KEY CONCEPTS IN APPLIED LINGUISTICS / CONCEPTOS CLAVE DE LA LINGÜÍSTICA APLICADA

THE SECOND LANGUAGE LEXICON

James Milton

Swansea University, UK

J.L.Milton@swansea.ac.uk

Summary articles which examine our understanding of the way words are learned to grow a second language lexicon usually begin by reflecting that academic interest in vocabulary has grown enormously in the last couple of decades after a period where it was largely disregarded. The lexicon has now found such a prominent place in applied linguistic and SLA research that Long and Richards (2009, xii) suggest that 'vocabulary can be viewed as the core component of all the language skills'. In suggesting this they are referring to two features which unite the diversity of work in the field of the second language lexicon. One is an appreciation of the centrality of the lexicon to language and to language learning. A lexicon of an appropriate size, and comprising appropriate elements, is essential to performance, something that is often overlooked in structural and other recent approaches to language teaching. The second is an understanding that the lexicon cannot be fully understood by approaches which view it as isolated words linked to separate and separable mean-

ings. The idea of the lexicon these days is that it is inextricably linked with other aspects of language such as grammar and syntax. The growth of a lexicon of appropriate size and quality may even drive the acquisition of other features of language. Wolter (2013) suggests that when language description is approached from a lexical perspective, the division between lexis and grammar breaks down.

Researchers in the second language lexicon approach the lexicon in a variety of different ways and for a variety of different purposes. Schmitt and McCarthy (1997) classify this research within three broad areas in order to make sense of this variety. These three areas are, firstly, research which focuses on description, secondly, research into the way elements of the lexicon are stored and accessed, and finally, there is research which focuses on the teaching and learning of the lexicon. Using these three headings is a convenient way of understanding the key ideas in our current understanding of the second language lexicon.

Description: what is a word?

It has been a goal of research into language description over many years, to arrive at workable and useful definitions of the units which makeup the lexicon. It seems that a learner's knowledge of the lexicon is not simply a list of words as they appear in a dictionary. A lexicon which is large and sophisticated enough for efficient second language communication is likely to include knowledge not just words but also of word parts and word combinations. As long ago as 1921 Palmer pointed out that learners of English will have to know word suffixes such as *–er* and *–ist* which can change a verb into an agent noun, so a *baker* is someone who *bakes* and *cyclist* is someone who *cycles*. They will have also to know common phases such as *of course* and *in spite of* which function like single words despite the fact they are made up of several words. It is possible to add further components to a definition of a unit of vocabulary knowledge.

It has become a commonplace to list the components which comprise word knowledge and these have grown over the years. Cronbach (1942) listed five components, Richards (1976) listed eight, and the most recent (Nation, 2001) lists nine elements each of which can be sub-divided into receptive and productive giving eighteen in all. This introduces such a degree of complexity into the testing of vocabulary knowledge that researchers rarely investigate all of these altogether but focus on one area of knowledge at a time or, for convenience, combinations that can be conveniently grouped together. A particularly prominent strand of research, therefore investigates *vocabulary breadth*, the number of words learners know or can recognise regardless of how much else they know about them. *Vocabulary breadth* is usually contrasted with a dimension called *vocabulary depth* (Anderson and Freebody, 1981), which is used to include what learners know about these words such as their collocational ability, their associations and the restrictions on their use. Some researchers introduce a further dimension of *fluency* or *automaticity* (Daller et al., 2007), which is the ease and speed with which second language vocabulary items can be used or recognised.

Storage, Access and Processing

These ideas about the multi-faceted nature of word knowledge have given rise to theories about how words are stored in the lexicon. There is growing agreement that learners store a base form of a word associated with a concept. Pienemann (1998), for example, places an understanding of the *lemma* at the heart of all language learning. It is the crucial insight from which all other understanding of the systems of language is derived. In using language, a base form of a word is retrieved from memory and can then be changed by rules for inflection or derivation to generate a different form of the word appropriate to the meaning and structure which is wanted. So, if a learner wants to talk about *cats* then *cat* is retrieved from memory, an *–s* is added to make the plural *cats*.

The different forms of a base word generated in this way are known as a *word family*. Learners do not need to learn all forms of the word family separately. If the rules are regular and frequent enough then learners need only learn one form to be able to generate all the others. The burden of learning is reduced in this way to manageable proportions. In English is seems that learners learn a form of word family called a *lemma*, which covers a base form and only the most frequent inflections which do not change the part of speech of a word. Many derived forms of English words are so irregular and infrequent, it seems, that they are learned as separate words. Other languages are much more regular in their word formation processes, Arabic for example, and larger word families can be learned. This is an important insight since by counting the lemmas or word families a learner knows, rather than a huge number of separate word forms, believable and understandable measurements of *vocabulary knowledge* can be gained.

Teaching, Learning and Assessment

Using a definition of a word as a word family or lemma then modern research is beginning to arrive at a consensus as to the scale of learning required to achieve proficiency in a language. In English as a foreign language it is thought that a minimum of 2,000 words, including the most frequent words, is absolutely essential before even gist understanding of normal language can be achieved. Nation (2001) suggests there is a threshold for language understanding around this level of vocabulary knowledge below which any sort of comprehension is nearly impossible. Learners have to engage in a substantial amount of learning therefore before they can realistically be expected to use the language meaningfully. Complete fluency might require substantially more; 8,000 or 9,000 words for trouble free reading and speech (Nation, 2006). It seems likely that other European languages will require vocabulary knowledge of the same order although work in this is still on-going (for example, Milton, 2010).

There are clear frequency effects in learning (Milton, 2007), that is to say, there is a tendency for the most frequent words in a language to be learned earliest. This links the learning of vocabulary to other more general theories of language learning. In social interactionist theory and competition theory (MacWhinney & Bates, 1989), for example, extensive language exposure is a requirement of learning and this will provide the frequency effects in input which are thought to govern sequences of uptake. Recent versions of generative grammar, such as the Minimalist Program, also view language uptake as being driven by frequency of occurrence in language input. Frequency of input significantly influences which words are learned in a language, and the quality of learning and the acquisition of these items in sufficient quantity triggers the setting of universal grammatical parameters. We have a less clear picture of the way the other dimensions of vocabulary knowledge develop except that depth and fluency very often develop in line with vocabulary breadth.

It seems, therefore, that we have developed a clear idea of what is being learned when a learner grows a lexicon in a second language. We have an idea, especially in English, of both the quantity of words needed, but also the lexicon's quality, which words are needed and are learned. Using standard tests to measure vocabulary knowledge (for example Meara and Milton's X-Lex 2003) a learner's knowledge and progress can be readily measured and it seems that vocabulary size works as a good predictor of performance in all of the language skills. Perhaps it should not be a surprise, therefore, that vocabulary size provides a very good guide to a learner's overall level of language knowledge and performance. Measuring the size of the second language lexicon produces a metric which might be more widely used by both teachers and learners to monitor their learning and their progress.

References

Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77-117). Newmark: International Reading Association.

Cronbach, L. J. (1942). An analysis of techniques for diagnostic vocabulary testing. *Journal of Educational Research, 36,* 206-217.

Daller, H., Milton, J., & Treffers-Daller, J. (2007). Editors' introduction: Conventions, terminology and an overview of the book. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 1-32). Cambridge: Cambridge University Press.

Long, M., & Richards, J. (2007). Series editors' preface. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. xii-xiii). Cambridge: Cambridge University Press.

MacWhinney, B., & Bates, E. (1989). *The crosslinguistic study of sentence processing*. New York: Cambridge University Press.

Milton, J. (2010). The development of vocabulary breadth across the CEFR levels. In I. Vedder, I. Bartning, & M. Martin (Eds.), *Communicative proficiency and linguistic development: Intersections between SLA and language testing research* (pp. 211-232). Rome: Second Language Acquisition and Testing in Europe Monograph Series 1.

Milton, J. (2007). Lexical profiles, learning styles and construct validity of lexical size tests. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 45-58). Cambridge: Cambridge University Press.

Nation, I.S.P. (2001). *Learning vocabulary in another language.* Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139524759

Nation, I.S.P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review, 63*(1), 59-82. http://dx.doi.org/10.3138/cmlr.63.1.59

Pienemann, M. (1998). *Language processing and second language development: Processability theory*. Amsterdam: John Benjamins.

Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quarterly, 10,* 77-89. http://dx.doi.org/10.2307/3585941

Schmitt, N., & McCarthy, M. (1997). *Vocabulary: Description, acquisition and pedagogy.* Cambridge: Cambridge University Press.

Wolter, B. (2013). Lexis: Overview. In C. A. Chapelle (Ed.), *The encyclopedia of Applied Linguistics.* (pp.3391-3394) London: Blackwell.

James Milton is Professor of Applied Linguistics at Swansea University UK. A long-term interest in measuring lexical breadth and establishing normative data for learning and progress has led to extensive publications including *Modelling and Assessing Vocabulary Knowledge* (CUP, 2007 with Michael Daller and Jeanine Treffers-Daller), *Measuring Second Language Vocabulary Acquisition* (Multilingual Matters, 2009) and *Dimensions of Vocabulary Knowledge* (Palgrave, 2014 with Tess Fitzpatrick).