Passive vocabulary size and speed of meaning recognition

Are they related?*

Batia Laufer and Paul Nation
University of Haifa / Victoria University of Wellington

This paper examines the relationship between fluency and vocabulary size, and also between fluency and word frequency level. Fluency was operationalised as the time learners need to recognize meanings of words sampled from different frequency levels. It was measured by a computerised vocabulary recognition speed test (VORST). The test was given to 488 native and non-native speakers who were divided by vocabulary size into four groups. The four groups were compared on speed of response to the 3000 level and University Word List (UWL) words. Speed was also correlated with vocabulary size. Additionally, response times to different frequency levels were compared for each subject. Results suggest that speed of retrieval is moderately related to vocabulary size and word frequency. Non-native speakers' increase in speed lags behind increase in vocabulary size. Non-native speakers also respond more slowly to less frequent words. Responses of native speakers, on the other hand, are more homogeneous across subjects and across vocabulary frequencies. Speed of retrieval cannot be fully predicted from vocabulary knowledge and therefore speed tests should supplement tests of vocabulary size and depth.

Background

The existing tests of vocabulary size and depth test only word comprehension and production. They do not take into account the length of time needed to recognize or recall the meaning or form of the test items. And yet, for vocabulary to play a significant role in everyday language use, word knowledge is not enough. This knowledge must be fluently available, if fluent conversation or

EUROSLA Yearbook 1 (2001), 7–28. ISSN 1568-1491

reading is expected. Because of the importance of fluency in language learning, it is thus important to measure the speed with which people retrieve the forms and meanings of words, both for diagnostic and for research purposes.

Dimensions of vocabulary knowledge

Discussions of what is involved in knowing a word (Richards 1976, Ringbom 1987, Nation 1990 and 1999) have typically seen it as consisting of a range of aspects of knowledge including knowledge of spoken and written form, morphological knowledge, knowledge of word meanings, collocational and grammatical knowledge, connotative and associational knowledge, and knowledge of constraints to be observed in the use of a word. Thus, knowing a word has been seen to involve a variety of different kinds of inter-related "sub-knowledges". Each of these "sub-knowledges" can be mastered to various degrees. Thus, each type of knowledge is usually seen as involving a cline, with the major dividing point in each cline often seen as being the receptive/productive distinction.

In other models, lexical knowledge is construed either as a continuum consisting of several levels, starting with superficial familiarity with the word, and ending with the ability to use the word correctly in free production (Faerch, Haastrup and Phillipson 1984, Palmberg 1987), or continua (Henriksen 1999) consisting for example of partial-precise knowledge, depth of knowledge, and the receptive-productive continuum. According to Henriksen, however, word use is not part of knowledge, but control over word access. Read (1997, 2000) criticizes the approach in the above models, which equate vocabulary knowledge with the learner's knowledge of particular words. He argues that vocabulary knowledge should be reconceptualized to fit the current view of language proficiency (Bachman 1990). Since proficiency is conceived in terms of learners' communicative skills, the lexical model should incorporate lexical communicative competence in addition to the knowledge of discrete lexical items.

Whether researchers have described vocabulary knowledge as a continuum, continua, or a taxonomy of "sub-knowledges;" whether they have viewed vocabulary use as a component of knowledge, or as control, they all distinguish between passive/receptive vocabulary and active/productive vocabulary. Moreover, since the construction of measurement instruments is based on the construct of what is being measured, i.e. the view of vocabulary knowledge, it is not surprising that most available tests are either global tests of passive knowledge, (Meara and Buxton 1987, Nation 1983) and active knowledge

(Arnaud 1992, Laufer and Nation 1995, 1999), or tests focusing on a specific aspect or aspects of word knowledge (Read 1993, Wesche and Paribakht 1996, Viberg 1993, Schmitt 1999).

A third dimension (in addition to passive and active) that has been given little attention with regard to vocabulary knowledge is the fluency dimension — how quickly a learner can access the various kinds of knowledge. Mezynski (1983), reviewing the relationships between vocabulary and reading comprehension for native speakers, suggests that automaticity of access can be a major factor influencing the effect of vocabulary learning on reading comprehension. If too much time is spent retrieving word meanings then comprehension will suffer. Mezynski also notes that fluency of access and breadth of knowledge are likely to be related (1983:274). Indeed, it is possible that there are two kinds of fluency that may complement each other. These two kinds can be distinguished in various theories of fluency development (Schmidt 1992). One kind of fluency is based on strength or power. It is achieved by repeating the same action over and over again until a very strong link is created between the items involved, for example, a form and its meaning. Another kind of fluency is based on richness of association. That is, a wide variety of links are made between the items involved, thus allowing access from any of a variety of starting points. These two kinds of fluency could be termed "the well beaten path" and "the rich and varied map" and they can be integrated with each other. Indeed, McLaughlin's (1990) restructuring hypothesis sees increases in fluency strength (the well beaten path) triggering the reorganisation of knowledge and the creation of new links. As we shall see later in this study, it is not easy to determine if fluency necessarily involves quantity of knowledge (knowing a lot of words) or if it basically involves knowing words well.

Fluency of access affects language use, and as Mezynski (1983) shows, the type of instruction can affect fluency of access. Nation (1999) suggests that a well-balanced language course should involve four strands: learning through meaning focused input (listening and reading), learning through meaning focused output (speaking and writing), deliberate language focused study, and fluency development across the four skills of listening, speaking, reading and writing. That is, about one-quarter of the time spent in language learning should involve fluency development activities where learners are working at a high level of performance with the language that they already know. Most language courses, especially where English is a foreign language, neglect the fluency development strand. Being able to show that learners have knowledge but lack fluency may be one way of convincing teachers and course designers

that fluency development activities, like easy sustained silent reading, repeated oral retellings of familiar events as in the 4/3/2 activity (Arevart and Nation, 1991), and continuous writing, should be part of a language course. To show this, we would need to develop a measure of vocabulary fluency. In addition to its pedagogical importance, a vocabulary fluency measure is of interest to researchers, since by measuring fluency, we may come closer to understanding how proficiency develops and how proficiency is lost. Does fluency necessarily lag behind knowledge? Is fluency dependent on an integrated system of knowledge or is it item related? Is loss of fluency an early sign of language attrition?

It is important to be clear about what is meant by fluency of access here. The traditional measure of fluency or latency with native speakers of a language is their speed in deciding if a given word form is a real word or not, or if it belongs to a certain category. For an outline of such tasks see Goldinger (1996) and research by de Groot (1993), Kempe and MacWhinney (1996), and Segalowitz, Watson and Segalowitz (1995). Meara (1996) suggested a measure of fluency for foreign learners which required them to recognise an English word in a string of letters. Our interest here, however, is different. We are interested in a vocabulary measure that more closely approximates language use by involving both word form recognition and the association of meaning to that form. This study, therefore, examines the relationship between vocabulary size and fluency by means of a new measure of vocabulary fluency, the VORST.

Description of the vocabulary fluency measure

The vocabulary recognition speed test (VORST) is a computerised version of the Vocabulary Levels Test (Nation 1983, 1990) together with a time-keeping device. The original paper version of the Vocabulary Levels Test measures passive vocabulary knowledge. It consists of five parts. Each part tests samples of words from one of five word frequency levels: the 2000, 3000, 5000 word levels, the University Word List (UWL), which contains 836 words, and the 10000 word level. Except for the UWL level, each level consists of 1000 words. So the 5000 word level for example represents the 4001st most frequent word to the 5000th most frequent word. The levels are based on the Thorndike and Lorge (1944) list, cross-checked against the Francis and Kucera (1982) list. More information on the design of the test can be found in Nation (1983, 1990). Each frequency level of the test comprises six blocks and each block includes six words and three definitions. Words at each frequency level of the test are representative of all the words at that level. The learners are instructed to

match target words with their corresponding definitions by writing the appropriate number next to each definition as in the following completed example:

1	business		
2	clock	6	part of a house
3	horse	3	animal with four legs
4	pencil	4	something used for writing
5	shoe		
6	wall		

The definitions are written using more frequent words than the tested words (see Nation 1983, 1990). The test consists of 90 items (18 in each of five levels), and the target words are tested out of context because context might provide clues to help choose the meaning of the words. The paper version of the test is available in four parallel versions.

The computerized version was based on the original Vocabulary Levels Test for which validation was carried out by Read (1988). However, research is continuing on this test (Schmitt, in preparation), especially on the 2000 and UWL levels which are important for learners with academic goals (Beglar and Hunt, 1999). The computerised version of the test is similar to the paper version in the number of levels, tested items and the basic testing procedure which consists of matching the meaning of three items (synonyms or definitions) with choices from six other items (the tested words). However, instead of displaying all three items at once, the computer displays only one definition on the left side of the screen and the six word block on the right, as in the following example:

	1. apply
	2. elect
make _____	3. jump
	4. manufacture
	5. melt
	6. threaten

When the learner has typed the answer in the form of a number, from 1 to 6, next to the word 'make', the computer displays the next screen with the same block of six words on the right, and a different word or phrase to match:

choose by voting _____

1. apply
2. elect
3. jump
4. manufacture
5. melt
6. threaten

The third screen repeats the same six words with a third word or phrase on the left side of the screen:

become like water _____

1. apply
2. elect
3. jump
4. manufacture
5. melt
6. threaten

Once the three definitions have been matched, the screen displays the following question:

Do you want to change your answers? Y/N

If 'No' is selected, the next group of three words is tested against the next block of six words. If 'Yes' is selected, the testing procedure of the three words that have been tested (as in the example above) is repeated. The position of the correct answer in the block of six words is randomly varied in order to average the search time across answer options. VORST differs from the Vocabulary Levels Test in that only one meaning choice is displayed at a time. In spite of this difference, the three words tested are still partly dependent on each other as in the Vocabulary Levels Test because the same six words are presented as choices for each of the items in the block. However, this interdependence does not invalidate the test. As will be explained later, the various statistics in the study did not rely on response times to individual items, but on the average response times in blocks of the three interdependent items and the average response time in the entire word frequency level.

The program checks the test answers against the key and scores each answer accordingly as correct or incorrect. For each level, it sums the total number of correct answers. In addition, a time measure records the response time to each target word in seconds. The response time is the number of seconds between the appearance of the tested item on the left of the screen and typing an answer in the form of a number (1–6) next to the item. If a subject decided to change

his answers in a block of 3 items, the altered answers were timed anew and the new response times were added to the response times of the first trial. Changes in answers were very infrequent. The computerised version of the test is partly adaptive. If the first three blocks (nine responses) at a frequency level are correct, the program does not continue testing words at the same level, but moves to the next level of word frequency to save time. If, however, one response is wrong, the entire number of items (eighteen) at a level is tested.

During the test, a log file is created for each test-taker in the form of a table which displays accuracy and speed scores. This table cannot be automatically viewed by the test-taker since it has to be accessed by a particular command. The following is an example of the results for the University Word List part of the test. The setting out of the table in Figure 1 is how it appears in the program.

Luwl 14R 4W 186.6(sec) 77.8%

1	55.3	9.34+	30.65	7.80+	17.14+	8.57
2	46.3	22.74	6.09+	14.67+	20.76+	10.38
3	14.1	4.07+	2.86+	4.45+	11.38+	3.79
4	25.4	4.55+	8.68	8.19+	12.74+	6.37
5	23.3	9.72+	7.53+	3.90+	21.15+	7.05
6	22.3	8.40+	3.57+	7.86	11.97+	5.99

Total time 95.14 Total average 7.21 (correct responses)

Figure 1. Log file of results for UWL

The information in the top line of Figure 1 should be read as follows: Luwl refers to the university word list level. 14R means that 14 items were answered correctly (R stands for 'right'); 4W means that four items were answered wrongly. 186.6 (sec.) represents the total time the learner spent on this level (UWL) of the test, and 77.8% is the percentage of correctly answered items on this level of the test. The next six rows represent the six blocks of words tested at each frequency level. In each block, three target words were tested and the records of the answers appear in columns 3, 4, 5. The figure represents the number of seconds it took to answer each item. So, for block 1 the first answer took 9.34 seconds and was correct. (A + next to the time figure means the item was answered correctly; lack of + means that the answer to the tested item was incorrect.) The second item in the first block took 30.65 seconds (a very long time) and was incorrect. Column 6 is the total time spent on correctly answered items in each block (17.14 seconds), column 7 is the average time spent on correctly answered items in a block (8.57 for block 1 (9.34 plus 7.80 divided by 2)). Since

the three items in a block are tested against the same six words, the second item may take less time than the first, and the third less than the first two because of the elimination that occurs with the first and the second answers. Because of this interdependence, an average response time to correct answers is calculated and presented in column 7. Column 2 shows the time spent on each block (55.3 seconds for block 1). This includes the time spent on correct answers, on incorrect answers and on the Y/N decision to the prompt asking the learner whether s/he wants to change his/her answers. The bottom line displays the total time for correctly answered items at the UWL level (95.14 in Figure 1) and the average response time to correctly answered items at the UWL level.

To sum up, the following information is provided in the log file:

1. Information about correctness: each response marked as correct or incorrect; the total number of correct and incorrect responses per level.
2. Information about speed of response: response time for each item, response time for correct items per block, average response time for correct items per block, total time per block, total time per frequency level, total time for correct answers per level, and finally the average response time for correct items at each frequency level of the test.

Having designed the VORST test, we set out to investigate (1) the relationship between the speed of access to word meaning and passive vocabulary size, and (2) the relationship between the speed of access and word frequency. Since speed is considered an indication of language fluency, we wanted to find out whether speakers with larger vocabularies were also more fluent than speakers with smaller vocabularies. Furthermore, we wanted to find out whether the same people would be just as fluent with words of different frequencies. In the present study, we measured the time it takes subjects to recognise word form, to recognise word meaning while selecting from choices, and to press the appropriate answer key. Although this is a complex activity, it is the same for all words and for all subjects involved in the experiment, both native speaker and non-native speaker.

The study

The research posed the following four questions:

1. Is there a difference in response times between groups with different vocabulary sizes?

2. What is the correlation between response times and vocabulary size?
3. What is the variance in response times of people with the same vocabulary correctness scores?
4. Do the same people have different response times to words at different frequency levels?

Questions 1, 2, and 3 relate to the issue of fluency as a function of vocabulary size, while question 4 relates to the issue of fluency as a function of word frequency.

Subjects

488 subjects were tested by VORST. Thirteen were educated native speakers, 475 were university learners of English as a foreign language of which 12 were English majors. However, 34 lower-level learners were eliminated from the sample as they were suspected of being random 'key hitters'. These 34 learners had very low correctness scores and yet the speed of their answers was faster than that of native speakers. The conclusion was that they were hitting the answer keys without any serious attempt to do the test honestly. Thus, after the 34 subjects were eliminated, our final sample consisted of 454 subjects.

The learners were native speakers of Hebrew, Russian and Arabic. All the test-takers were familiar with the use of computers, because CALL exercises constitute a part of EFL courses, and all university students are taught basic computer skills. The native speakers, who were mostly members of staff and graduate students, used computers on a daily basis. The EFL learners and the English majors were mostly graduates of Israeli high schools and some were immigrants from the former Soviet Union. But their knowledge of English was comparable to the non-immigrants as they all passed the same standardised University entrance exam which includes an English reading comprehension sub-test. The English majors also passed a special departmental written entrance examination. Based on our teaching experience and knowledge of students, we expected a wide range of vocabulary size among the EFL students and a much higher vocabulary size among the English majors. The native speakers, who were English majors and some staff members were expected to have the highest vocabulary size. We were aiming at the widest possible range of vocabulary size scores since our objective was to examine the relationship between different vocabulary sizes and fluency.

Procedure

The EFL learners were given the test in a CALL lab during their normal English as a foreign language classes. The test was administered by an assistant in the presence of the class teachers. English majors and native speakers were invited on a personal basis to the office of one of the researchers and used her computer for the test. These are slightly different conditions for the two groups, but the comparison is not the main purpose of this study. Before the beginning of the test, an example — an item on the left to be matched with one of the six alternatives on the right — was written out on the blackboard and the testing procedure was explained. However, during the test it became evident that in spite of the explanation, the test takers needed some time to get used to the test. Therefore the first items took much longer to answer than the following ones. None of the subjects in the experiment were told that the responses would be timed. They were told, however, not to get distracted during the test, not to interrupt the test in the middle and to finish it as fast as possible. They were told they could leave after the completion of the test. Although we cannot be fully certain they were working at their maximal speed, we have no reason to believe they did not want to finish the test as fast as possible. More importantly, we were not interested in the absolute speed but in the relationship between speed of response and vocabulary size. If learners did not work at maximum speed, this must have been equally true for learners with different vocabulary sizes. The results of the tests were collected in the log files. A SAS program was written in such a way that it could read and process the data from the original log files as in Figure 1.

Results

Kuder-Richardson 21 reliability values for the correctness scores at the five levels and the whole test are as follows: 2000 level .82, 3000 level .80, UWL level .73, 5000 level .79, 10000 level .74, and the entire test .93. The entire test shows high reliability and the levels show moderate reliability.

Vocabulary size and speed of access-1

Research question 1 asked whether there was a difference in response times between groups of different vocabulary sizes. Therefore, the sample including

learners and native speakers was split according to the correctness score (vocabulary size) into four groups. Learners who scored lower than 15 out of 90 were not included in this comparison. A score of less than 15 meant that the learner did not even know the first 2000 words. A total score of 15 represented knowledge of about 1,500 word families. It seemed misleading to test learners with such low knowledge on response time to 2000 level items, let alone less frequent words. Since the maximum score on the test is 90, the division was as follows: group 1 was the group of subjects whose score was in the range of 15–32; group 2 had scores from 33 to 50; group 3, from 51 to 68; group 4 , from 69 to 90. There may be a disagreement among researchers as to how accurately the actual vocabulary size score can be extrapolated from test scores. Yet a higher score on the test reflects a larger vocabulary. The implicational scaling found by Read (1988) indicates that if two students received an identical score, they could not have answered items from different frequency levels each. A possible exception could be learners of a Romance mother tongue (Meara, personal communication). They may know more infrequent cognates than common English words. However, we did not have such subjects in our study. We will therefore refer to the four divisions according to test scores as four vocabulary size groups. Group 1 had the lowest vocabulary size, group 4 , the highest.

The four vocabulary size groups were compared on the speed of response to words at the 3000 frequency level and at the UWL level by means of a one-way ANOVA. (We do not present the comparison at the 2000 level since response times included an additional factor: familiarization time). In the two cases, the differences between the different vocabulary size groups were found to be significant (At the 3000 level: $F(414,3) = 36.97$, $p < .001$; at the UWL level: $F(414,3) = 13.52$, $p < .001$). Table 1 presents the results of Duncan's Multiple Range test showing the differences in response time between the four groups. Groups marked by different letters are significantly different from one another. Table 2 presents the same results for the University Word List.

Table 1 shows that the first two groups are not significantly different from

Table 1. Response time to the words at the 3000 level

Duncan Grouping	Mean	n	Group	Vocabulary size score
A	18.44	188	1	15–32
A	16.63	86	2	33–50
B	14.45	89	3	51–68
C	8.19	55	4	69–90

Table 2. Response time to the words at the University Word List level

Duncan grouping	Mean	n	Group	Vocabulary size score
A	23.09	188	1	15–32
A	22.995	86	2	33–50
A	20.50	89	3	51–68
B	12.97	55	4	69–90

each other in the speed of response to 3000 level words. Groups 3 and 4 are significantly different from (faster than) groups 1 and 2. They are also significantly different from each other. With regard to the UWL (Table 2), group 4, the best in terms of vocabulary size, is significantly different from the other three. Group 3 is not significantly different in the speed of response from group 2 (even though the two were significantly different in speed of response to the 3000 level words). Tables 1 and 2 show that total vocabulary size has to be quite a distance beyond the tested level in order for there to be a significant difference in fluency. Fluency on the 3000 level words increased when the average vocabulary size reached an average size score of 60 (See the score ranges in Table 1). Speed of response did not change within the 15–50 size score range. Fluency on the University Word List, which includes vocabulary from the 4000 and 5000 frequency levels, increased when the average size score was 80. That is, fluency seems to lag significantly behind size of vocabulary.

Vocabulary size and speed of access-2

Research question 2 was about the correlation between response times and the vocabulary size of the test-taker. Tables 3 and 4 present the correlation values between response time at different frequency levels and the learners' vocabulary size, i.e. the total correctness score. In Table 3, the entire sample is included. Table 4 presents the correlations for non-native speakers only.

Tables 3 and 4 show moderate and significant correlations between vocabulary size and speed of response at the different word frequency levels. Negative correlations mean that the larger the size, the shorter the response time. In other words, people who know more words in general are likely to come up with the meanings of words more quickly than people with smaller vocabularies.

Table 3. Correlations between the vocabulary size score and speed of response at different frequency levels for the whole sample

	2000	3000	5000	UWL	10000
r	-0.39	-0.52	-0.50	-0.37	-0.69
p	.0001	.0001	.0001	.0001	.0001
n	454	344	250	268	52

Table 4. Correlations between the vocabulary size score and speed of response for non-native speakers only

	2000	3000	5000	UWL	10000
r	-0.38	-0.40	-0.50	-0.31	-0.67
p	.0001	.0001	.0001	.0001	.0001
n	441	331	237	255	39

Vocabulary knowledge and speed of access

The third research question was "What is the variance in response times of people with the same vocabulary correctness scores?". This question was examined for the separate samples of native and non-native speakers. At each frequency level, the response times of people who received the same vocabulary size score at that level were checked. In the case of native speakers, all the subjects answered all the items correctly (i.e. scored 18 points per level). In the case of the non-native speakers, we looked for groups which satisfied two requirements: (a) they had a large number of correct answers (14 at least) at each frequency level (i.e. they could be considered as knowledgeable of the respective vocabulary levels), and (b) they consisted of a reasonable number of subjects. The outcome was a group of 54 subjects with 18 correct answers at the 3000 level, 29 subjects with 14 correct answers at the 5000 level, 16 subjects with 14 correct answers at the UWL level. We could not find a group of learners that would satisfy the two requirements at the 10000 level. The 2000 level was not examined because of the familiarization factor mentioned earlier. It introduced an additional variance into the response time, which was probably unrelated to the speed of access, but was dependent on individual speed of familiarization with the testing format. In spite of the removal of the 2000 level, we felt that the examination of variances at the other frequency levels provided useful information for research question 3. Table 5 shows the time results of native speakers, Table 6, the results of non-native speakers.

Table 5. Response time of subjects with the same correctness scores: Native speakers

		Correct Response time			
Level	n	answers	Mean	SD	Variance
3000	13	18	4.57	1.37	1.88
5000	13	18	4.78	1.56	2.43
10000	13	18	5.44	1.5	2.77
UWL	13	18	4.79	1.38	1.92

Table 6. Response time of subjects with the same correctness scores: Non-native speakers

		Correct Response time			
Level	n	answers	Mean	SD	Variance
3000	54	18	12.2	6.4	40.8
5000	29	14	17.75	5.94	35.32
UWL	16	14	17.96	7.3	53.4

The variance values in Table 5 point to a homogeneous group in terms of speed of response, irrespective of the frequency level. The standard deviation is smaller than a third of the mean. The variance values in Table 6, on the other hand, show that subjects whose correctness scores are identical, i.e. have the same vocabulary size at a certain vocabulary level, differ widely in the time it takes them to retrieve the meaning of this vocabulary. For these learners, knowledge and speed in accessing the same items are not closely related.

Vocabulary frequency and response times

Research question 4 concerned the effect that word frequency had on the speed of access to the meaning of the word. It asked whether the same people had different response times to words at different frequency levels. This question was examined for the separate samples of native and non-native speakers. The 2000 frequency level was removed from the comparison because, as mentioned earlier, some of the response time had to do with familiarization with the test procedure. In spite of this limitation, we felt that comparing speed of response at the remaining levels of frequency would be revealing in terms of research question 4. In each group of subjects, the response times for different word

frequency levels were compared by the Repeated Measures procedure. In the group of native speakers, all subjects did the entire test. Therefore all of them participated in the comparison of the four levels of frequency. No significant difference among the four means was found. Comparison of responses for individual pairs of frequency levels revealed a significant difference only between the 3000 and the 10000 frequency words, at the .05 level of significance.

Tables 7 and 8 present differences in speed of response between pairs of vocabulary frequency levels. The significance of the difference is indicated by * ($p < .05$), ** ($p < .01$), *** ($p < .001$). The sign > means that it took longer to respond to the frequency level on the left.

Table 7. Differences in mean response times between pairs of frequency levels: Native speakers

Frequency levels	Difference
5000 > 3000	0.21
UWL > 3000	0.22
UWL > 5000	0.01
10000 > 3000	0.83*
10000 > 5000	0.62
10000 > UWL	0.61

In the group of non-native speakers, not all the subjects were able to do all the levels of the test. Furthermore, only subjects who had at least 6 correct responses at a vocabulary level were entered into the comparison. A Repeated Measures comparison of four levels of word frequency (3000, 5000, UWL, 10000) processed only 35 subjects (the best non-native learners). This comparison showed that there was a significant difference between the four response speed means ($F(32,3) = 23.09$, $p < .001$). Post hoc multiple range tests revealed that all the four means were significantly different from each other except for the difference between response times to the UWL and 5000 word frequency levels. In order to compare more learners on their response speed at different frequency levels, a second Repeated Measures analysis was performed comparing mean response speed at 3000, 5000, and UWL levels without the 10000 level. In this analysis, 178 subjects were processed. In this analysis too, there was a significant difference between the three response speed means ($F(176,2) = 61.69$, $p < .001$). Post hoc multiple range tests revealed that all the three means were significantly different from each other. And finally, 225 subjects who could do the 3000 and the UWL sections of the test were compared on their speed of

Table 8. Differences in mean response times between pairs of frequency levels: Non-native speakers

Frequency levels	Difference		
	A	B	C
	n=35	n=178	n=225
UWL>3000	5.15***	6.28***	6.6***
5000>3000	3.57**	3.84***	
UWL>5000	1.58	2.43***	
10000>3000	9.84***		
10000>5000	6.27***		
10000>UWL	4.68**		

A: Learners who did the entire test.
B: Learners who did 3000, 5000 and UWL
C: Learners who did 3000 and UWL

responses to the two levels by a paired t-test. The difference was found to be significant ($t=12.04$, $p<.001$). Table 8 shows the mean differences in response times between each pair of word frequencies. The figures clearly show that in the non-native speaker population, the less frequent words take longer to match with their meanings.

In sum, the results comparing the subjects themselves on the response speed to words at different frequency levels show that native speakers' responses slow down only with the least frequent vocabulary. The response speed of the non-native speakers, however, changes at each level of word frequency, showing a relationship between speed and strength of knowledge.

Discussion

The aim of this paper was to investigate how the speed of meaning recognition is related to vocabulary size and to the frequency of the tested words. To do this, we used VORST — a relative (not an absolute) measure of the speed with which meanings rather than form of words at different frequencies are recognized. VORST, in its present form, is a preliminary version. We hope that it will inspire colleagues interested in vocabulary measurement to improve the test technology, e.g. adding a familiarisation session, a 'don't know' answer option, and a stopping decision device.

Speed of meaning recognition: how does it relate to vocabulary size?

Vocabulary size and speed of response were found to be moderately correlated, except for the less frequent vocabulary. It would be too simplistic, however, to state that people with larger vocabularies will be faster in accessing word meaning. The data for research question 1 shows that the two lower groups in terms of vocabulary size did not differ significantly in response times to the 3000 level words. This means that learners across a wide range of scores (15–50) were similar in the speed of their responses to the most common words. Apparently speed of response times to the 3000 level picks up only when the learners have progressed well beyond this level, i.e. they belong to vocabulary size groups 3 and 4. Comparison of response times to the UWL suggests a similar conclusion. Group 3, with a vocabulary score of 50–68, was significantly different from group 2 (33–50 on the vocabulary size score) on the response rates to the 3000 level words. However, the two groups did not differ significantly in the response time to UWL. This suggests that for the speed to increase on the less frequent words of the UWL, the learner has to acquire a vocabulary beyond this level. The group that is significantly more fluent is the group with the top range of size scores (69–90). The conclusion of the two comparisons is therefore that increase in automaticity lags behind increase in size.

Our study was not designed to find out why this lag occurs. We can only speculate about the reasons. It is possible that the lag between vocabulary size and fluency is symptomatic of vocabulary development as such, i.e. a lot of vocabulary has to be acquired before some of it becomes automatic. Yet it is equally plausible that this lag is symptomatic of the learning conditions of our subjects and represents the kind of practice that language learners get. That is, recently learned vocabulary is not practised to a high level of fluency in carefully designed fluency activities. Learners gain a kind of fluency practice with this vocabulary when they move on to texts containing this vocabulary and other new vocabulary. These are not ideal conditions for vocabulary development. Hulstijn (in press) suggests that fluency can best be developed not by exposure to i+1 input (a little beyond the learner's level), but by practising language at the i-1 input level, i.e. without the presence of any new language features. This has not been empirically demonstrated yet. Therefore, it would be interesting to see if fluency can be more effectively developed through the use of graded texts, such as graded readers. Students would be required to read a number of graded readers at the same vocabulary difficulty level, thus creating conditions for rehearsing the vocabulary while not burdening the learner with new words.

An interesting question related to size and fluency is whether it is possible to reach a high degree of fluency with most of the words within a limited vocabulary? If, for example, learners did substantial fluency practice with graded readers within a 2000 word vocabulary, could they reach a high level of fluency even though they knew little vocabulary beyond the most frequent 2000 words? Or is it absolutely necessary to have a reasonably large well-integrated vocabulary, say beyond 4000–5000 words, before it is possible to become fluent with the most frequent 2000 words? The answers to these questions can have a profound influence on pedagogy and on the setting of learning goals.

Speed of response and the word frequency effect

The data to answer research question 4 clearly showed that learners respond more slowly to less frequent words. Responses to almost all frequency levels were significantly different from one another. Apparently words that appear more often in the input to the learners and are more frequently used by the learners have also become better automatized. By contrast, native speakers' vocabulary size is large and well established enough to make differences in speed with higher frequency words largely insignificant.

Both the effect of word frequency and the "lagging behind" phenomenon discussed in the preceding section, and native speakers' uniformity in response times across frequency levels, point to the same pedagogical implication. Improving the fluency of a certain level of vocabulary requires increased exposure to this vocabulary, practice of its use and acquisition of vocabulary far beyond this level.

Speed of response: native and non-native speakers

Native speakers and foreign learners of a language exhibit different patterns of response times. Not only are native speakers faster on all responses, but they are also much more homogeneous in their speed than foreign language learners. Table 7 shows the homogeneity of native speakers in their response time across words of different frequencies. Table 8, on the other hand, shows that non-native speakers' responses became less automatic with decreasing word frequency. The homogeneity of native speakers' and the variance of non-native speakers' speed of access are also reflected in the variances in Tables 5 and 6.

Unlike native speakers, non-native speakers with an identical correctness score at a certain word frequency level differ widely in response time to the same

words. The reason for this variation may be related to the variation in vocabulary size score. (Even though subjects in Table 6 had the same size score at a particular frequency level, they may have had different total vocabulary size scores). As mentioned earlier, learners with larger vocabularies are faster than learners with smaller vocabularies. Native speakers in the study were graduate students and academics. Their vocabularies were probably homogeneous too and must have far exceeded the 10000 word level which was the limit of the test. Native speakers who are high school graduates are reported to know between 15000 and 20000 word families (Goulden, Nation and Read, 1990). No wonder that the responses to the 10000 word level exhibited well-developed automaticity.

Another possible reason for the variance in the non-native speakers' sample has to do with the different learners' experience with these words. The amount of written input learners receive varies according to their field of study. Some areas of study, for example Psychology and Sociology, require a lot of reading in English. Other subjects, e.g. Hebrew Literature, or Biblical Studies require much less. This difference of experience is partly signalled in the differences in the total vocabulary size described earlier in this study. But even in groups of learners with similar vocabulary size, some learners may have improved in fluency due to repeated exposures to words in their reading material. Hence the lack of homogeneity in response times.

Conclusion

In this paper we introduced the first version of a measure of passive vocabulary fluency, the VORST (Vocabulary Recognition Speed Test). This measure was operationalised as the speed with which a subject matches a target word with its meaning. Using the VORST, we examined the relationship between fluency and vocabulary knowledge reflected in the passive vocabulary size score, and also the relationship between fluency and the frequency of words being accessed. The study was carried out with learners of English as a foreign language and with native speakers.

Native and non-native speakers vary considerably in their response patterns. Not surprisingly, non-native speakers are much slower to recognise vocabulary meaning than native speakers. They also form a heterogeneous group with regard to fluency and this is reflected in the large variances of response times. The relationship between fluency and vocabulary knowledge of learners was found to be non-linear, i.e. an increase in vocabulary size did not

automatically result in an increase in response speed. Speed on a frequency level increased only when learners' size progressed far beyond that level. Moreover, within each learner, fluency is not uniform but changes according to the frequency of words. Less frequent words require more time for recognition. Native speakers, on the other hand, form a homogeneous group with regard to speed of response and exhibit uniformity in response times across the frequency levels measured in this test.

We do not know whether the lag of learners' fluency behind vocabulary knowledge has to do with the nature of vocabulary development, or the conditions in which vocabulary was learnt. A study similar to ours with learners in a different learning setting, for example in the language speaking environment, could provide an answer. Similarly, we do not know whether it is possible to achieve fluency within a limited vocabulary size. To answer this question, we would need to study subjects (immigrants, foreign workers) whose language needs involve fluent communication at a basic vocabulary level. In order to carry out such studies, a measure of vocabulary fluency is needed. In addition to acquisition studies, the measure can be useful in investigating the early stages of attrition where problems of accessibility precede deterioration of knowledge. It can also be helpful in studies of re-learning, for comparing novice learners with "second-timers" and checking whether re-learning is advantageous for fluency.

Concluding remarks on VORST

The pedagogical importance of a test like VORST lies in diagnosing learners whose fluency is further behind others of a similar vocabulary size. This would allow learners to gain appropriate practice to help remedy this lack. In general, VORST and other similar tests seem to have a very useful role to play in sorting out theoretical issues and in providing practical guidance for teachers and learners. As the vocabulary research instruments increase, we can expect to have better data on which to base theory and practice.

Notes

* The authors would like to thank Ms Na'ama Tal, a graduate student at the University of Haifa, for assistance with data collection and the University of Haifa CALL centre for technical assistance.

References

Arevart, S. and Nation, I.S.P. 1991. " Fluency improvement in a second language". *RELC Journal* 22: 84–94.

Arnaud, P. 1992. "Objective lexical and grammatical characteristics of L2 written compositions and the validity of separate component tests". In *Vocabulary and Applied Linguistics*, P. Arnaud and H. Bejoint (eds), 133–145. London: Macmillan.

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

Beglar, D. and Hunt, A. 1999. "Revising and validating the 2000 word level and the university word level vocabulary tests". *Language Testing* 16: 131–162.

de Groot, A.M.B. 1993. "Word-type effects in bilingual processing tasks: support for a mixed-representational system". In *The Bilingual Lexicon*, R. Schreuder and B. Weltens (eds), 27–51. Amsterdam: John Benjamins.

Faerch, K., Haastrup, K. and Phillipson, R. 1984. *Learner Language and Language Learning*. Clevedon: Multilingual Matters.

Francis, W.N. and Kucera, H. 1982. *Frequency Analysis of English Usage*. Boston: Houghton Mifflin.

Goldinger, S.D. 1996. "Auditory lexical decision". *Language and Cognitive Processes* 11: 559–567.

Goulden, R., Nation, P. and Read, J. 1990. "How large can a receptive vocabulary be"? *Applied Linguistics* 11: 341–363.

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

Hulstijn, J.H. In press. "Intentional and incidental second-language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity". In *Cognition and Second Language Instruction*, P. Robinson (ed.). Cambridge: Cambridge University Press.

Kempe, V. and MacWhinney, B. 1996. "The crosslinguistic assessment of foreign language vocabulary learning". *Applied Psycholinguistics* 17: 149–183.

Laufer, B. and Nation, P. 1995. "Vocabulary size and use: Lexical richness in L2 written Production". *Applied Linguistics* 16: 307–322.

Laufer, B. and Nation, P. 1999. "A vocabulary-size test of controlled productive ability". *Language Testing* 16: 33–51.

McLaughlin, B. 1990. "Restructuring". *Applied Linguistics* 11: 113–128.

Meara, P. 1996. "The third dimension of vocabulary knowledge". Paper presented at the 11th AILA World Congress. Jyvaskyla, Finland, August 5–9.

Meara, P. and Buxton, P. 1987. "An alternative to multiple choice vocabulary tests". *Language Testing* 4: 142–151.

Mezynski, K. 1983. "Issues concerning the acquisition of knowledge: effects of vocabulary training on reading comprehension". *Review of Educational Research* 53: 253–279.

Nation, I.S.P. 1983. "Testing and teaching vocabulary". *Guidelines* 5: 12–25.

Nation. I.S.P. 1990. *Teaching and Learning Vocabulary*. Rowley, Mass.: Newbury House.

Nation, I.S.P. 1999. *Learning Vocabulary in Another Language*. E.L.I. Occasional Publication No. 19, Victoria University of Wellington.

Palmberg, R. 1987. "Patterns of vocabulary development in foreign language learners". *Studies in Second Language Acquisition* 9: 202–221.

Read, J. 1988. "Measuring the vocabulary knowledge of second language learners". *RELC Journal* 19: 12–25.

Read, J. 1993. "The development of a new measure of L2 vocabulary knowledge". *Language Testing* 10: 355–371.

Read, J. 1997. "Vocabulary and testing". In *Vocabulary: Description Acquisition and Pedagogy*, N. Schmitt and M. McCarthy (eds), 303–320. Cambridge: Cambridge University Press.

Read, J. 2000. *Assessing Vocabulary.* Cambridge. Cambridge University Press.

Richards, J. C. 1976. "The role of vocabulary teaching". *TESOL Quarterly* 10: 77–89.

Ringbom, H. 1987. *The Role of the First Language in Foreign Language Learning.* Clevedon: Multilingual Matters.

Schmidt, R. 1992. "Psychological mechanisms underlying second language fluency". *Studies in Second Language Acquisition* 14: 357–385.

Schmitt, N. 1999. "The relationship between TOEFL vocabulary items and meaning, association, collocation and word-class knowledge". *Language Testing* 16: 189–216.

Schmitt, N. In preparation. "Developing New Versions of the Vocabulary Levels Test".

Segalowitz, N., Watson, V. and Segalowitz, S. 1995. "Vocabulary skill: single-case assessment of automaticity of word recognition in a timed lexical decision task". *Second Language Research* 11: 121–136.

Thorndike, E. and Lorge, I. 1944. *The Teachers' Word Book of 30,000 Words.* Teachers College Columbia University.

Viberg, A. 1993. "Crosslinguistic perspectives on lexical organization and lexical progression". In *Progression and Regression in Language*, K. Hyltestam and A. Viberg (eds), 340–385. Cambridge: Cambridge University Press.

Wesche, M. and Paribakht, T.S. 1996. "Assessing second language vocabulary knowledge: Depth versus breadth". *The Canadian Modern Language Review* 53: 13–40.