publications with regard to this article are: *Exploring Spoken English* (Cambridge University Press, 1997) and *Cambridge Grammar of English* (Cambridge University Press, 2006) (both with Michael McCarthy) and (with David Nunan) *The Cambridge Guide to Teaching English to Speakers of Other Languages* (Cambridge University Press).

Acknowledgements

To Cambridge University Press for permission to quote extracts from the CANCODE corpus (The Cambridge and Nottingham Corpus of Discourse in English), part of the 1 billion word Cambridge International Corpus (CIC) and to Michael McCarthy, co-director with me of the CANCODE project, for permission to draw on material for this paper which we have jointly collected and written.

A vocabulary size test

Paul Nation
LALS, Victoria University of Wellington
David Beglar
Temple University Japan, Osaka

Sponsored by Compass Publishing

Keywords: vocabulary testing, vocabulary size, assessment, proficiency

The Vocabulary Size Test was developed to provide a reliable, accurate, and comprehensive measure of a learner's vocabulary size from the 1st 1000 to the 14th 1000 word families of English. There are several reasons for wanting to measure a non-native speaker's vocabulary size. One reason is to see how close the learner is to having enough vocabulary to be able to perform certain tasks such as read a novel, read newspapers, watch movies, and listen to friendly conversations. There is now data on the vocabulary sizes needed to perform such receptive tasks (Nation, 2006) and this indicates that learners need to have a vocabulary close to 8,000 word families to do this. This assumes that proper names do not need to be known. Table 1 summarises the data from Nation (2006) showing the vocabulary sizes needed to reach 98% coverage of various kinds of text. When a learner has 98% coverage of a text, adequate unassisted comprehension is possible (Hu & Nation, 2000).

The goal of around 8,000 word families is an important one for learners who wish to deal with a range of unsimplified spoken and written texts. It is thus helpful to know how close learners are to this critical goal.

A second reason for measuring vocabulary size is to be able to chart the growth of learners' vocabularies. There is virtually no information on how quickly non-native speakers' vocabularies grow. Measuring this requires a test capable of measuring a large range of vocabulary sizes both longitudinally and across a group of learners.

A third reason for measuring vocabulary size is to be able to compare non-native speakers with native speakers. Such comparisons could be useful in comparing rates of growth in second language learning settings (Do non-native speakers increase their vocabulary knowledge at a faster, slower, or similar rate to native speakers?) and in determining if there is a threshold, as suggested by the 98%

Table 1. Vocabulary sizes needed to get 98% coverage (including proper nouns) of various kinds of texts

Texts	98% coverage	Proper nouns
Novels	9,000 word families	1-2%
Newspapers	8,000 word families	5-6%
Children's movies	6,000 word families	1.5%
Spoken English	7,000 word families	1.3%

coverage figure, where further increase brings no noticeable effects on comprehension, reading pleasure, or reading speed.

Other existing tests

The Vocabulary Levels Test (Nation, 1983; Schmitt, Schmitt & Clapham, 2001; Beglar & Hunt, 1999) is currently widely used to determine whether learners need to focus on high frequency words, academic words, or low frequency words. It is a diagnostic test that looks at separate slices of a learner's vocabulary (the 2nd 1000, the 3rd 1000, the 5th 1000, the Academic Word List, and the 10th 1000). There is no Academic Word List section in the Vocabulary Size Test and words from the Academic Word List can be found from the 1st 1000 to the 10th 1000 of the Vocabulary Size Test. The Vocabulary Size Test has a different purpose. It is not a diagnostic measure like the Vocabulary Levels Test, but is a proficiency measure used to determine how much vocabulary learners know.

Teachers and researchers have felt the need for a more comprehensive test of vocabulary size, and this has been reflected in attempts to fill the gaps in the Vocabulary Levels Test (at the 4th 1000, and 6th to 9th 1000 levels) through estimates of knowledge in these gaps by extrapolating from the scores on each side of the gap, for example using a learner's 3rd 1000 and 5th 1000 scores to estimate what the 4th 1000 might be.

Meara and Jones (1987, 1990) developed a computerized vocabulary size test (*the Eurocentres Vocabulary Size Test 10ka*) based on the first 10,000 words of Thorndike and Lorge's (1944) frequency count. They used a yes/no item type with one non-word item for every two real words. The non-word items were nonsense words which were used to measure the reliability of the learners' answers. Their test consisted of ten 1000 word levels.

The Vocabulary Size Test differs from the Eurocentres test in several important ways. It uses a different test format, multiple-choice compared with the yes/no format of the Eurocentres test. It puts the tested word in a short non-defining context, and it is based on a different set of word frequency lists from the Thorndike and Lorge lists. The Vocabulary Size Test is also freely available to teachers and researchers.

The source of the words used in the vocabulary size test

The development of the Vocabulary Size Test has been greatly helped by the development of the fourteen 1000 BNC word lists (Nation, 2006). The lemma-based Thorndike and Lorge (1944) and Leech, Rayson, and Wilson (2001) counts are not suitable for measures of receptive vocabulary size because the more inclusive word family is a more appropriate unit for such a receptive measure. The word family is more appropriate because learners beyond a minimal proficiency level have some control of word building devices and are able to see that there is both a formal and a meaning relationship between regularly affixed members of a word family. There is also increasing evidence that the word family is a psychologically real unit (Nagy, Anderson, Schommer, Scott, & Stallman, 1989; Bertram, Baayen, & Schreuder, 2000; Bertram, Laine, & Virkkala, 2000).

The word family unit used in the fourteen 1000 BNC word family lists (available for free download from <www.vuw.ac.nz/lals/staff/paul-nation/nation.aspx> along with the Range program) is set at level 6 of Bauer and Nation's (1993) scale of levels. This is a very inclusive definition of a word family making some families very large. However, level 6 only includes bases which are free forms. For example, *romance* and *romantic* (both in the 5th 1000 level) could not be members of the same family because their base is a bound form, not a free form. In addition, all of the family members at level 6 meet the criteria of regularity, frequency, productivity, and predictability used in developing the word family levels.

The word lists used to choose and sequence the test items differ from those described in Nation (2006). The lists in Nation (2006) were based on the whole 100,000,000 token British National Corpus. It was clear from this study that the largely formal written nature of the British National Corpus strongly affected the high frequency levels, meaning that items like *cat*, *hello*, *sun*, *worse* occurred in the 4th 1000 rather than at a higher frequency level. Similarly there were very formal words like *civil* and *commission* occurring in the 1st 1000 words. As a result the first twelve 1000 word lists were revised using word family range and frequency figures from only the 10 million token spoken section of the British National Corpus. This resulted in a more sensible ordering although the changes were not large. The Vocabulary Size Test is based on this spoken corpus ordering and the lists available from <www.vuw.ac.nz/lals/staff/paul-nation/nation.aspx> are the lists based on the spoken corpus.

Although it may seem a little strange to use a spoken corpus-based ordering for a test of written receptive knowledge, it was felt that the spoken

ordering more closely represented the order in which the intended test-takers might learn the words. Frequency of occurrence however is only one factor, although a very important one, affecting order of acquisition.

The nature of the Vocabulary Size Test

The Vocabulary Size Test samples from the most frequent 14,000 word families of English. The test consists of 140 items (ten from each 1000 word level) (see the Appendix). Here is a sample item from the 5th 1000 word level.

1. miniature: It is a **miniature**.
   - a a very small thing of its kind
   - b an instrument for looking at very small objects
   - c a very small living creature
   - d a small line to join letters in handwriting

The Vocabulary Size Test is a measure of written receptive vocabulary size. In order to answer the items, the test-takers have to have a moderately developed idea of the meaning of the word. This makes it a slightly more difficult test than the Vocabulary Levels Test (Schmitt, Schmitt, & Clapham, 2001), because the correct answer and the distractors usually share elements of meaning.

1. innocuous: This is **innocuous**.
   - a cheap and poor in quality
   - b harmless
   - c not believable
   - d very attractive-looking

The multiple-choice format was chosen:

1 to allow the test to be used with learners from a variety of language backgrounds.

2 to control the level of difficulty of the answers. That is each item attempts to tap roughly the same degree of knowledge of a word.

3 to make marking as efficient and reliable as possible.

4 to make learners demonstrate knowledge of each item.

The test words were all put in a simple non-defining context. Research by Henning (1991) has shown that the use of such a context is a desirable feature. This is probably because it indicates the part of speech of the word, orients the test-taker to view it as an item of language use, and provides a little extra associational help in accessing the meaning.

The items measure receptive knowledge of vocabulary. That is, the learners are provided with the word form and have to access the meaning of the word. They need to have a moderately developed idea of the meaning of the word in order to be able to choose it from the four options. The Vocabulary Levels Test by contrast has distractors which are not related in meaning or form to the tested word, and thus in the Vocabulary Levels Test learners can get the answer correct if they have a small amount of knowledge of the word. The Vocabulary Size Test is thus a little more demanding than the Vocabulary Levels Test.

The fourteen levels of the test are a way of organising the items in the test so that the test begins with the items more likely to be known. It is not necessary to make learners sit all fourteen levels when the test is used with elementary or intermediate learners, but they should sit a few levels beyond their present level. This is because frequency level is not a perfect indicator of which words are likely to be known. Frequency level is strongly related to the likelihood of a word being known (scores at each level drop as a learner progresses through the test), but there are other factors involved in knowing a word and frequency counts can give differing results depending on the size and nature of the corpus used. In other words, a learner with a vocabulary size of 3,000 words is likely to know some words beyond this level and is likely not to know some words within this level.

Making the multiple-choice test items

The choices were all written using a restricted vocabulary. For the first and second 1000 sample, only words from the first 1000 of West's (1953) General Service List were used. As far as possible, the words in the definitions were of higher frequency than the item being defined, but for the highest frequency items, this was not always possible, for example, there was no possibility for defining *time* except with words of lower frequency (e.g., *hours*). For words in samples from 3000 upwards, the defining words were drawn from the first 2000 of West's General Service List. Occasionally it was necessary to use some other

item, but the frequency of the defining word and the item were always checked in the British National Corpus, and the defining word was always significantly more frequent than the item being defined. An example of this is *haunt*, where it was necessary to use *ghost* in the definition.

Each item was put in a non-defining context. All the definitions were then required to be substitutable in the context sentence. The contexts were chosen to reflect the most frequent environments for the item. Thus with *instance*, it was clear that the very high frequency of this arose from the phrase *for instance*, and so this was used as the context. Where the plural of an item was significantly more frequent than the singular, the context was made plural (e.g., *standard*). The part of speech chosen for the item was also a reflection of the highest frequency environment. A lot of care was put into making the distractors so that they were genuine choices and they were carefully checked in pilot testing.

Interpreting the test results

Because there are ten items at each 1000 word level, each item in the test represents 100 word families. If a test-taker got every item correct, then it is assumed that that person knows the most frequent 14,000 word families of English. A test-taker's score needs to be multiplied by 100 to get their total vocabulary size up to the 14th 1000 word family level.

Because the test is a measure of *receptive* vocabulary size, a test-taker's score provides little indication of how well these words could be used in speaking and writing. In addition, although vocabulary knowledge is the most important factor affecting the readability of a text (Klare, 1974), a test-taker's score is only a rough indication of how well a learner can read.

The greatest value of the test will be in measuring learners' progress in vocabulary learning. The most frequent 14,000 words of English along with proper nouns account for over 99% of the running words in written and spoken text (Nation, 2006). Although adult native speakers' vocabularies are much larger than 14,000 words, these 14,000 words include all the most important words.

Initial studies using the test indicate that undergraduate non-native speakers successfully coping with study at an English speaking university have a vocabulary of around 5,000-6,000 word families. Similarly competent non-native speaking doctoral students have around a 9,000 word vocabulary. At present data is being gathered to assess the reliability and validity of the test. The test is available at <www.vuw.ac.nz/lals/staff/paul-nation/nation.aspx> and <www.lextutor.ca>.

References

Bauer, L., & Nation, I. S. P (1993). Word families. *International Journal of Lexicography*, *6*(4), 253-279.

Beglar, D., & Hunt, A. (1999). Revising and validating the 2000 word level and the university word level vocabulary tests. *Language Testing*, *16*(2), 131-162.

Bertram, R., Baayen, R., & Schreuder, R. (2000). Effects of family size for complex words. *Journal of Memory and Language*, *42*, 390-405.

Bertram, R., Laine, M., & Virkkala, M. (2000). The role of derivational morphology in vocabulary acquisition: Get by with a little help from my morpheme friends. *Scandinavian Journal of Psychology*, *41*(4), 287-296.

Henning, G. (1991). *A study of the effects of contextualization and familiarization on responses to the TOEFL vocabulary test items*. Princeton, NJ: Educational Testing Service.

Hu, M., & Nation, I. S. P. (2000). Vocabulary density and reading comprehension. *Reading in a Foreign Language*, *13*(1), 403-430.

Klare, G. R. (1974). Assessing readability. *Reading Research Quarterly*, *10*(1), 62-l02.

Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English*. Harlow: Longman.

Meara, P., & Jones, G. (1987). Tests of vocabulary size in English as a foreign language. *Polyglot*, *8*(1), 1-40.

Meara, P., & Jones, G. (1990). *Eurocentres vocabulary size test. 10KA*. Zurich: Eurocentres.

Nagy, W. E., Anderson, R., Schommer, M., Scott , J. A., & Stallman, A. (1989). Morphological families in the internal lexicon. *Reading Research Quarterly*, *24*(3), 263-282.

Nation, I. S. P. (1983). Testing and teaching vocabulary. *Guidelines*, *5*(1), 12-25.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, *63*(1), 59-82.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the vocabulary levels test. *Language Testing*, *18*(1), 55-88.

Thorndike, E. L., & Lorge, I. (1944). *The teacher's word book of 30,000 words*. New York: Teachers College Columbia University.

West, M. (1953). *A general service list of English words*. London: Longman, Green & Co.

Appendix

The appendix can be viewed online at <jalt-publications.org/tlt/resources/2007/0707a.pdf>