P. 1990. *Teaching and Learning Vocabulary*. New York: Heinle and Heinle
30 Publishers.
31
32 Nation, P. 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge
33 University Press.
34
35 Nation, P. 2006. How large vocabulary is needed for reading and listening? *The*
36 *Canadian Modern Language Review* 63: 59-82.
37
38 Nation, P. and D. Beglar. 2007. A vocabulary size test. *The Language Teacher* 31, no.
39 7: 9-12.
40
41 Pallant, J. 2007. *SPSS Survival Manual*. Maidenhead: Open University Press.
42
43 Palmer, H. E. 1917. *The Scientific Study and Teaching of Languages*. London: Harrap.
44
45 Pellicer-Sánchez, A. and N. Schmitt. 2012. Scoring Yes–No vocabulary tests: Reaction
46 time vs. nonword approaches. *Language Testing* 29, no. 4: 489-509.
47
48 Read, J. 2000. *Assessing Vocabulary*. Cambridge: Cambridge University Press.
49
50 Sawalha, M. and E. Atwell. 2011. *Accelerating the processing of large corpora: Using*
51 *grid computing technologies for lemmatizing 176 million words Arabic internet*
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

corpus. Paper presented at the Advanced Research Computing Open Event. University of Leeds, Leeds, UK.

Schmitt, N., D. Schmitt and C. Clapham. 2001. Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing* 18, no. 1: 55-88.

Seashore, R. and L. Eckerson. 1940. The measurement of individual differences in general English vocabulary. *Journal of Educational Psychology* 31: 14-38.

Sharoff, S. 2006. Open-source corpora: Using the net to fish for linguistic data. *International Journal of Corpus Linguistics* 11, no. 4: 435-462.

Thorndike, E. and I. Lorge. 1944. *The Teacher's Word Book Offered 30,000 Words*: New York: Teachers College, Columbia University.

Vermeer, A. 2004. The relation between lexical richness and vocabulary size in Dutch L1 and L2 children. In *Vocabulary in Second Language*, eds. P. Bogaard and B. Laufer, 173-189). Amsterdam: Benjamin's Publishing Company.

Versteegh, K. 1997. *The Arabic Language*. Edinburgh: Edinburgh University Press.

Wilson, S., A. Thorne, M. Stephens, J. Ryan, S. Moore, J. Milton and G. Brayley. forthcoming. English vocabulary size in adults.

Appendix 1. Arabic-Lex A – student version

Arabic-Lex Vocabulary Size Test A (50K)

Please look at these words. Some of these words are real Arabic words and some are invented but are made to look like real words. Please tick the words that you know or can use. Here is an example.

book ✓

Thank you for your help.
من فضلك انظر الى هذه الكلمات. بعض هذه الكلمات هي كلمات عربية حقيقية والبعض الآخر كلمات تم تركيبها لتبدو وكأنها كلمات حقيقية. فضلاً اختر الكلمات التي تعرفها أو تستطيع استخدامها. استخدم هذا المثال.

✓ كتاب

شكراً لمساعدتكم،

نشر	أكاد	حُكِمَ	تركب	استمرت
عُفِّ	زَنَقَ	اختر لهم	انقضى	تسرع
لَقَّ	تقارق	تصور منا	استبهاض	ناقد
شبهة	مخفل	لرغي	الغاض	الاسمى
سائق	مخاف	أعفى	سائق	امكانيه
تفوح	فتراف	مناقصة	القصند	مترار
جزو	منطى	منطوب	يوضح	الاقباء
عمل	عنمن	منزلهم	البيخاق	النشر
الاربعاء	أغير	استدعي	الانبياء	اوصاعها
خمر	اطاق	خافق	أخر	عقاد
مخاور	ضام	أخوض	تهشم	تحققه
قفر	اختمالك	سكيم	خروج	إعجاب
عطف	كثيرة	القرى	رميت	ألقها
مترقى	أزجركم	مؤجى	تفيض	شذراج
تجول	خرف	احترامهم	لايكفى	اواجه
تكفل	يشافون	الفضائل	مؤسسية	نظام
تالق	مستحضرات	مهالك	ذليعا	الافنديت
أختر اذ	كتاب	زواج	اتصال	الزلام
الغدير	مائدة	يشاد	أجتمعا	تلقا
صنم	يخافط	صناعة	أفصاه	مقطب
كلا	التخال	الإنتقام	أرشح	خلاف
فر	القام	يلكمه	مستائي	نغرة
استخدامات	الأسرة	تغير لهم	اشقى	استطاب
إجماع	مناهجات	صانج	إندراج	تغريز

Appendix 2. Arabic-Lex B – student version

Arabic-Lex Vocabulary Size Test B (50K)					
Please look at these words. Some of these words are real Arabic words and some are invented but are made to look like real words. Please tick the words that you know or can use. Here is an example.					
Thank you for your help.					
من فضلك انظر الى هذه الكلمات. بعض هذه الكلمات هي كلمات عربية حقيقية والبعض الآخر كلمات تم تركيبها لتبدو وكأنها كلمات حقيقية. فضلاً اختر الكلمات التي تعرفها أو تستطيع استخدامها. استخدم هذا المثال.					
✓ كتاب					
book ✓					
شكراً لمساعدتكم،					
شع	تتكلف	غبار	الأفات	ألوف	
تجاذوزها	مخطط	اختراتي	احتضان	زائع	
هوى	منقي	صنفل	غابر	أراد	
تقرض	الآلانية	أينل	عائق	بيجل	
مجردة	تثاغي	قنار	طعم	صمود	
فلوح	غلاب	أسر داد	قفر	نكم	
منابحه	تغفل	الأفدس	قسنينة	علماني	
ينزدد	مصنوف	الغالي	يخضع	أيد	
افتر اجنبي	الإستبازات	الإحتياجات	أشد	منني	
بقهاء	خملأ	نشف	عبد	وضوء	
النوائج	غقب	تستهدف	ضغوط	صفر	
شكور	صاجر	طرانس	مطلب	فاهر	
افتر اعل	الوكهم	نر	نغن	الشهر	
وولم	متناجرة	القصيدير	يشاولون	بيجع	
غاب	مؤليماً	يقتلون	اشفوت	رهي	
لصبة	مؤسستانية	يؤاني	أم	كاز	
مقرزة	خذور	ادام	الزرس	طقيف	
متهاكل	قنار	استز هاء	جسوف	خشم	
الغسمة	تكالفي	استرجاع	اختيار	امنية	
علام	مؤثدياً	منعته	الثرام	ألوب	
مؤلي	الكر	الأنسجة	مغابير	لك	
استطاب	لنفذ	الجمالية	ايمنأ	مليء	
مسنج	أجرة	تعبيرهم	تاغل	تألف	
منابق	مترصد	صنوق	زمنج	صوار	

Appendix 3. Arabic-Lex A – teacher's version highlighting non-words

Arabic-Lex Vocabulary Size Test A (50K)

Please look at these words. Some of these words are real Arabic words and some are invented but are made to look like real words. Please tick the words that you know or can use. Here is an example.

book ✓

Thank you for your help.
من فضلك انظر الى هذه الكلمات. بعض هذه الكلمات هي كلمات عربية حقيقية والبعض الآخر كلمات تم تركيبها لتبدو وكأنها كلمات حقيقية. فضلاً اختر الكلمات التي تعرفها أو تستطيع استخدامها. استخدم هذا المثال.

✓ كتاب

شكراً لمساعدتكم،

استغفرت	نزلت	خاتم	أكاد	نشر
نُسُوع	انتقمض	أخيار لهم	زئبق	غلف
ناقد	استنهاض	لُصُوصاً	طازق	ذيق
الاسفنى	العلمس	يزغى	مخفل	شبهة
اغكاتبه	سالم	أغفى	مخفل	سائق
بمزار	القمند	مناقبة	فتراف	شُجُوح
الإقتداء	يُشَح	منطوب	منطى	جُزُو
النشر	انسحاق	منزلهم	غمضن	عمال
أوضاعها	الامتلاء	استدعي	أغفر	الأريماء
عُيَاد	آخر	خافق	أغفاق	خُفر
لُحَقَّة	تهشم	أخوضن	ضام	مخاور
إغتكاب	ذُروغ	سكبه	اختيماك	فُفر
افاقها	روشت	القرى	كثيرة	عطف
نقزاح	قُبعض	مُوحى	الزجركم	مُوقى
أواجه	لايكنى	احترامهم	ذرف	تجول
نعام	مُؤسستانية	الفتيل	بشطاءون	تقل
الأمنيات	نابما	مهالك	مُستخبرات	تألق
أفرو لام	إتحال	زواج	كتاب	اختراد
نلقث	أجتمعت	ينشد	مائدة	الغدير
مقلب	القضاء	صدقة	يخافظ	صنم
خلف	أرتبج	الإنتقام	التخال	كلا
نغزة	مسائي	بالكملة	أفلام	فر
استطاب	اشفى	تغيزهم	الأسرة	استخذاث
تغريز	النزاج	صاجر	مُمتاهاذات	إجماع

Appendix 4. Arabic-Lex B – teacher’s version highlighting non-words

Arabic-Lex Vocabulary Size Test B (50K)					
Please look at these words. Some of these words are real Arabic words and some are invented but are made to look like real words. Please tick the words that you know or can use. Here is an example.					
Thank you for your help.					
من فضلك انظر الى هذه الكلمات. بعض هذه الكلمات هي كلمات عربية حقيقية والبعض الآخر كلمات تم تركيبها لتبدو وكأنها كلمات حقيقية. فضلاً اختر الكلمات التي تعرفها أو تستطيع استخدامها. استخدم هذا المثال.					
book ✓					
✓ كتاب					
شكراً لمساعدتكم،،					
شع	تشكلف	عغار	الأفات	الأوف	
تتجاوز ما	مخطط	اخواتي	احتضان	رائع	
هوى	مخفي	صنديل	غابر	أراد	
تعرض	الأائية	أيتل	غانق	ييجل	
مجرة	تداغي	قار	طلم	صمود	
فلوح	عقاب	إسرداد	قفر	لخم	
منابه	ثقل	الأفس	قمنية	علماني	
يزرد	مصانف	العالي	يخضع	أيد	
إقتر اضبي	الإستبزازات	الإحتياجات	أشد	مبنى	
فقهاء	خطا	ثشف	عند	وشوء	
النوايج	غيب	تستهدف	مشغوط	صفر	
شخور	صاحح	طرانس	مطلب	فاهر	
أقتراش	ألفهم	نر	شمن	اشهر	
زولام	مقاجزة	القصدور	يتناولون	بذيع	
غاب	مؤلماً	يقفون	اشقرت	رسمي	
لصبة	مؤسستية	يقوانى	أم	كثّر	
مفرزة	خذور	أدام	الزرس	طفيف	
متهاكل	قيلان	استنر هاء	جسوف	خشم	
المشيمة	تكاليف	استرجاع	إختبار	امينة	
علام	مزديبا	نضغه	القام	أنوب	
مؤلى	الكر	الأنسجة	مغابير	لكل	
استطاب	لنفاذ	أجمالية	ايضا	مليء	
مسنج	أجرة	تعبير هم	تأمل	تألف	
مناطق	مترصد	صندوق	زمنج	صوار	

List of tables

Table 1. Study participants

Group	Level	Year	<i>N</i>	Mean age
1	Intermediate school	7	43	13
2	Intermediate school	8	42	14
3	Intermediate school	9	37	15
4	High school	10	37	16
5	High school	11	47	17
6	High school	12	48	18
7	Undergraduate	UG	53	20
8	Postgraduate	PG	32	31

Table 2. Descriptive statistics for schoolchildren and undergraduate students

Year	Mean	<i>N</i>	<i>SD</i>	Min	Max	<i>SE</i>
7	9337.21	43	2229.99	5000	12500	340.07
8	11226.19	42	1715.07	8500	15000	264.64
9	13891.89	37	3186.63	7000	19000	523.88
10	16135.14	37	2807.84	11000	23000	461.60
11	17425.53	47	6047.79	10000	29500	882.16
12	21531.25	48	5987.93	14000	31500	864.28
UG	26490.57	53	6334.04	16500	39500	870.04

Table 3. Paired sample statistics of the Arabic-Lex test scores

		Mean	<i>N</i>	<i>SD</i>	<i>SE</i>
Pair 1	Form A	46187.50	32	3049.99	539.17
	Form B	45906.25	32	2985.01	527.68
Pair 2	A Odds	23218.75	32	1453.24	256.89
	A Evens	23390.63	32	1574.67	278.36
Pair 3	A Top-half	23546.88	32	1509.96	266.92
	A Bottom-half	23328.13	32	1365.59	241.39

Table 4. Paired sample test (t-scores) comparing the mean scores difference of parallel forms, odds-evens, and top-bottom halves

		<i>Cronbach's Alpha</i>	<i>t</i>	<i>df</i>	<i>p</i>
Pair 1	Test A-B	.95	1.235	31	0.226
Pair 2	Odds-Evens	.94	-1.407	31	0.169
Pair 3	Top-Bottom	.89	1.352	31	0.186

Table 5. Mean ranks for hierarchy bands

Frequency bands	Mean Ranks
Band 1	4.90
Band 2	3.68
Band 3	3.11
Band 4	2.06
Band 5	1.25

Table 6. Factor loadings

Variable	Level 3	Level 4	Level 2	Level 5	Level 1
Factor loading	.947	.926	.9	.885	.761

Note: Level = frequency level.

List of figures

Figure 1

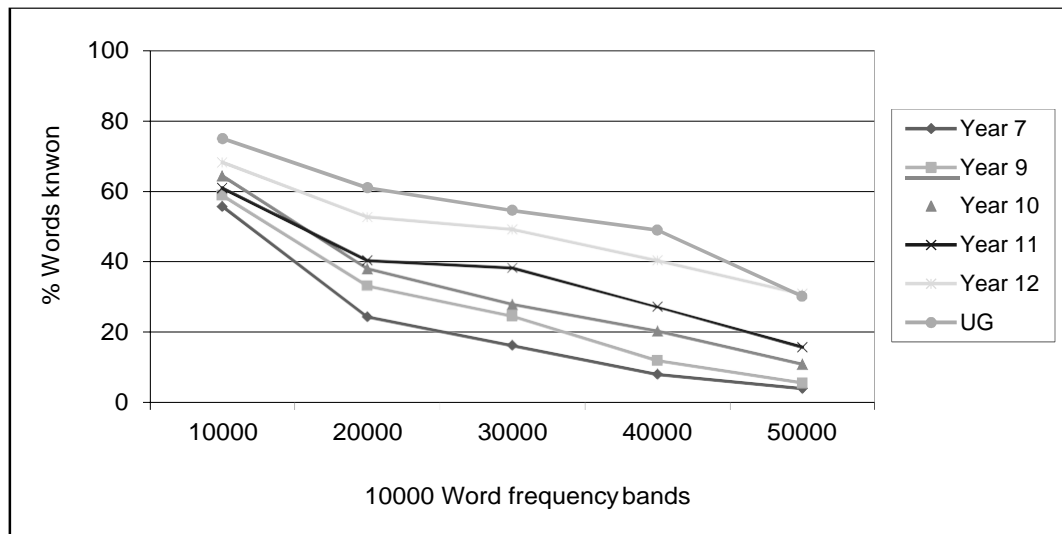
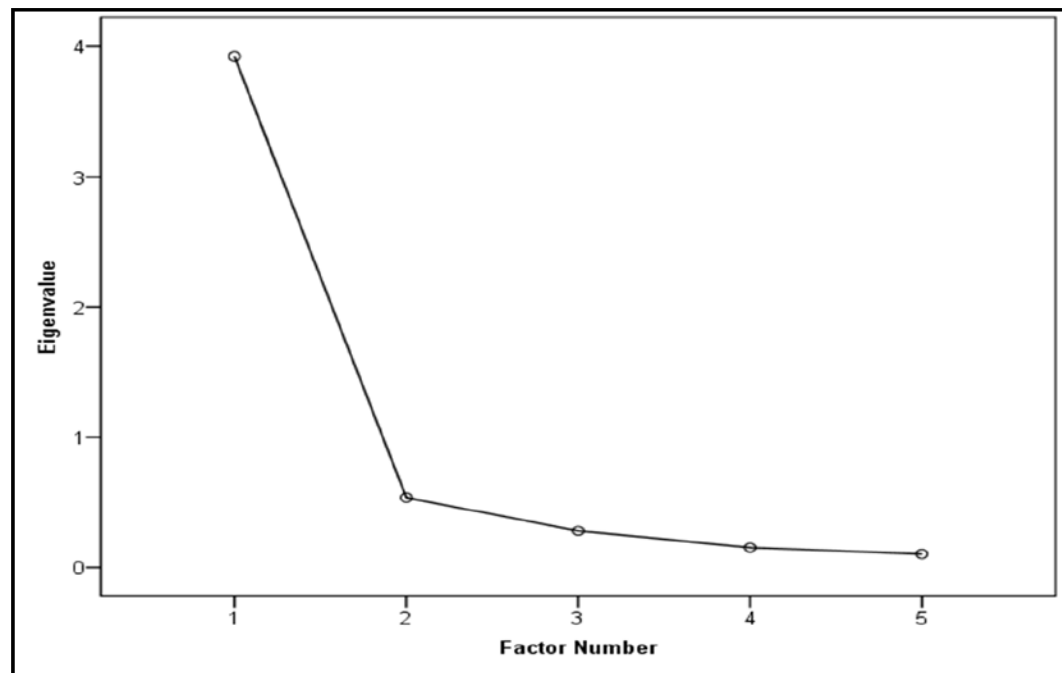


Figure 2



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figures captions

Figure 1. Frequency profiles for school-aged learners and undergraduate students

Figure 2. The scree plot