

The development of vocabulary breadth across the CEFR levels.

A common basis for the elaboration of language syllabuses, curriculum guidelines, examinations, and textbooks across Europe.

James Milton
Swansea University, UK

This chapter attempts to attach measurements of vocabulary breadth, the number of words a learner knows in a foreign language, to the six levels of the Common European Framework of reference for Languages (CEFR). The details of the Framework document (Council of Europe, 2001) indicate that vocabulary breadth ought to be a useful metric in the description of the levels and that, broadly, it would be expected that as language level increases so would the learner's knowledge of vocabulary and the sophistication with which that vocabulary can be used. The evidence we have from vocabulary size tests is reviewed and confirms this assumption, and suggests the actual volumes of vocabulary that are associated with each CEFR level. This information should be very useful to learners, teachers and other users of the CEFR is helping to link language performance to the CEFR levels. The evidence also appears to suggest that vocabulary breadth may vary from one language to another but it is not yet clear whether this reflects differences between the languages themselves, or differences in the construction of the corpora from which vocabulary size tests are derived.

1. Introduction

This chapter addresses the principal aim of SLATE, which is to determine 'which linguistic features of learner performance (for a given target language) are typical at each of the six CEFR levels?' (see Hulstijn, Alderson, & Schoonen, this volume; see also "Aims of SLATE," n.d.). It attempts to identify, the scale of vocabulary knowledge which is typical at each of the six levels of the Common European Framework of Reference for foreign languages (CEFR). It addresses, therefore, an issue which the creators of the CEFR themselves raise in pointing out that 'users of the Framework may wish to consider ... what size of vocabulary (i.e. the number of words and fixed expressions) the learner will need...' in seeking to attain a particular level of performance (Council of Europe, 2001, p. 150). And the CEFR document further suggests, 'an analysis

of the ... vocabulary necessary to perform the communicative tasks described on the scales could be part of the process of developing a new set of language specifications' (Council of Europe, 2001, p. 33). In addressing this issue, therefore, this chapter also addresses the second of the research issues SLATE identifies and attempts to contribute to a linguistic tool kit for diagnosing learners' proficiency levels by examining the number of words in their foreign language that learners at each CEFR level typically know. This is potentially very useful for teachers and learners and will make the process of assigning learners to CEFR levels quicker and, potentially, more accurate. It should help, too, to make the CEFR more robust by adding detail to the levels descriptors.

This chapter will begin by considering what the CEFR framework says about vocabulary knowledge and the way it is expected to develop as learners improve in competence. Broadly, this suggests that language learners, as they progress through the levels of the CEFR, will grow increasingly large, and increasingly complex, lexicons in the foreign language. This relationship between vocabulary knowledge and overall competence in a foreign language is supported by research that suggests that vocabulary knowledge is key to both comprehension and communicative ability (e.g. Stæhr, 2008). While vocabulary knowledge and general linguistic performance are separable qualities, given that the number of words a learner knows is not the sole determiner of how good he or she is in communication, they are not entirely separate qualities. A learner's vocabulary can be expected to become measurably larger and more sophisticated as communicative competence increases. The potential for this as a diagnostic tool is obvious since if vocabulary knowledge can be measured, then learners may be quickly and easily linked to the relevant CEFR level. Such a measure would not provide details of every aspect of linguistic performance, of course, but might in addition to providing a placement within the framework for vocabulary knowledge be a useful general measure. The methodology for measuring vocabulary knowledge will be explained and this involves an understanding of what is meant by 'word' in this context. Current methodology allows the numbers of words learners know in a foreign language to be estimated with some confidence, and these measurements appear particularly useful in making broad assessments of learner level. The measurements we have of vocabulary size and which are linked to the CEFR levels will be presented and examined.

2. Vocabulary within CEFR descriptors

Some of the early materials relating to the CEFR contained great detail about the vocabulary associated with performance at some of the six levels. At what is now called the B1 level, given several names at the time such as *Threshold* and

Niveau Seuil, there are several word lists available for this level (for example, Coste, Courtillon, Ferenczi, Martins-Baltar, & Papo, 1987; Van Ek & Trim, 1991). These lists typically contain about 2000 words. At what is now A2 level, called *Waystage* at the time in English, materials also included wordlists (for example Van Ek, 1980) and these were, as might be expected, smaller in size than the B1 level lists with about 1000 words. In each case the wordwere derived from notional functional areas which were deemed appropriate to these levels, such as clothing and what people wear, personal identification, and routines in daily life. Adumbrating the words that should be known in word lists had the serious drawback, however, of prescribing the language for each level in a way that restricted the flexibility of the system and its ability to be applied across the huge variety of language courses and language learning that takes place in Europe, and even across the different languages that are used in Europe. The 2001 CEFR document makes the argument that ‘descriptors need to remain holistic in order to give an overview; detailed lists of micro-functions, grammatical forms and vocabulary are presented in language specifications for particular languages (e.g. Threshold level, 1990)’ (Council of Europe, 2001, p. 30). The word lists have not been abandoned or disowned in anyway by the CEFR, therefore, but a different and more all-inclusive approach to language description has been adopted. Current descriptions of the CEFR level have, therefore, defined the levels in terms of skills, language activities or communicative goals (Council of Europe, 2001). The current descriptions are flexible and inclusive and by being general they can apply across different languages more readily than the separate lists for individual languages were capable of doing.

The new levels descriptors sometimes include reference to the vocabulary that might be expected of learners performing certain skills and this is illustrated in samples of A1 and B1 level descriptors, provided in Table 1, which are taken from the Council of Europe’s (2001) description of the CEFR. These include, in the A1 listening and reading descriptors, reference to the recognition and comprehension of ‘familiar words’, and in the B1 reading descriptors reference to the understanding of ‘high frequency or everyday job-related vocabulary’. The terminology is couched in a form to give a broad characterisation but may be hard to apply in practice. What are these familiar words and what is everyday vocabulary?

The CEFR document also includes details of the vocabulary range and vocabulary control which are expected of learners at each level of the hierarchy. The vocabulary range criteria are presented in Table 2. This is likewise a series of general characterisations, for example, how broad should a lexical repertoire be before it is broad enough to fit the C level descriptors? Would a few thousand words be sufficient or is the learner expected to know the several tens of thousands which native speakers are reputed to have (D’Anna, Zechmeister, &

Table 1. A1 and B1 level descriptors from Council of Europe (2001, pp. 26–27)

LEVEL	LISTENING	READING	WRITING
A1	I can recognise familiar words and very basic phrases concerning myself, my family and immediate concrete surroundings when people speak slowly and clearly.	I can understand familiar names, words and very simple sentences, for example on notices and posters or in catalogues.	I can write a short, simple postcard, for example, sending holiday greetings. I can fill in forms with personal details, for example entering my name, nationality and address on a hotel registration form
B1	I can understand the main points of clear standard speech on familiar matters regularly encountered in work, school, leisure etc.	I can understand texts that consist of mainly high frequency or everyday job-related language. I can understand the description of events, feelings and wishes in personal letters.	I can write simple connected text on topics which are familiar or of personal interest. I can write personal letters describing experiences and impressions.

Hall, 1991; Goulden, Nation, & Read, 1990)? Again, at what point does a learner's vocabulary knowledge pass from being sufficient for self-expression, at B1 level, to being good at B2 level? A further question arises as to how learners are to demonstrate this knowledge when the tasks presented to them, written essays or oral interviews for example, only allow them to produce a few hundred words, and most of these will be highly frequent and common to most learners (Milton, 2009). Daller and Phelan (2007) demonstrate that raters can be quite inconsistent in applying these kinds of criteria. While judgements of vocabulary range appear to be one of the more reliably applied sets of criteria in this data, it appears that raters can be misled by non-vocabulary factors such as accent in making their judgements (Li, 2008).

The value of the CEFR lies in the ability of its users to apply these criteria consistently and accurately but in the absence of more detailed criteria this may be difficult to do in practice. This difficulty is implicitly recognised in the CEFR document with the suggestion that vocabulary size details might usefully be added to the descriptors. The potential value of a form of assessment which is able to put some numbers, or more precise measurements, to these characterisations is very clear. If a learner possesses many thousand words, including idiomatic and colloquial expressions, and is comparable to a native speaker in his or her foreign language vocabulary knowledge then this would be good evidence that he or she would be at C2 level, at least in terms of vocabulary range. A learner with only a few hundred foreign language words would probably be at A1 level in terms of vocabulary range and almost inevitably would be much more limited in their skill in using the foreign language. It is exactly the kind of development which the writers of the CEFR foresee and

Table 2. Vocabulary range criteria from Council of Europe (2001, p. 112)

VOCABULARY RANGE	
C2	<i>Has a very good command of a very broad lexical repertoire including idiomatic expressions and colloquialisms, shows awareness of connotative levels of meaning.</i>
C1	<i>Has a good command of a broad lexical repertoire allowing gaps to be readily overcome with circumlocutions; little obvious searching for expressions or avoidance strategies. Good command of idiomatic expressions and colloquialisms.</i>
B2	<i>Has a good range of vocabulary for matters connected to his or her field and most general topics. Can vary formulation to avoid repetition, but lexical gaps can still cause hesitation and circumlocution.</i>
B1	<i>Has a sufficient vocabulary to express him/herself with some circumlocutions on most topics pertinent to his/her everyday life such as family, hobbies and interests, work, travel and current events.</i>
A2	<i>Has sufficient vocabulary to conduct routine, everyday transactions involving familiar situations and topics. Has a sufficient vocabulary for the expression of basic communicative needs. Has a sufficient vocabulary for coping with simple survival needs.</i>
A1	<i>Has a basic vocabulary repertoire of isolated words and phrases related to particular concrete situations.</i>

which SLATE is embracing in its diagnostic tool kit. A set of figures for the size of vocabulary learners possess, and a straightforward method for determining these, would appear to be a very useful addition to the more general descriptive criteria on vocabulary range in particular.

3. Vocabulary knowledge and language skill

While the idea that the bigger and better your vocabulary in a foreign language is, the better you will be in your foreign language seems obvious, it is worth asking what research evidence we have to demonstrate that this is in fact the case. There is now a quite extensive body of research evidence which supports this idea and even provides some information as to the scale of vocabulary needed for different levels of performance and even which words are required to attain the highest levels in the CEFR framework.

The principle underlying these studies is an instrumentalist view of vocabulary that words are the primary carriers of meaning (Vermeer, 2001) and that as a consequence vocabulary provides the 'enabling knowledge' (Laufer & Nation, 1999) to be successful in other areas of language communication and proficiency. These studies show repeatedly that estimates of vocabulary knowledge correlate with reading comprehension (for example, Beglar & Hunt, 1999; Laufer, 1992; Qian, 1999; Stæhr, 2008), with writing ability (for example, Astika, 1993; Laufer, 1998; Stæhr, 2008), with listening comprehension (Milton, Wade, & Hopkins, 2010; Stæhr, 2008; Zimmerman, 2004), and with oral fluency (Milton et al., 2010; Zimmerman, 2004). The correlations are usually quite good and are typically between 0.6 and 0.8. Large vocabularies, there-

fore, typically associate with good performance in the communicative skills, and low vocabularies associate with poor performance. Perhaps not surprisingly therefore, the research also shows that tests of vocabulary knowledge can discriminate between groups of learners at different ability levels (Meara, 1992) and can be a useful tool for assigning learners to the correct level in an institutional program (Laufer, 1992; Laufer & Nation, 1999; Schmitt, 1994). The research also shows the presence of thresholds in vocabulary knowledge; volumes of vocabulary knowledge which appear essential if certain levels of comprehension, communication or performance in a foreign language are to be attained. Vocabulary thresholds have been suggested for reading comprehension (Alderson, 1984) and for reading and writing ability (Stæhr, 2008).

Recent research in this area goes beyond searching for correlations between vocabulary knowledge and the scores on individual languages skills and seeks to use regression analysis to calculate the scale of the contribution which vocabulary knowledge makes to performance in these skills. These studies establish that vocabulary knowledge, and vocabulary size in particular, is a major contributor to communicative performance in a foreign language. Stæhr (2008), for example, examines the relationship between examination grades on listening, reading and writing papers, and the vocabulary size of the testees, using scores on Nation's (1990, 2001) Vocabulary Levels Test as an indicator of vocabulary knowledge. His results suggest a link between vocabulary knowledge and all three elements of exam performance and a strong link with reading in particular. The correlations are given in Table 3.

Table 3. Spearman correlations between vocabulary size scores and listening, reading and writing scores (N=88) (Stæhr, 2008, p. 144)

	Listening	Reading	Writing
Vocabulary size	.69**	.83**	.73**

** Correlation is significant at the 0.01 level

Stæhr (2008) goes on to divide his exam results into two groups; below average, and average and above average. He carries out a binary logistic regression analysis using this division and concludes that as much as 72% of variance in the ability to score an average mark or above on the reading test can be explained by vocabulary size – the number of words a learner knows. Vocabulary may be less important than this in writing and listening but the contribution of vocabulary knowledge still appears sizeable. Stæhr's results suggest that up to 52% of variance in the ability to score average or above in writing, and 39% of variance in listening, can be explained through vocabulary knowledge.

Milton et al. (2010) investigate the contribution of two types of vocabulary knowledge, orthographic vocabulary size and phonological vocabulary size, to the scores and sub-scores on the International English Language Testing System (IELTS) test (see <http://www.ielts.org/default.aspx>). The IELTS test provides sub-scores for each of the four skills: reading, writing, listening and speaking. The orthographic vocabulary size test used is X_Lex (or XLex), the Swansea Levels Test (Meara & Milton, 2003). The phonological vocabulary size test used is Aural Lex, known as ALEX (Milton & Hopkins, 2005), which is an aural version of XLex where the tests words are heard but not seen. Details of this test are given later in this chapter.

The Spearman correlations which emerge from these scores are provided in Table 4 and reveal an interesting pattern to the way vocabulary knowledge interacts with scores on the IELTS skill sub-scores.

Table 4. Spearman correlations between vocabulary size scores and IELTS scores (N=30) (Milton et al., 2010, p. 91).

	ALex	reading	listening	writing	speaking	overall
XLex	.46*	.70**	.48**	0.76**	.35	.68**
ALex		.22	.67**	.44*	.71**	.55**

** Correlation is significant at the 0.01 level

* Correlation is significant at the 0.05 level

The modest correlation between ALEX and XLex suggests that two different aspects of vocabulary knowledge are being tested and that they are not strongly connected. Orthographic vocabulary (XLex) scores correlate well with reading and writing skills, while phonological vocabulary (ALex) scores correlate well with speaking scores. Both vocabulary scores correlate with listening scores perhaps because the test for this skill involves both reading and listening. It is the orthographic vocabulary (XLex) scores which correlate particularly well with overall IELTS scores and which therefore appear to link best with overall language performance. Linear regression analysis suggests that nearly 60% of variance in the IELTS writing scores, 48% of variance in reading scores, and 58% of variance in overall IELTS scores can be explained by differences in orthographic vocabulary size. 51% of variance in listening scores can be explained by a combination of orthographic and phonological vocabulary scores. Using a binary logistic regression, where learners are divided into groups scoring IELTS 5 or better or below 5, 61%, variance in speaking scores can be explained through differences in phonological vocabulary (ALex) scores.

But research also suggests which words might be particularly relevant in these frameworks. Stæhr (2008) suggests that knowledge of the most frequent 2000 words in English in particular represents a threshold which must be crossed if learners are to gain an average score or above on the language tests he uses, and from this he suggests that the most frequent 2000 words are essential for learners to progress to intermediate level and beyond, presumably the B and C levels in the CEFR, and this is supported by Nation (2001, p. 16). Nation's (2006) study of coverage in English and comprehension further suggests that knowledge of the most frequent 5000 words, and overall vocabulary knowledge of perhaps 8000 or 9000 words, is essential for the highest levels of fluency and understanding in English as a foreign language. Where vocabulary knowledge measures tie so closely to performance and skill in the foreign language, it might be expected that vocabulary knowledge will link to levels within the CEFR.

These are interesting results and very relevant to this chapter since they suggest a very strong association between a learner's vocabulary size, and in particular the number of words a learner recognises in written form, and the communicative level and performance that the learner attains. This lends weight to the idea that particular vocabulary sizes might be associated with the grades of the CEFR, and confirms the attention paid in the CEFR document to vocabulary range, and in particular vocabulary size, as measured by the tests used in Stæhr (2008) and Milton et al. (2010) in particular.

Even from a brief review of this kind two general truths emerge. One is, as the CEFR hierarchy suggests, that progress through the hierarchy is closely related to vocabulary knowledge and knowing more and more words in the foreign language. High level performers tend to have extensive vocabulary knowledge and elementary level performers do not. The second is that knowledge of the most frequent words in the foreign language appears crucial to successful performance.

4. Vocabulary knowledge

Attention is paid in the CEFR description to several aspects of vocabulary knowledge and the terms vocabulary range, vocabulary control and vocabulary size are all used. How do these terms fit into the terminology for vocabulary knowledge which is more commonly used by researchers in this area and are there tests for these qualities? A common characterisation in vocabulary studies is to consider vocabulary knowledge as a number of contrasting dimensions. On one dimension is vocabulary size, also called lexical breadth, which is 'the number of words a learner knows regardless of how well he or she knows them' (Daller, Milton, & Treffers-Daller, 2007, p. 7). A calculation of vocabulary size

hinges, therefore, on the number of words a learner can recognise as words in the foreign language, and it may not matter for this calculation whether a meaning or a translation can be attached to the word and whether the word can be used with any great subtlety. This is the type of knowledge which is widely used by researchers when searching for the connection between vocabulary knowledge and language skills such as reading and writing, as do Stæhr (2008), Nation (2006) and Milton et al. (2010) in the previous section, and is explicitly mentioned by the CEFR description as a potentially useful calculation (Council of Europe, 2001, p. 150). Much of the Vocabulary range criterion, with its characterisations of basic vocabulary and broad lexical repertoire appears to be a function of this size or breadth dimension. I would argue too that Vocabulary control, with its emphasis on the learner's ability to select the right word for the meaning intended, is also largely a function of vocabulary size.

Vocabulary size contrasts with other dimensions of vocabulary knowledge. It contrasts with the knowledge a learner may have about how these words may work, their nuances of meaning and subtleties of combination, which is known as vocabulary depth. Knowledge of vocabulary depth is often calculated by estimating the degree to which learners can appropriately combine words as in collocations (Gyllstad, 2007), or can be selected for their appropriateness in given situations as in the use of idioms and colloquialisms, but the concept may also include partial word knowledge and knowledge of polysemy (Read, 2000). The Vocabulary range criterion (Council of Europe, 2001, p. 112) includes elements of vocabulary depth in addition to vocabulary size or breadth by including, at C level, reference to idiomatic expressions, colloquialisms and connotative meaning. The vocabulary control criterion (Council of Europe, 2001, p. 112) also appears to include elements of vocabulary depth in its reference, again at C level, to the 'appropriate use of vocabulary'. There need be no ambiguity in the CEFR's characterisations here since vocabulary size test scores correlate well with vocabulary depth scores. The two qualities are very closely inter-related and a test of one dimension inevitably tests the other. Vermeer (2001) argues that depth is a function of vocabulary size and that effectively they are the same dimension.

Vocabulary size contrasts too with the ease and speed with which these words can be called to mind and used in communication, which is usually characterised as productive vocabulary knowledge or lexical fluency. If these three dimensions of vocabulary knowledge seem, superficially, easy to understand and potentially useful in characterising learner language, they have proved rather harder in practice to operationalise. Of the three dimensions only vocabulary size has a generally accepted definition which can give rise to standard tests capable of measuring knowledge in this field. Fortunately, a vocabulary size test would seem to capture most of what the CEFR terms Vocabulary range and Vocabulary control. However, it has taken some time for standard tests to

emerge, because measuring a foreign language learner's word knowledge requires decisions to be made about what should be counted as a word. Early counts of the number of words a person knows gave some very large figures; sometimes in the hundreds of thousands for native speakers (for example Seashore & Eckerson, 1940). Word learning on this scale appeared to set a super-human task for foreign language learners who aspired to native-like levels of competence and performance. But figures such as these are a product, at least in part, of the definition of what a word is. These counts tended to be made on the basis of dictionary counts where polysemous words or homonyms might have several entries. A word such as *bank* might, therefore, include several entries as a noun: the *bank* of a river or the *bank* which dispenses money. It might also include several entries as a verb: *to bank* as in the turn of an aircraft, to put money into a bank, or to rely on something. But it might also include separate and additional entries for the various derived and inflected forms of these words. *Bank* and its plural *banks* might, conceivably, have separate entries. So too might other related forms such as *banker* or *bankable*. There is a tendency in these, older, counts for every different form of a word to be counted as a different word, and for every different meaning of the same form to be counted differently.

A word count made in this way may not usefully characterise the nature or scale of learning that foreign language learners undertake. Inflected and derived forms of words, in English these include regular plurals made by adding *-s* and regular past tenses made by adding *-ed* for example, often appear to be rule based. Once a learner knows a rule, it can be applied to a huge number of other words in English without the need for separate learning of a new form. It may be more useful, therefore, to characterise a word as a base form, in the example above *bank*, and a variety of related forms: *banks*, *banking* and *banked* for example. A base form and its related forms are known as a word family and we have evidence that words are stored and used in this way by native speakers (Aitchison, 1987, Chapter 11, for example summarises this evidence). And we have evidence that words are learned in this way by foreign language learners. Schmitt and Meara's (1997) research among Japanese L2 learners of English suggests that inflectional suffixes in particular are learned comparatively early and that a base form and rule-based variants are a feature of the developing lexicon, even if inaccuracies in use persist and knowledge of derived forms are added much later in the course of learning. If the definition of a word family is broadly drawn to include a base form and all derivations and inflections then it is estimated that an educated native speaker might know some 17,000 to 20,000 words (D'Anna et al., 1991; Goulden et al., 1990). While this is still a huge volume of learning for a non-native speaker to aspire to, it is substantially more approachable in scale than the hundreds of thousands of words suggest-

ed by earlier counts. There are many reputable and useful word counts and word lists which have been compiled using a definition of word as a large word family (for example, Coxhead, 2000).

This is not the only way to define a word family, however, nor the most appropriate for the purpose of building vocabulary size counts into the levels of CEFR. As Gardner (2007, pp. 260–261) points out these constructs need not be absolutely rigid in their definitions but might usefully be varied, for example to match the levels of knowledge of the learners being examined. A feature of the large word family is that it includes many comparatively infrequent derivations in its definition of a word, and these derivations are rarely known by non-native speakers at least until they achieve advanced levels of performance (Schmitt & Meara, 1997). A word count based on this definition seems likely to mischaracterise the knowledge of lower level learners, therefore. A rather more useful definition of a word family, which addresses this issue, is the lemma. A lemma includes a base form of a word and only the most frequent and regular inflections; related forms of a word which do not differ in part of speech from the base form. In English the lemma would include regular plurals and genitives in nouns, regular inflections -s, -ed, -ing and -en past participle forms in verbs, and comparative and superlative -er and -est endings in adjectives. Research evidence suggests these tend to be acquired early in the process of learning and this definition matches the kind of knowledge which elementary and intermediate level learners have. It is not surprising, therefore, that vocabulary size tests often draw on word frequency information where the words being counted are lemmatised words, for example, Nation's (1990, 2001) Vocabulary Levels Test and Meara and Milton's (2003) XLex. Nation's test has been described by Meara (1996) as the nearest thing we have to a standard test in the area, but the vocabulary size tests Meara has subsequently developed, such as XLex, appear to offer additional qualities to the tester. It seems likely, then, that these tests might be most useful for the basis of the size estimates that the creators of the CEFR feel would add useful detail to the levels descriptors.

5. Measuring vocabulary size

The previous sections have reported the use of vocabulary size tests which provide useful characterisations of the foreign language vocabulary size of learners. What are these measures and how do they work? Nation's (1990, 2001) Vocabulary Levels Test is one widely used test. It tests words in the 2000, 3000, 5000 and 10000 word frequency ranges, in addition to words from the University Word List (Nation, 1990) in order to estimate overall lexical competence. Each level tests 30 test words in a multi-item matching task where testees

are provided with six test words and a selection of three explanations which must be matched up with the test words. An example is given in Figure 1.

Figure 1. Vocabulary Levels Test example taken from Nation (2001, p. 416)

This is a vocabulary test. You must choose the right word to go with each meaning. Write the number of that word next to its meaning.	
1 business	
2 clock	_____ part of a house
3 horse	_____ animal with four legs
4 pencil	_____ something used for writing
5 shoe	
6 wall	

This format allows rather more than passive recognition of word forms to be tested and should allow an estimate of knowledge of words and their meanings to be formed. It is quite a complex test, however, where success relies not just on a learner's knowledge of the test words (on the left hand side) but also on knowledge of the words in the explanations (on the right hand side) and it is not completely clear which items are being tested. Further, each question contains multiple test items and the learner's knowledge of some of the items is likely to have an impact on the ability to work out the answers to other items where these are not known. We know that learners often try to maximise their scores by making educated guesses in these circumstances, it is called economy of practice, but we do not know the effects of guesswork on the scores that emerge. Kamimoto (2005), for example, reports think aloud protocols conducted among learners taking this test and the feedback he received suggests that a considerable amount of guesswork and calculation goes on in answering the questions and that the learner's choice of guessing strategy can produce considerable differences in score. The Levels Test might have much more variation according to guesswork than most users ever imagine. However, there is no explicit way in the test for taking account of this phenomenon or compensating for it. It is not completely clear how scores on this test might equate to vocabulary size although it would be a surprise if scores on this test and vocabulary size did not correlate well. It is also heavily weighted in its content to the infrequent ranges of vocabulary. It will be recalled that Stæhr (2008) draws attention to the importance of the most frequent 2000 words for learners but the most frequent 1000 words, the frequency band which includes so much structure and function vocabulary, is absent from consideration in this test. The majority of test items lie at or beyond the 5000 word level and learners will need to be highly

accomplished in English before they have significant knowledge in this area. A test in this format is probably unsuitable as a measure to tie into the levels of the CEFR since it seems unlikely that it will be able to distinguish well between learners at the elementary levels of knowledge and performance.

The better alternative is the XLex check list test format (Meara & Milton, 2003) which corresponds to tests of passive receptive vocabulary within a specified vocabulary range. The XLex tests estimate knowledge of the most frequent 5000 lemmatised words in a language and estimate overall knowledge of this vocabulary. They are Yes/No tests which present learners with test words, one by one, and learners have to indicate whether they know each word. There is a combination of real words and false words which are designed to look and sound like words in the target language. The number of Yes responses to the real words allows a preliminary estimate of the proportion of words known from the most frequent 5000 words, and the number of Yes responses to the false words allows this preliminary estimate to be adjusted for guessing and overestimation of knowledge. The tests give an overall score of words known out of the most frequent 5000 words. These tests can draw on frequency lists in different languages, where these exist, to allow comparable tests of vocabulary to be constructed. In Meara and Milton's (2003) XLex, for example, the English tests are based on data available from Nation (1984) and Hindmarsh (1980), Greek tests draw on the Hellenic National Corpus (<http://hnc.ilsp.gr/en/>; see also Milton & Alexiou, 2010), and the French tests draw on Baudot's (1992) frequency lists. The 5000 words content includes the most frequent 2000 words, the importance of which elementary and intermediate learners Stæhr (2008) has pointed to, and the most frequent 5000 words which Nation (2006) has pointed to as so crucial for advanced level comprehension and communicability. Studies on these tests in English format (Milton, 2005), in French format (David, 2008), and in Greek format (Milton & Alexiou, 2010) conclude that they are both reliable and valid measures of receptive vocabulary size. An illustration of the format of this type of test is given in Figure 2.

Figure 2. Example of a checklist test (French version from Milton, 2009, p. 257)

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

chien

Thank you for your help.

de distance abattre absurde achevé mancher

This format has certain features which make it attractive for calculating the vocabulary size that learners have in relation to CEFR levels. The test is simple in format and comparatively large numbers of items can be tested in a short space of time. Results on these tests are usually very reliable. The test format also has the benefit of allowing the creation of parallel or even equivalent forms of the test in different languages to be made rather more straightforwardly than would be the case with multiple choice or Levels Test formats. Examples of the tests can be found in Milton (2009, pp. 255–261).

6. Scores on vocabulary size measures and CEFR levels

The research literature contains several examples where vocabulary size estimates have been linked explicitly to the levels of the CEFR. Milton and Meara (2003) tested students taking and passing Cambridge exams at every level of the CEFR and estimated their vocabulary sizes using the XLex tests. These exams test all the 4 skills and vocabulary size is therefore linked to a learners' level in the widest possible sense. Their results are shown in Table 5.

Table 5. Approximate vocabulary size scores associated with CEFR levels (adapted from Meara & Milton, 2003, p. 8)

CEFR Levels	Cambridge exams	XLex (5000 max)
A1	Starters, Movers and Flyers	<1500
A2	Kernel English Test	1500 - 2500
B1	Preliminary English Test	2750 - 3250
B2	First Certificate in English	3250 - 3750
C1	Cambridge Advanced English	3750 - 4500
C2	Cambridge Proficiency in English	4500 - 5000

Milton and Alexiou (2009) used three different language versions of XLex and collected data from over 500 learners of English, Greek and French as second and foreign languages with a view to comparing the levels of vocabulary knowledge at each CEFR level. As broad a range of languages and learners as possible was sought and the authors drew on contacts from Vocabulary Research Group in order to collect the data from a variety of locations. The EFL learners in Hungary were drawn from 144 students in two schools in Szeged and the 88 EFL learners in Greece from a private language school in central Greece. The learners of French in UK were 155 students at a comprehensive school in South

Wales, and the learners of French in Spain were all 50 students of French at a Secondary school in northern Spain. The Greek learners of French were all 65 students of French at a private language school in central Greece. The learners of Greek as a second language comprised all 64 learners at a university centre in Thessaloniki. The CEFR levels were determined in this study by teachers who placed the learners being tested into streams for study at each of the CEFR levels. The mean scores and standard deviations collected from learners at each level are summarised in Table 6. The standard deviation scores are included to indicate the degree of overlap between the groups at each level.

Table 6. Summary of mean scores (and standard deviations) for each CEFR level in three foreign languages

CEFR Level	EFL in Hungary	EFL in Greece	French in Spain	French in Greece	Greek in Greece
A1		1477 (580)	894 (604)	1125 (620)	1492 (705)
A2		2156 (664)	1700 (841)	1756 (398)	2237 (538)
B1	3135 (434)	3263 (434)	2194 (717)	2422 (517)	3338 (701)
B2	3668 (666)	3304 (666)	2450*	2630 (251)	4012 (415)
C1	4340 (471)	3690 (471)	2675 (643)	3212 (473)	
C2		4068 (261)	3721 (416)	3525 (883)	

* in this field there is only one student score and no SD can be calculated

These figures are broad generalisations but one striking feature emerges. Progressively higher vocabulary scores are associated with progressively higher levels in the CEFR hierarchy. Milton and Alexiou (2009, pp. 200–204) use ANOVA and Tukey analyses to demonstrate that the differences between the mean scores at each CEFR level are statistically significant, and linear regression modelling to show that, with the exception of the Hungarian data, some 60 to 70% of variance in CEFR levels can be explained by differences in vocabulary size. It is apparent from this data that the assumption made in the CEFR literature, that as learners progress through the CEFR levels their foreign language lexicons will increase in size and complexity, is broadly true. There is individual variation and overlap between the scores that learners attain within the CEFR levels, however. It will be recalled that vocabulary size and communicative performance are separable qualities and one interpretation of this variation is that students with the same or similar vocabulary sizes may make different use of this knowledge to communicate more or less successfully.

A further notable feature of these estimates is just how much vocabulary is needed to progress beyond the most basic levels of performance. Something like 2000 words, of the most frequent 5000, might be needed to attain A2 level in all the languages tests, and this recalls Stæhr's (2008) comments, although about EFL in particular, about the importance of the most frequent 2000 words and the threshold that knowledge of these words represents in attaining above average marks in his writing and reading tests. It appears that approximately 3000 words, of the most frequent 5000 words, might be needed to progress beyond the elementary A1 and A2 levels and achieve any kind of independence in English and Greek as foreign languages. The advanced levels of performance on the CEFR, C1 and C2 level, are associated as Nation (2006) suggested with almost complete recognition of the most frequent 5000 words. Scores of 4000 or better are associated with these levels and these suggest that learners' overall vocabularies must be very large indeed, and this fits well with Nation's (2006) figures, derived from coverage calculations, that 8000 or 9000 words overall might be needed for full comprehension of written texts.

However, the figures associated with each level in the three different languages are different. The figures for French as a foreign language are consistent with each other and suggest that at each CEFR level the learners have smaller vocabulary sizes than would be the case for English. The data available for Greek as a foreign language suggests that achieving each CEFR level requires a rather larger vocabulary size than would be the case for either English or French. Milton and Alexiou (2009) are at pains to point out that the data available come from comparatively small samples and that caution should be exercised in reaching conclusions as a consequence. Nonetheless, there are reasons why differences of this kind might be expected when comparing the vocabulary sizes of learners across different languages. This raises the question of how vocabulary sizes can be meaningfully compared in this way.

7. Making cross-linguistic comparisons

One of the virtues of a checklist test format, where words are chosen from across frequency bands, is that it is possible to construct tests in different languages which are arguably comparable with each other. The tests used in the previous section can all claim to estimate the number of words the learners know out of the most frequent 5000 lemmatised words in those languages. It is not inevitable, however, that knowing the same number of words in several languages will mean you can perform identically in those languages. Languages differ in important ways and one effect of this may be that it is possible to do more with fewer words in one language than in another.

There are several reasons for thinking this which are discussed in Milton (2009). One is that languages can inflect and derive words very differently and this may affect frequency calculations. The most frequent words in English, for example, include pronouns and prepositions but not every language shares this quality. Agglutinative languages, like Hungarian, Finnish and Turkish, handle many of the functions and meanings which these words convey in English rather differently. Typically these meanings are conveyed by the addition of suffixes to the root form of a verb or noun with the result that a single word family might include many more word forms than would be the case in English. This should affect the volumes of vocabulary required for coverage and comprehension. It may be possible to do more with, say, 1000 words in Finnish than in English. A second reason might lie in the historical development of languages. Speakers of English, for example, often appear to have a variety of words available for much the same idea in a way that is quite different from other languages. For historical reasons English differentiates, for example, between many farmyard animals and the meat which comes from them, between *pork* and *pig* and between *sheep* and *mutton*, in a way that other languages do not. This implies that comprehension of English might require a larger vocabulary than would be needed for equivalent understanding in another language. Further, languages can differ considerably in their word formation processes. Some languages, like German, favour compounding and combine existing words to make a new word. By contrast, other languages favour creating or deriving words for new concepts and ideas. The effect of this word combining process may be compounded since not all languages are as clear as Western European languages in signalling where word boundaries occur. Chinese, for example, which combines ideographs to make words, does not mark word boundaries in writing so it may not immediately be clear where one word ends and the next begins, or whether an expression should be treated as two or three words or as one. Different decisions would, when systematised across a whole corpus, produce different word counts.

A further difference may result from cultural differences in how languages choose to express themselves across different registers. We have some empirical evidence from English and French that these differences can affect coverage. It appears that in French the most frequent vocabulary does the service of everyday language and, in addition, can also be used in formal and academic registers where, in English, a specialist academic vocabulary is required. Cobb and Horst (2004, p. 30) point to the coverage provided of academic texts by the most frequent 2000 words in French. The figure they quote of nearly 89% would be equivalent to the General Service Word List of 2000 words plus Coxhead's Academic Word List, a further 570 carefully selected rather than purely frequency-based words in English. It appears that it really may be possible to do more communicatively with fewer words in French than is the case in English, as the figures in Table 6 suggest.

This observation leads Milton and Alexiou (2009) to suggest that the figures for vocabulary size in the CEFR levels may be tied to coverage, and that this will provide a way of both explaining and predicting how vocabulary size will vary across the CEFR levels in different languages. They note (2009, pp. 207–208) that the figures for coverage in Greek suggest that larger vocabularies will be needed to achieve the CEFR's various communicative requirements at each level, and that the data from Table 6 confirms that this is the case. Milton and Alexiou (2008) also suggest, however, that this is not the only possible interpretation of this data and that these differences might equally well be the result of differences in the ways the various corpora used to derive information on coverage were created. Generally, language corpora aspire to draw on a wide variety of general source material such as newspapers, novels and periodicals so as to be as representative of the normal use of the language as possible. But there are no conventions to indicate how large a sample from these genres is appropriate for the corpus to be representative of the language as a whole. A corpus compiled from a large number of small samples drawn from a wide variety of these sources may well contain a greater diversity of words than corpora drawn from a smaller number of texts where the material is likely to be thematically much more unified. This may explain why the Greek corpus (<http://hnc.ilsp.gr/en/>), at least in the incarnation used for the Greek XLex tests, appeared to contain a much larger proportion of singly occurring words than the other two corpora. The English language corpora used to create the EFL XLex tests (Hindmarsh, 1980; Nation, 1984) include material drawn from EFL teaching texts, a feature which is absent from both the French corpus (Baudot, 1992) and Greek corpus used in this study. This may have influenced the number of words that learners recognise, since the English XLex tests potentially contain words more closely related to the material they have encountered in class although it is not clear that this will inevitably be the case. None of these corpora, at the time they were used, contained a spoken sub-corpus and while this is unlikely to have affected the words which emerge as most frequent, it is likely to affect the comparative frequencies of these words and, potentially, the frequency bands they occur in therefore. Language corpora, constructed in a much more strictly equivalent manner, are needed before doubts about the comparability of corpora, and the tests derived from them, can be allayed.

8. Conclusions

The evidence from studies of vocabulary size has confirmed that vocabulary size measurements can be tied to the levels of the CEFR with some confidence, although it seems at present as though the actual vocabulary sizes may depend

on the language being tested and perhaps the source of the words which form the test. The higher up the hierarchy of the CEFR learners progress, the more words they are likely to require and the greater vocabulary size they will have. The relationship between vocabulary size and the CEFR levels is sufficiently strong, notwithstanding some individual variation, for figures for vocabulary size to be attached to the CEFR level. This relationship is not just a matter of academic interest and one of the most useful benefits in linking vocabulary knowledge and the CEFR in this way is to add detail to the framework in the manner that the CEFR itself anticipates. Users of the system often find it difficult to match learners or materials to the levels with any precision and different people, different examiners, even different national examination systems, can apply the CEFR's levels descriptors very differently. The interest SLATE has in creating a diagnostic tool kit which links linguistic features of performance to the CEFR levels looks to be important, therefore, and the use of vocabulary size measurements, and the tests to derive such measurements, appear to add detail to the CEFR rather as the creators of the framework anticipate. The presence of these kinds of more precise descriptors should help users of the system in different schools or countries apply grading criteria more consistently and confidently. Vocabulary size measurements have much to add to the CEFR hierarchy therefore. It seems that vocabulary size can be a very useful part of a linguistic tool kit and even by itself is a good predictor of CEFR level (Milton & Alexiou, 2009).

References

- Aims of SLATE. (n.d.). Retrieved from <http://www.slate.eu.org/aims.htm>
- Aitchison, J. (1987). *Words in the mind*. Oxford: Blackwell.
- Alderson, J. C. (1984). Reading in a foreign language: A reading or a language problem? In J. C. Alderson & A. H. Urquhart (Eds.), *Reading in a foreign language* (pp. 1–24). London: Longman.
- Astika, G. G. (1993). Analytical assessment of foreign students' writing. *RELC Journal*, 24(1), 61–72.
- Baudot, J. (1992). *Fréquences d'utilisation des mots en français écrit contemporain*. Montréal: Les Presses de l'université de Montréal.
- Beglar, D., & Hunt, A. (1999). Revising and validating the 2,000 word level and university word level vocabulary tests. *Language Testing*, 16, 131–162.
- Cobb, T., & Horst, M. (2004). Is there room for an AWL in French? In B. Laufer & P. Bogaards (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 15–38). Amsterdam: John Benjamins.
- Coste, D., Courtillon, J., Ferenczi, V., Martins-Baltar, M., & Papo, E. (1987). *Un niveau seuil*. Paris: Didier.

- Council of Europe. (2001). *Common European Framework of Reference for Languages*. Cambridge: Cambridge University Press.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34, 213–238.
- Daller, H., Milton, J., & Treffers-Daller, J. (Eds.). (2007). *Modelling and assessing vocabulary knowledge*. Cambridge: Cambridge University Press.
- Daller, H., & Phelan, D. (2007). What is in a teacher's mind? Teacher ratings of EFL essays and different aspects of lexical richness. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 234–244). Cambridge: Cambridge University Press.
- D'Anna, C. A., Zechmeister, E. B., & Hall, J. W. (1991). Toward a meaningful definition of vocabulary size. *Journal of Reading Behavior*, 23, 109–122.
- David, A. (2008). Vocabulary breadth in French L2 learners. *Language Learning Journal*, 36, 167–180.
- Gardner, D. (2007). Validating the construct of word in applied corpus-based vocabulary research: A critical survey. *Applied Linguistics*, 28, 241–265.
- Goulden, R., Nation, I. S. P., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics*, 11, 341–363.
- Gyllstad, H. (2007). *Testing English collocations – developing receptive tests for use with advanced Swedish learners* (Doctoral dissertation). Lund University, Sweden.
- Hindmarsh, R. (1980). *Cambridge English lexicon*. Cambridge: Cambridge University Press.
- Kamimoto, T. (2005, September). *The effect of guessing on vocabulary test scores: A qualitative analysis*. Paper presented at The European Second Language Association 2005 conference (EUROSLA 15), Dubrovnik, Croatia.
- Laufer, B. (1992). How much lexis is necessary for reading comprehension? In H. Bejoint & P. Arnaud (Eds.), *Vocabulary and applied linguistics* (pp. 126–132). London: Macmillan.
- Laufer, B. (1998). The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics*, 19, 255–271.
- Laufer, B., & Nation, P. (1999). A productive-size test of controlled productive ability. *Language Testing*, 16, 33–51.
- Li, H. (2008). *Measuring and assessing English L2 spoken vocabulary* (Unpublished doctoral dissertation). Swansea University, UK.
- Meara, P. (1992). *EFL vocabulary tests*. University College Swansea, Centre for Applied Language Studies.
- Meara, P. (1996). The dimension of lexical competence. In G. Brown, K. Malmkjaer, & J. Williams (Eds.), *Performance and competence in second language acquisition* (pp. 35–53). Cambridge: Cambridge University Press.
- Meara, P., & Milton, J. (2003). *X_Lex, the Swansea Levels Test*. Newbury: Express.
- Milton, J. (2005). Vocabulary size testing in the Arabic-speaking world. In P. Davison, C. Coombe, & W. Jones (Eds.), *Assessment in the Arab world* (pp. 337–353). Dubai: TESOL Arabia.

- Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters.
- Milton, J., & Alexiou, T. (2008). Vocabulary size in Greek as a foreign language and the Common European Framework of Reference for Languages. *Journal of Applied Linguistics*, 24, 35–52.
- Milton, J., & Alexiou, T. (2009). Vocabulary size and the Common European Framework of Reference for Languages. In B. Richards et al. (Eds.), *Vocabulary studies in first and second language acquisition* (pp. 194–211). Basingstoke: Palgrave.
- Milton, J., & Alexiou, T. (2010). Developing a vocabulary size test for Greek as a foreign language. In A. Psaltou-Joycey & M. Mattheoudakis (Eds.), *Selection of papers for the 14th International Conference of Applied Linguistics (GALA)* (pp. 307–318). Aristotle University of Thessaloniki, GALA.
- Milton, J., & Hopkins, N. (2005). *Aural Lex*. Swansea University.
- Milton, J., Wade, J., & Hopkins, N. (2010). Aural word recognition and oral competence in a foreign language. In R. Chacón-Beltrán, C. Abello-Contesse, M. Torreblanca-López, & M. D. López-Jiménez (Eds.), *Further insights into non-native vocabulary teaching and learning* (pp. 83–97). Bristol: Multilingual Matters.
- Nation, I. S. P. (Ed.). (1984). *Vocabulary lists: Words, affixes and stems*. English University of Wellington, English Language Institute.
- Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston: Heinle & Heinle.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, 63, 59–82.
- Qian, D. D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *The Canadian Modern Language Review*, 56, 283–307.
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Schmitt, N. (1994). Vocabulary testing: Questions for test development with six examples of tests of vocabulary size and depth. *Thai TESOL Bulletin*, 6, 9–16.
- Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework – word associations and verbal suffixes. *Studies in Second Language Acquisition*, 19, 17–36.
- Seashore, R. H., & Eckerson, L. D. (1940). The measurement of individual differences in general English vocabulary. *Journal of Educational Psychology*, 31, 14–38.
- Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal*, 36, 139–152.
- Van Ek, J. A. (1980). *Waystage English*. London: Pergamon Press.
- Van Ek, J. A. & Trim, J. L. M. (1991). *Threshold 1990*. Strasbourg: Council of Europe.
- Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics*, 22, 217–234.
- Zimmerman, K. J. (2004). *The role of vocabulary size in assessing second language proficiency* (Unpublished master's thesis). Brigham Young University, USA.

