(2009). In M.H. Long & C.J. Doughty (Eds.), The handbook of language teaching (pp. 543-559). West Sussex, UK: Wiley-Blackwell)

28 Teaching and Testing Vocabulary

PAUL NATION AND
TERESA CHUNG

A well-planned vocabulary component of a language course has the following features: it focuses on the appropriate level of vocabulary, it provides a balanced range of opportunities for learning, and it monitors and assesses the learners' vocabulary knowledge in useful ways. Put another way, a well-planned program answers these questions: What vocabulary? How should vocabulary be taught and learned? How should vocabulary knowledge and growth be assessed?

What Vocabulary?

An important step in planning the vocabulary component of a language course is deciding what vocabulary goals to set. A useful way of doing this is to find out how much vocabulary is needed to do certain tasks without assistance, like read a newspaper, take part in a conversation, or watch a movie. Research with second language learners by Hu and Nation (2000) and with first-language learners by Carver (1994) indicates that at least 98 percent coverage of the running words (tokens) is needed for unassisted reading. This means that there should not be more than one unknown word in every 50 running words.

In earlier studies there was a lot of guesswork and extrapolation involved in making such calculations (Hirsh & Nation, 1992; Sutarsyah, Nation, & Kennedy, 1994). However, there are now lists of the most frequent 14,000 words of English that can be used with the RANGE program (http://www.vuw.ac.nz/lals/staff/paul-nation/nation.aspx), and that are also incorporated into a web-based text analysis program (http://www.lextutor.ca/vp/). These programs can be used to determine the amount of vocabulary needed to reach 98 percent coverage (see Table 28.1, based on Nation, 2006).

The figures in Table 28.1 assume that proper nouns are known or that not knowing them does not hinder the comprehension of a text. The figures indicate that a vocabulary size of around 8,000–9,000 words is needed to deal successfully with a range of language uses (Adolphs & Schmitt, 2004; Nation, 2006). Even the

Table 28.1 Vocabulary size needed to reach 98 percent coverage in a variety of texts

Type of text	*Vocabulary size*
Children's movies	6,000 words
Conversation	7,000 words
Newspapers	8,000 words
Novels	9,000 words

Table 28.2 Percentage coverage by fourteen 1,000-word family lists of three newspaper corpora

Vocabulary level	Independent *(UK)*	New York Times	Dominion Post *(NZ)*
1st 1,000	76.59	75.72	75.11
2nd 1,000	8.68	8.38	8.96
3rd 1,000	2.86	2.66	3.23
4th 1,000	2.21	2.64	2.47
5th 1,000	1.27	1.24	1.38
6th 1,000	0.82	0.86	0.89
7th 1,000	0.55	0.66	0.56
8th 1,000	0.42	0.49	0.53
9th 1,000	0.31	0.35	0.31
10th 1,000	0.21	0.30	0.30
11th 1,000	0.27	0.20	0.23
12th 1,000	0.12	0.16	0.13
13th 1,000	0.13	0.13	0.09
14th 1,000	0.11	0.11	0.10
Proper nouns	4.29	4.62	4.15

most optimistic measures of the vocabulary growth of foreign or second language learners indicate that it takes at least a year, and usually much longer, to increase vocabulary size by a thousand words (Laufer & Paribakht, 1998; Nurweni & Read, 1999). It is thus necessary to be very strategic about what vocabulary is learned first, and what next, when trying to reach this goal. Table 28.2 provides data to reinforce this idea, by showing the vocabulary coverage of each successive 1,000 frequency and range-related words for various collections of newspaper texts.

The three newspaper corpora consist of roughly equal amounts of text from *The Independent* (a British newspaper), *The New York Times*, and *The Dominion Post* (a New Zealand newspaper) within a three-month period. Each of the three corpora contains around 194,000 words.

The figures in Table 28.2 show that by far the largest amount of text coverage comes from the first 1,000 words, and that the coverage figures drop rapidly, so that on average, the fifth 1,000 words cover less than 1.4% of the tokens, and the ninth 1,000 0.35% or less of the tokens. For the *Independent*, a 5,000-word family vocabulary plus proper nouns provides 96% coverage. A 9,000-word vocabulary plus proper nouns is needed to reach 98% coverage.

There are two important conclusions to be drawn from these figures. First, it makes sense in terms of cost–benefit analysis to learn the vocabulary of English roughly in order of its frequency of occurrence. There is much more value to be gained from learning the second 1,000 words than the fourth 1,000 words. Secondly, the very high coverage provided by the first and second 1,000 words suggests that these words deserve a great deal of intensive attention of various kinds. Most of these 2,000 words occur with high frequency, no matter what kind of text is being focused on.

Vocabulary levels

A way to increase the efficiency of the vocabulary focus is to make use of specially designed vocabulary lists. A common way to do this is to distinguish four vocabulary levels: high-frequency, academic, technical, and low-frequency words.

(1) *High-frequency words.* These make up a group of around 2,000 word families. The classic list is Michael West's *A General Service List of English Words* (1953). This list needs updating (it does not contain words like *computer, email, internet*), but it still works reasonably well, and was made with young learners of English in mind. The first 2,000 words from the British National Corpus have a more adult and more formal flavor because of the nature of the British National Corpus (Nation, 2004a). Typically, the first 2,000 words of English cover between 80% and 90% of the running words in a text, depending on the type of text.

(2) *Academic words.* For learners who wish to do academic study through the medium of English in senior high school or in tertiary education, the next step after the high-frequency words is Coxhead's (2000) Academic Word List. This is a list of 570 word families that are very common across a wide range of academic disciplines. It covers around 10% of the running words in academic text, around 4% of the running words in newspapers, and less than 2% of the running words in novels. It is clearly a somewhat specialized vocabulary. There have been several textbooks produced to teach the Academic Word List; see, for example, Schmitt and Schmitt (2005) and Huntley (2006).

(3) *Technical words.* For learners who have very specific study or work purposes, the next level of vocabulary consists of technical words. These are words that are very closely associated with a specialist area, in the way the words *dwang, truss, nog, eaves* are associated with building, or that the words *negotiation, interaction, phoneme, token* are associated with applied linguistics.

Research on technical vocabulary (Chung & Nation, 2003, 2004) shows that the amount of technical vocabulary in specialized texts has generally been

underestimated. Chung found that around 30% of the running words in an anatomy text were technical words, and around 20% of the words in an applied linguistics text were technical. Technical vocabulary can be found with around 90% accuracy by comparing the frequency of vocabulary in a technical text with the frequency of the same vocabulary in a large mixed corpus that does not contain any texts from the technical area being studied. Vocabulary that only occurs in the technical text and vocabulary that is over 50 times more frequent in the technical text is likely to be technical vocabulary. Technical vocabulary can come from high-frequency vocabulary, academic vocabulary, or low-frequency vocabulary.

(4) *Low-frequency words.* The remaining words of the language are low-frequency words. There are thousands of these of varying frequency, and as we have seen by looking at the coverage of the British National Corpus lists, many need to be learned, so that learners can reach the 98% coverage of text required for unassisted language use.

Multi-word units

Focusing on individual words has been criticized in course design because words often do not make much sense unless they are in phrases or larger units. Also, learning words in multi-word units means that they are learned with the collocations and grammar with which they need to be used.

There are four major productive effects of learning multi-word units.

1 Learners will be able to produce grammatically correct utterances.
2 Learners will be able to produce utterances that are nativelike.
3 Learners will be able to produce utterances fluently.
4 Learners will be able to communicate very early in their language learning.

Behind all these effects is the idea that learners will be able to make use of instances of grammatical features without having to have control of the system that might be needed to make creative use of the features. These effects also have their receptive equivalents.

The effects have been pointed out by writers like Palmer (1925): "What is then the fundamental guiding principle of . . . those who are anxious to become proficient in foreign conversation? – *Memorize perfectly the largest number of common and useful word groups!*" This idea is also supported by Pawley and Syder (1983), and was popularized by Lewis (1993, 1997) in his lexical approach, which advocates the learning of multi-word units as one of the major focuses of language learning.

A fundamental problem with discussion about multi-word units has been the proliferation of terminology that has been very poorly defined. Without good definitions, it is not possible to have replicable corpus searches. Where clear and consistent criteria were set up to classify multi-word units, the criteria tended to be ones of convenience to meet the needs of computer searches. More recently,

however, there have been attempts to establish clear criteria and to apply them in a systematic way.

Grant and Bauer (2004) used the two criteria of compositionality (Can the meaning of the whole be related to the meaning of the parts?) and figurativeness to classify three major kinds of multi-word units: core idioms, figuratives, and literals. The two criteria are applied following the steps in Figure 28.1. First the item is checked to see if it is compositional, that is, do the meanings of the parts make up the meaning of the whole? If they do, it is classified as a literal, for example, *at six o'clock*. Literals are thus compositional and non-figurative. The meaning of the whole can be understood from the meaning of the parts. There are thousands of these in English, and further criteria like frequency and grammatical well-formedness, are needed to distinguish them from items that are not usefully regarded as multi-word units.

If the item is non-compositional, it is then checked to see if it is a figurative. A figurative is a multi-word unit where the meanings of the parts do not give you the actual meaning of the whole. When a learner meets the clause, "We decided to kill two birds with one stone," in context, this may be initially confusing because the text so far has had nothing to do with birds, stones, or killing. To a figurative is non-compositional. However, by applying a strategy of interpretation, it is possible to see how the literal meaning of *kill two birds with one stone* reflects its figurative meaning "to do two jobs at once." Some figuratives are quite easy to interpret, e.g., *We want to make sure we are singing from the same hymn sheet*, while others require considerable background knowledge, e.g., *You have to take it with a grain of salt*. There are a lot of figuratives in English, and new ones are created every day.

If the item is not a figurative, it is classified as a core idiom. A core idiom is a multi-word unit where the meaning of the parts does not make up the meaning

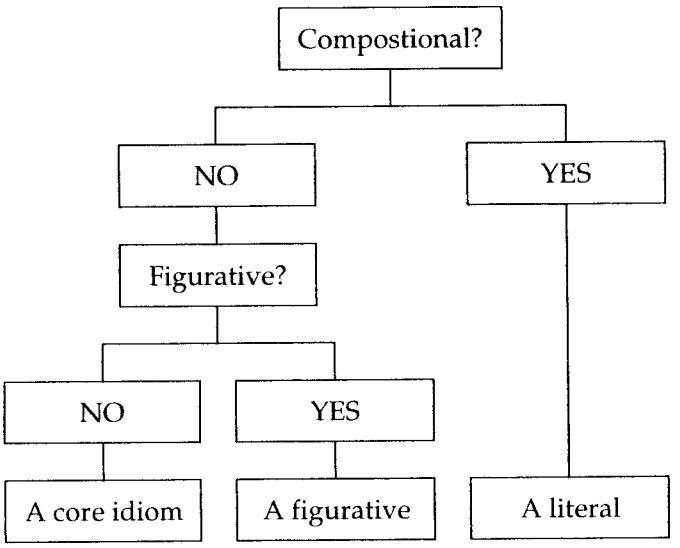

Figure 28.1 Classification of multi-word units

of the whole. In addition, the meaning of the whole cannot be arrived at by applying a figurative interpretation to the unit. So, if we meet the phrase *as well as*, as in *you as well as me*, knowing the meanings of *as*, and *well* does not help us understand the phrase. The term *core idiom* is used to separate it from the many looser uses of the word *idiom*. When the criteria of compositionality and figurativeness are carefully applied, there turn out to be just over 100 core idioms in English (Grant & Nation, 2006).

The researcher who has done the most to investigate the teaching and learning of figuratives is Frank Boers. The research of Boers and his colleagues has provided very useful guidelines that not only apply to figuratives but also to core idioms and literals.

1 The use of figurative sequences in speaking results in higher oral proficiency scores (Boers et al., 2006).
2 The source domains of idioms and figuratives differ from one language to another and reflect the culture and history of the speakers of that language (Boers, Demecheleer, & Eyckmans, 2004).
3 Figuratives vary considerably in the transparency of the expression (for example, *rock the boat* compared with *out for a duck*) (Boers, Eyckmans, & Stengers, in press).
4 Giving deliberate attention to figurative sequences can help their learning (Boers et al., 2006).
5 Associating a figurative with its source domain (sailing, cricket, war, etc.) and its etymology helps in understanding the expression and helps learning (Boers, Eyckmans, & Stengers, 2006).
6 Around 20 percent of figurative expressions make use of alliteration (*through thick and thin*) and assonance (*high and dry*) to some degree, and focusing on this helps learning.

The research of Boers and his colleagues challenges some of the conventional wisdom. It is often suggested that multi-word units should be learned as unanalyzed wholes. The research does not support this. Etymological elaboration helps learning. It is also suggested that items should be learned through a meaning focus. The research shows that giving conscious attention to form (alliteration and assonance) also helps learning.

The different nature of core idioms, figuratives, and literals suggests different learning focuses. Because core idioms cannot readily be analyzed, they need to be learned as largely unanalyzed wholes. It is possible in some cases, however, to create etymologies for them and these will help learning. Figuratives can be very effectively learned through an interpretive strategy that relates the literal and figurative meanings. While literals may seem straightforward, they can be subdivided into those where the learners' first language has a parallel word-for-word expression and those where the first-language expression is different. For example, the Thai equivalent of *good morning*, *good afternoon*, and *good evening* is one word *sawasdee*. Thus, for Thai learners, these three literals require a bit more

learning effort than the expression *next week* which has a word-for-word equivalent in Thai, *athit na*, although with Thai word order. Shin and Nation (2006) found that about one third of the high-frequency, grammatically well-formed English multi-word units did not have word-for-word equivalents in Korean.

Boers and his colleagues developed three computer-based activities to help the learning of figurative sequences and to use in their studies. These three activities make up their *Idiom teacher*. The first two activities are multiple-choice with feedback. Once the item is answered, the correct answer is given or the learner keeps choosing until the correct answer is chosen. Activity one is *Identify the source*, where learners have to choose from four choices the domain that the phrase comes from, for example, that *close to the wind* comes from sailing. The second activity focuses on comprehension; learners have to choose which of four meanings describes the meaning of the multi-word unit. The third activity provides the expression in a defining context, but with a word missing from the expression which learners have to complete. These test/teach activities proved very effective in getting learners to think about and learn figuratives.

Biber, Conrad, and Cortes (2004) used frequency and range to identify "lexical bundles," and found in the total corpus of over 14 million tokens, 138 four-word sequences occurring at least 40 times per million words. The bundles did not have to be complete structural units (grammatically well-formed). Biber et al. say that "only 15 percent of the lexical bundles in conversation can be regarded as complete phrases or clauses, while less than five percent of the lexical bundles in academic prose represent complete structural units" (p. 377). Well-conducted research like this is providing a good basis for the systematic incorporation of multi-word units into language syllabus design.

How Should Vocabulary Be Taught and Learned?

One way to make sure that opportunities for vocabulary learning are properly balanced in a course is to see that there is a roughly even allocation of learning activities to each of the four strands of meaning-focused input, meaning-focused output, language-focused learning, and fluency development (Nation, 2007). Each strand should receive about the same amount of time in a well-balanced course. This includes what happens both inside and outside the classroom.

Meaning-focused input

Meaning-focused input involves learning via comprehensible input obtained through listening and reading. A very important part of the meaning-focused input strand is an extensive reading program, where learners read large amounts of interesting texts. There has been continued and growing interest in extensive reading. Day and Bamford's (1998) book was followed by a very substantial collection of activities (Bamford & Day, 2004) aimed at encouraging, organizing and monitoring extensive reading. Along with these publications, the Extensive

Reading Foundation has been established (www.erfoundation.org/) with the goal of promoting extensive reading and encouraging the production of high-quality graded readers. The foundation has set up a process for judging and giving awards to various levels of graded readers each year.

This promotion of extensive reading has been accompanied by high-quality research on vocabulary learning from extensive reading. Waring and Takaki's (2003) study of vocabulary learning from one graded reader showed that there are various kinds of levels of vocabulary learning, which, if viewed together, represent a significant increase in learning. Waring and Takaki used three vocabulary tests of differing difficulty to measure incidental vocabulary learning. One was a word-form recognition test, where learners chose words they had seen in the text from those that did not occur in the text. Learners scored about 16 out of 25 on this rather easy test. Another test was a receptive multiple-choice test, where learners were provided with words from the text with four-item first-language choices. Learners scored about 12 out of 25 on this test. The most difficult was a translation test, where learners saw words from the text and had to provide first-language translations for them. Learners scored about 4 out of 25 on this test. Taken together, this means that of the words tested, four were known reasonably well, an additional eight were partly known, and an additional four were on the first step towards being known. This is a very reasonable outcome from a piece of incidental learning that took around 56 minutes and undoubtedly resulted in other kinds of learning and skill improvement, as well. In a very detailed case study, Pigada and Schmitt (2006) found large amounts of incidental vocabulary learning of various strengths from sustained extensive reading. Horst (2005) also found evidence of substantial amounts of learning.

One way of helping reading and encouraging vocabulary learning from reading is to provide glosses of words that are likely to be unknown. A gloss is a short note on the meaning of a word provided after the text or in the margin near where the word occurs. There has been a renewed interest in glossing as a result of the growth of computer-assisted language learning. Reading texts presented on the computer provide opportunities for hypertext links and even direct links to dictionaries. In this way, glossing and dictionary use have moved closer to each other.

In many ways, research on glossing deals with issues that are very important in the wider teaching and learning of vocabulary. These include:

1 The relative advantages of the first language compared with the second language in communicating meaning. Research on first-language learners' understanding of definitions indicates that short, clear definitions are the best (McKeown, 1993). Second language translations are likely to meet this condition for second language readers, particularly lower-proficiency readers. Taylor (2006) suggests that the effectiveness of the language of the gloss may be dependent on the language used to test comprehension of the text, with L2 glosses being more effective if comprehension of the text is tested in L2.
2 The role of deep processing in vocabulary learning. In a very detailed qualitative study, Rott (2005) compared three-item multiple-choice translation glosses

with single translation glosses. Each target item occurred four times in the text, so the effect of repeated encounters was also examined. Multiple-choice glosses led to better long-term retention and more complete comprehension. Rott's study allowed her to explain this advantage which seemed to occur through the use of multiple information sources and multiple processing strategies. Rott's study thus made a major contribution to understanding not only how glossing should be carried out, but also the processes that underlie effective vocabulary learning.

3 The giving of deliberate attention to language features in the context of a message-focused task. Glossing tends to increase vocabulary learning.
4 The effect of multiple encounters with a word. Generally, repetition helps learning, but its effects are not guaranteed.
5 The role of textual enhancement (highlighting, bolding) in learning. Such enhancement tends to improve learning of the form rather than the meaning.
6 The length of retention after the first encounters. Rott, Williams, and Cameron (2002) found that five weeks' delay was too long between encounter and testing.

The activities that encourage vocabulary learning through meaning-focused input include listening to stories where the teacher notes unfamiliar words on the board or quickly explains their meaning, extensive reading, and taking part in interactive speaking and reading activities.

Meaning-focused output

Meaning-focused output involves learning through speaking and writing. If vocabulary is used in generative or creative ways, then memory for these words is strengthened (Joe, 1998). Activities like retelling, role plays, rewriting for a different purpose, and group and pair work involving negotiation can be very effective sources of vocabulary learning (Joe, Nation, & Newton, 1996).

Language-focused learning

Language-focused learning involves the deliberate learning and the deliberate study of vocabulary and vocabulary-learning strategies. In terms of efficiency, the most effective deliberate learning of vocabulary involves the use of small word cards with the target word or phrase on one side and the first-language translation on the other. Nation (2001, pp. 296–316) describes this learning strategy in detail. Strategy training is also a very efficient use of the language-focused learning strand in the classroom. Learners can benefit from training in guessing from context (Walters, 2004, 2006), dictionary use, using word cards, and using word parts.

The direct teaching of vocabulary is not a particularly efficient use of class time, but can be usefully done during intensive reading (Nation, 2004b) and as unknown words occur in the range of classroom activities. Nation (2008) suggests a variety of vocabulary-teaching activities that require various levels of preparation.

Computer-assisted vocabulary learning can be an effective way of getting help with vocabulary. Computer-assisted vocabulary learning covers a wide range of possibilities. First, it can include computer-based analysis of vocabulary, which has resulted in the creation of frequency-based word lists (Nation, 2006), the Academic Word List (Coxhead, 2000), and procedures for determining technical vocabulary (Chung & Nation, 2003; Chujo & Utiyama, 2006). Second, it can include the analysis of texts to determine their suitability for particular learners or to indicate how they should be adapted or sequenced (www.lextutor.ca/vp/). Such analysis can also be used as a way of monitoring the lexical richness of texts produced by learners for research and assessment purposes (Laufer & Nation, 1995; Morris & Cobb, 2004). Third, it can include programs designed for the deliberate learning of vocabulary (Cobb, 1999; Horst, Cobb, & Nicolae, 2005; Mondria & Mondria-de-Vries, 1994). Fourth, it can include the use of text-linked aids, such as spoken support, hypertext, glosses, concordances, dictionary look-ups, and electronic dictionaries, which support reading (Cobb, 1997; Cobb, Greaves, & Horst, 2001). Fifth, it can include the use of word-processing tools, such as highlighting, track changes, comments and hypertext links, to provide feedback on electronically submitted written work (Gaskell & Cobb, 2004). Not surprisingly, there has been an increasing amount of research and the development of applications in each of these areas as computer technology has developed. In this review we will focus on the use of text-linked aids because this is closest to what is commonly considered computer-assisted vocabulary learning, and because there has been useful innovation and research in the area, much of it involving the research of Tom Cobb and his colleagues.

Cobb's website (www.lextutor.ca) has an ever-increasing variety of research tools, resources and learning programs that are freely available to those who want to use them. Let us look at one set of programs before focusing on concordancing, which has interested both researchers and teachers since electronic concordancers first became available. The Hypertext program (www.lextutor.ca/hypertext/) on Cobb's website involves the learner working with an electronic text that can be pasted into the program. The program then provides the following support for reading.

1 *Spoken form.* Clicking once on any word provides the spoken form of the word. In a separate program under the Text-to-Speech heading on Cobb's web page, it is possible to link written texts to their spoken form where these exist. Useful sites for texts available in both spoken and written form can also be found from the Compleat Lexical Tutor website.
2 *Examples in context.* Clicking twice on a word brings up several instances of the word in context (a concordance).

> of the Arkansas. It was hoped that to this post would flow a large quantity of furs from the west, principally do
> among Mr. Coward's more memorable works. The melodies flow along pleasantly, as Mr. Coward's songs usually do,
> be if it was made by a rusty tool; this would stop the flow and also prevent infection. My lawyer told me that his

These extra contexts can be used to help guess the meaning of the unknown word, to gain information about the use of the word (grammar and collocates), and to gain information about the range of senses of the word.

3 *Meaning*. Clicking on a link brings up a substantial dictionary entry for a word from a range of possible dictionaries, including learner dictionaries.

4 *Revision*. Holding down the Alt-key and clicking puts the word in a box at the top of the screen for later revision. These revision activities can include (a) a dictation test, where the word is heard and the learner has to write the word; (b) a meaning test, where a concordance appears but the pivot word is missing and must be chosen from the list in the box.

The concordancer and dictator are also available as stand-alone programs not linked to a text.

Concordancers have long been suggested as tools for vocabulary learning. Study of concordances can provide a range of information about a word, including its meaning, the range of forms it can have, its grammar, its collocates, its frequency and the relative frequency of its various forms, uses and meanings. Seeing a concordance of a word is in some ways similar to meeting examples of the word while reading. The advantage of the concordance is that all these examples can be readily compared with one another and more deliberate generalizations made from them. It also means that a guess at the meaning of a word is more likely to be successful, as there is much more data available. The disadvantages are that the encounters are not spaced (spaced repetition is better for retention than massed repetition), and the study of concordances is time-consuming. Cobb (1997) compared learners using concordances with those doing the same kinds of activities without concordances and found a 12 percent vocabulary-learning advantage for concordance users. There is evidence that students generally enjoy using concordances, but concordance use is still a largely under-researched area.

Fluency development

The fourth strand of a well-balanced course is the strand of fluency development. Fluency development needs to occur in each of the four skills of listening, speaking, reading, and writing. Speaking fluency activities include 4/3/2, where learners deliver the same talk to three successive listeners in a four-minute, then a three-minute, and then a two-minute time frame, and pyramid-ranking activities, where learners deal with the same items to rank in pairs, and then groups, and then as a whole class. Reading fluency activities include timed texts with questions, and repeated reading. Writing activities include ten-minute writing, where learners get positive feedback on quantity and content, but not on form, and writing about things that have already been read about and talked about. Fluency development tends to be a neglected strand in most courses, but it is important that learners not only learn new language items but also are able to access and use them fluently.

How Should Vocabulary Knowledge Be Monitored and Assessed?

Vocabulary testing can be used to see what level of vocabulary learners should be focusing on (a diagnostic goal), to see how much vocabulary learners know and how well they know it (a proficiency goal), and to see what vocabulary they have recently learned and how well they have learned it (an achievement goal).

Diagnostic testing

The *Vocabulary Levels Test* (Read, 2000, pp. 118–26; Schmitt, Schmitt, & Clapham, 2001) was designed to look at high-frequency, academic, and low-frequency vocabulary. Although it has been used to measure vocabulary size, it was not designed to do that. Because the Vocabulary Levels Test is a monolingual test, there is no section testing the first 1,000 words, as the meanings of these words cannot be represented using more frequent vocabulary than the tested words. The first 1,000 words is an extremely important group of words of which many learners may have poor knowledge. Because of this, some bilingual tests of the first 1,000 and the second 1,000 words have been developed, where the meanings of the tested words are given in the first language of the test takers. The Vocabulary Levels Test is also available online at www.lextutor.ca.

Proficiency testing

Vocabulary proficiency testing has typically involved the measurement of vocabulary size, but there are now innovative approaches to measuring how well vocabulary is known and how diverse a vocabulary learners actually use. The distinction between how many words are known and how well they are known has been described as the distinction between breadth of vocabulary knowledge and depth of vocabulary knowledge. Read (2004) has usefully distinguished between several meanings of depth – precision of meaning, comprehensive word knowledge, and network knowledge.

Read's (1993, 2000, pp. 180–6) Vocabulary Associates Test is a well-researched example of a network knowledge measure. It involves items like the following, where learners need to choose elements of meaning and collocates of the test item. Each item involves choosing four words out of the eight provided.

sound
(A) logical (B) healthy (C) bold (D) solid (E) snow (F) temperature (G) sleep (H) dance

The test and variants of it have been found to be reliable and results to correlate well with a measure of reading comprehension (Qian & Schedl, 2004).

There has been a range of innovative attempts to measure the diversity of vocabulary used in learners' writing and speech. One approach is to calculate the number of different words (types or lemmas) used in relation to the number of running words (tokens) used. An early measure of this was the type–token ratio, but this has been found to be strongly related to text length. Malvern and Richards (1997; Duran et al., 2004) have developed a much more sophisticated measure, D, that avoids the weaknesses of the type–token ratio. This measure has fuelled a resurgence of research in this area (Daller, van Hout, & Treffers-Daller, 2003). Software using D (vocd) is available at http://childes.psy.cmu.edu as a part of the CLAN suite of programs for the analysis of children's language. D measures "the extent to which the active vocabulary is employed and how richly it is deployed" (Duran et al., 2004). See Jarvis (2002) for a positive critique of D.

The lexical frequency Profile (Laufer & Nation, 1995) attempts to relate vocabulary diversity to word frequency levels. In this way it may have diagnostic value as well as being a proficiency measure. There has been debate about the relative merits of this kind of measure and those based on modeling (Laufer, 2005; Meara, 2005). Meara has also suggested other innovative methods for measuring productive vocabulary (Meara & Bell, 2001; Meara & Fitzpatrick, 2000).

Achievement testing

Laufer and her colleagues (Laufer & Goldstein, 2004; Laufer et al., 2004) examined four kinds of tests that can be used in both monolingual and bilingual versions. In the examples below, the monolingual example is given first. Laufer and Goldstein used a bilingual test, Laufer et al. monolingual tests.

1 Active recall (supplying a form for a given meaning; the first letter of the tested word is given to prevent learners from supplying non-target words)

Turn into water *m*_____ (monolingual)
*m*_____ mencairkan (bilingual)

2 Passive recall (supplying the meaning for a given form; the first letter of the translation is given)

When something *melts it turns into* _____.
Translate the following words into Indonesian.
melt

3 Active recognition (choosing the target word form from four options)

Turn into water a. elect b. blame c. melt d. threaten
Select the correct translation for the following words.
mencairkan a. elect b. blame c. melt d. threaten

4 Passive recognition (choosing the meaning of the target word from four options)

Melt a. choose b. accuse c. make threats d. turn into water
Melt a. menolong b. mencairkan c. memeriksa d. memandang

In the two studies, Laufer et al. found:

1 There was a hierarchy of difficulty in the order given above from the most difficult, active recall, to the least difficult, passive recognition.
2 The four formats were significantly different from each other. Active recall was very difficult compared with the other formats; the difference between active and passive recognition was much smaller.
3 Knowledge did not seem to grow at an even pace in the four strength measures. The more difficult recall formats seemed to take much longer to show growth.
4 Vocabulary knowledge is not an all-or-nothing phenomenon. Knowledge of words develops cumulatively and there is value in having tests that show this change in strength of knowledge.

An important message from this research for achievement testing is that there is a variety of test formats that could be used and which differ from each other in difficulty. It is thus very important to consider the level of knowledge that learners are likely to have when choosing a format to measure this knowledge. Choosing a very difficult format, such as active recall, could underestimate learning. Choosing a format that is too easy may not give credit for additional knowledge that learners have. When considering the type of item to choose, the following factors are important.

1 How large is the learners' vocabulary? If their vocabulary size is small, bilingual items would be better than monolingual items.
2 What kind of learning did the learners do? If the learning was through reading or listening, then passive measures are more suitable. If substantial attention was given to the learning and active retrieval was used, then active measures are more suitable.
3 How long ago did the learning occur? The longer the time gap, the easier the item format needs to be.
4 How difficult does the teacher want the test to be?

This review of teaching, learning and testing vocabulary has shown the increasing effect of computing technology on vocabulary research and teaching. This effect is apparent in corpus-based research, in providing aids to assist reading and learning vocabulary from reading, and in the deliberate learning of vocabulary. Undoubtedly, this effect will increase as ways of dealing with multi-word units become more sophisticated.

REFERENCES

Adolphs, S. & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics* 24, 4, 425–38.

Bamford, J. & Day, R. R. (eds.) (2004). *Extensive reading activities for teaching language.* Cambridge: Cambridge University Press.

Biber, D., Conrad, S., & Cortes, V. (2004). "If you look at . . .": Lexical bundles in university teaching and textbooks. *Applied Linguistics* 25, 3, 371–405.

Boers, F., Demechleer, M., & Eyckmans, J. (2004). Cross-cultural variation as a variable in comprehending and remembering figurative idioms. *European Journal of English Studies* 8, 3, 375–88.

Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting the Lexical Approach to the test. *Language Teaching Research* 10, 3, 245–61.

Boers, F., Eyckmans, J., & Stengers, H. (2006). Presenting figurative idioms with a touch of etymology: More than mere mnemonics. *Language Teaching Research* 11, 43–62.

Carver, R. P. (1994). Percentage of unknown vocabulary words in text as a function of the relative difficulty of the text: Implications for instruction. *Journal of Reading Behavior* 26, 4, 413–37.

Chujo, K. & Utiyama, M. (2006). Selecting level-specific specialized vocabulary using statistical measures. *System* 34, 255–69.

Chung, T. M. & Nation, P. (2003). Technical vocabulary in specialised texts. *Reading in a Foreign Language* 15, 2, 103–16.

Chung, T. M. & Nation, P. (2004). Identifying technical vocabulary. *System* 32, 2, 251–63.

Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System* 25, 3, 301–15.

Cobb, T. (1999). Breadth and depth of vocabulary acquisition with hands-on concordancing. *Computer Assisted Language Learning* 12, 4, 345–60.

Cobb, T., Greaves, C., & Horst, M. (2001). Can the rate of lexical acquisition from reading be increased? An experiment in reading French with a suite of on-line resources. In P. Raymond & C. Cornaire, *Regards sur la didactique des langues secondes* (pp. 133–53). Montréal: Éditions logique.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly* 34, 2, 213–38.

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

Day, R. R. & Bamford, J. (1998). *Extensive reading in the second language classroom.* Cambridge: Cambridge University Press.

Duran, P., Malvern, D., Richards, B., & Chipere, N. (2004). Developmental trends in lexical diversity. *Applied Linguistics* 25, 2, 220–42.

Gaskell, D. & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System* 32, 3, 301–19.

Grant, L. & Bauer, L. (2004). Criteria for redefining idioms: Are we barking up the wrong tree? *Applied Linguistics* 25, 1, 38–61.

Grant, L. & Nation, I. S. P. (2006). How many idioms are there in English? *ITL – International Journal of Applied Linguistics* 151, 1–14.

Hirsh, D. & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language* 8, 2, 689–96.

Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *Canadian Modern Language Review* 61, 3, 355–82 .

Horst, M., Cobb, T., & Nicolae, I. (2005). Expanding academic vocabulary with an interactive on-line database. *Language Learning and Technology* 9, 2, 90–110.

Hu, M. & Nation, I. S. P. (2000). Vocabulary density and reading comprehension. *Reading in a Foreign Language* 13, 1, 403–30.

Huntley, H. (2006). *Essential academic vocabulary*. Boston: Houghton Mifflin.

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

Joe, A. (1998). What effects do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied Linguistics* 19, 3, 357–77.

Joe, A., Nation, P., & Newton, J. (1996). Vocabulary learning and speaking activities. *English Teaching Forum* 34, 1, 2–7.

Laufer, B. (2005). Lexical frequency profiles: From Monte Carlo to the real world. A response to Meara (2005). *Applied Linguistics* 26, 4, 582–8.

Laufer, B., Elder, C., Hill, K., & Congdon, P. (2004). Size and strength: Do we need both to measure vocabulary knowledge? *Language Testing* 21, 2, 202–26.

Laufer, B. & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning* 54, 3, 399–436.

Laufer, B. & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics* 16, 3, 307–22.

Laufer, B. & Paribakht, T. S. (1998). The relationship between passive and active vocabularies: Effects of language learning context. *Language Learning* 48, 3, 365–91.

Lewis, M. (1993). *The lexical approach*. Hove, UK: Language Teaching Publications.

Lewis, M. (1997). *Implementing the lexical approach*. Hove, UK: Language Teaching Publications.

Malvern, D. & Richards, B. (1997). A new measure of lexical diversity. In A. Ryan & A. Wray (eds.), *Evolving models of language* (pp. 58–71). Clevedon, UK: Multilingual Matters.

McKeown, M. G. (1993). Creating effective definitions for young word learners. *Reading Research Quarterly* 28, 1, 17–31.

Meara, P. (2005). Lexical frequency profiles: A Monte Carlo analysis. *Applied Linguistics* 26, 1, 32–47.

Meara, P. & Bell, H. (2001). P_Lex: A simple and effective way of describing the lexical characteristics of short texts. *Prospect* 16, 3, 5–19.

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

Mondria, J.-A. & Mondria-de Vries, S. (1994). Efficiently memorizing words with the help of word cards and "hand computer": Theory and applications. *System* 22, 1, 47–57.

Morris, L. & Cobb, T. (2004). Vocabulary profiles as predictors of the academic performance of Teaching English as a Second Language trainees. *System* 32, 75–87.

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

Nation, I. S. P. (2004a). A study of the most frequent word families in the British National Corpus. In P. Bogaards & B. Laufer (eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 3–13). John Benjamins, Amsterdam.

Nation, I. S. P. (2004b). Vocabulary learning and intensive reading. *EA Journal* 21, 2, 20–9.

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

Nation, I. S. P. (2007). The four strands. *Innovation in Language Learning and Teaching* 1–12.
Nation, I. S. P. (2008). *Teaching vocabulary: Strategies and techniques*. New York: Thomson Heinle.
Nurweni, A. & Read, J. (1999). The English vocabulary knowledge of Indonesian university students. *English for Specific Purposes* 18, 2, 161–75.
Palmer, H. E. (1925). Conversation. In R. C. Smith (1999) *The Writings of Harold E. Palmer: An Overview* (pp. 185–91). Tokyo: Hon-no-Tomosha.
Pawley, A. & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (eds.), *Language and communication* (pp. 191–225). Longman, London.
Pigada, M. & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language* 18, 1, 1–28.
Qian, D. & Schedl, M. (2004). Evaluation of an in-depth vocabulary knowledge measure for assessing reading performance. *Language Testing* 21, 1, 28–52.
Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing* 10, 3, 355–71.
Read, J. (2000). *Assessing vocabulary*. Cambridge : Cambridge University Press.
Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined? In P. Bogaards & B. Laufer (eds.), *Vocabulary in a second language: Selection, acquisition and testing* (pp. 209–27). Amsterdam: John Benjamins.
Rott, S. (2005). Processing glosses: A qualitative exploration of how form–meaning connections are established and strengthened. *Reading in a Foreign Language* 17, 2, 95–124.
Rott, S., Williams, J., & Cameron, R. (2002). The effect of multiple-choice glosses and input-output cycles on lexical acquisition and retention. *Language Teaching Research* 3, 183–222.
Schmitt, D. & Schmitt, N. (2005). *Focus on vocabulary*. White Plains, NY: Longman Pearson Education.
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.
Shin, D. & Nation, I. S. P. (2006). Beyond single words: The most frequent collocations in spoken English. *ELT Journal* 62, 4, 339–48.
Sutarsyah, C., Nation, P., & Kennedy, G. (1994). How useful is EAP vocabulary for ESP? A corpus based study. *RELC Journal* 25, 2, 34–50.
Taylor, A. (2006). Factors associated with glossing: Comments on Ko (2005). *Reading in a Foreign Language* 18, 1, 72–3.
Walters, J. (2004). Teaching the use of context to infer meaning: A longitudinal survey of L1 and L2 vocabulary research. *Language Teaching* 37, 243–52.
Walters, J. (2006). Methods of teaching inferring meaning from context. *RELC Journal* 37, 176–90.
Waring, R. & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language* 15, 2, 130–63.
West, M. (1953). *A general service list of English words*. London: Longman, Green & Co.