

V_LINKS: BEYOND VOCABULARY DEPTH

PAUL MEARA AND BRENT WOLTER

ABSTRACT

This paper argues that the distinction made by many people between vocabulary breadth and vocabulary depth is an unfortunate one. Although the dichotomy is appealing, it forces us to look at vocabulary development in an unhelpful way. Depth of vocabulary knowledge can only be assessed by means of more and more detailed tests and the logistics of testing implies that this work can be done only with fewer and fewer words. The paper argues that **vocabulary size** and **vocabulary organisation** is a much more productive way of looking at vocabularies. It outlines some of our work on vocabulary organisation, and reports some preliminary results with a tool designed to assess the way core vocabularies are organised in L2 speakers.

Introduction

Recent work on vocabulary acquisition has tended to make a broad distinction between **vocabulary breadth** and **vocabulary depth**. Vocabulary breadth has generally been interpreted as the number of words that learners know, whereas vocabulary depth is generally taken to mean how well they know these words. Most of the research in this framework goes back to a seminal article by Richards published in 1976, though the ideas have been picked up and developed by other writers since that time (e.g. Nation 1990, Nation 2001, af Trampe 1983; Blum-Kulka 1981; Madden 1980, McNeill 1996, and others). Richards' paper identifies a number of different aspects of word knowledge and the most important of these are summarised in Figure 1.

- Knowing a word means knowing the degree of probability of encountering a word in speech or print. For many words we also know the sort of words most likely to be found associated with the word.
- Knowing a word implies knowing the limitations imposed on the use of the word according to variations of function and situation.
- Knowing a word means knowing the syntactic behaviour associated with a word.
- Knowing a word entails knowledge of the underlying form of the word and the derivatives that can be made from it.
- Knowing a word entails knowledge of the network of associations between the word and the other words in the language.
- Knowing a word means knowing the semantic value of word.
- Knowing a word means knowing many of the different meanings associated with the word. (P 83).

Figure 1. Aspects of word knowledge from Richards (1976)

A number of people have tried to develop formal tests which measure depth of vocabulary knowledge in these terms. Wesche and Paribakht (1996) for instance developed a rating scale approach, in which test-takers are invited to rate their knowledge of target words on a five point scale, generating definitions for the target words or sentences containing the target words to confirm their self-ratings where appropriate. Although VKS was initially developed as a way of measuring specific gains in vocabulary as a result of reading, it has often been taken as a more general test for measuring depth of vocabulary knowledge. A further example of this approach is Schmitt and Meara (1997). This paper developed an instrument which assessed test-takers' ability to generate derivative forms of target words, attempting to show that this ability was independent of vocabulary breadth. Other examples of vocabulary depth tests, which adopt the same general approach include Schmitt (1994) and Read (1995).

This work clearly takes the idea of 'knowing a word' some way further than the measures of vocabulary breadth which are currently available. These latter measures tend to be relatively superficial: Meara's Yes/No tests, for example (Meara and Milton 2003) simply ask test-takers to say whether they can recognise that a word exists or not, and Nation's Vocabulary Levels Test (Nation 2001; Schmitt, Schmitt and Clapham 2001) requires test-takers merely to match words to simple definitions. The depth tests, in contrast, require test-takers to show that their knowledge of the target words is not limited to superficial knowledge of this sort.

It seems to us, however, that this enterprise is fundamentally doomed. The problem is that testing vocabulary depth in this way requires us to carry out extensive testing of individual words, and this makes it all but impossible to design experiments which can tell us very much about the larger characteristics of whole vocabularies – a classic example of not being able to see the wood for looking at the trees. The logic of testing vocabulary depth using the vocabulary knowledge framework implies that we need to test very many words in ever-increasing detail, and this very quickly leads us into serious logistical problems which constrain the types of hypotheses that we can test. Suppose, for example, that we take Richards' list at face value, and suppose that we want to test how well a group of L2 speakers knows a list of 50 target words. To carry out this work, we would need to develop a set of perhaps a dozen subtests for each of the words we are interested in – at least one subtest for each feature in the framework. If we want to test 50 words in this way, then this implies that we would need a minimum of 600 test items before we can make even basic statements about a student's depth of vocabulary knowledge for these words. And this in

turn assumes that we could develop a single test item able to assess depth of knowledge in a meaningful way. On purely logistic grounds, a test battery of this size is completely infeasible: in practical terms, it would be very difficult indeed to get large groups of learners to take a 600 item test. In any case, it is highly unlikely that we could develop single test items that would reliably access a learner's depth of knowledge for target words – it is very difficult to think of any way of testing how well a learner knows the syntactic behaviour of a word using a single test item, for instance – and this implies that we would actually need several test items for each of the facets listed in Figure 1. A 'solution' which involved even more test items would result in even larger and even less feasible tests, this in turn implies that we must reduce dramatically the number of target words we test. Suppose, then, that we reduce our hypothetical list of target words to 10 items, and suppose that we develop a set of 20 sub-tests for each word. Even a minimal list of this sort would still require a battery of 200 subtests, and the nature of the material would probably require each subtest to be separately developed and validated. This does not feel like an attractive proposition to us. Furthermore, even if a testing program of this sort could be developed and deployed, we would still be left with the far from negligible problem of how we can generalise from our 10 target words to the rest of the vocabulary.

Put in its simplest terms, then, the prevailing approach to depth of vocabulary knowledge requires us to develop more and more finely tuned tests for fewer and fewer words. We do not think that this is a productive way to go, and our own thinking has led us in a rather different direction. Most people would agree that the recent growth in vocabulary research has largely been driven by the development of simple tests for vocabulary breadth – though for reasons which will become clear later, we prefer to call it **vocabulary size**. Typically in a test of this sort we give the test takers a large number of words and evaluate whether they 'know' these words or not. At first sight, this work looks as though we are primarily concerned with single words, but actually things are more complicated than this. If the target words are well-chosen, then we can extrapolate from the target words to an estimate of the test-taker's overall vocabulary size, and most tests of vocabulary breadth do just this. Thus, although we are ostensibly testing individual words, what really interests us is using this data to generate a description of the test takers' overall vocabulary size. **Vocabulary size** is not a feature of individual words: rather it is a characteristic of the test taker's entire vocabulary. This is a subtle shift of focus but an important one, and it has considerable implications for the way we approach measures of vocabulary depth.

We believe that the attempts made by researchers such as Wesche and Paribakht, and Schmitt focus in too much detail on knowledge of individual words, and neglect the larger picture. We believe that a better approach to vocabulary development would be to look at features which are characteristic of a learner's whole lexicon, rather than features which are characteristic only of single words. Ideally what we would like is a characteristic which scales in much the same way as vocabulary size measures scale. Vocabulary size is a good measure, with highly desirable measurement characteristics: vocabulary size measures start at zero, and they have a wide range, typically several thousand, and this means that they are very easy to work with, and very easy to interpret. Ideally we would like to develop a 'depth' characteristic with similar features.

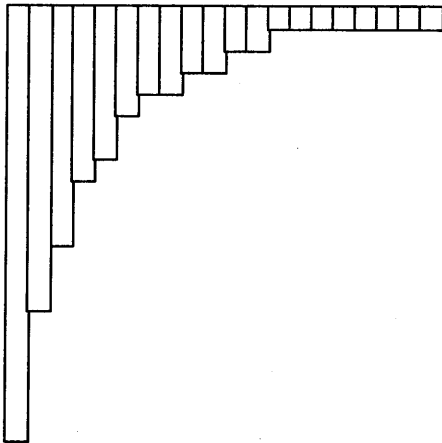
An alternative to breadth and depth

Our current view is that depth of vocabulary knowledge is rather more than the sum of the learners' knowledge of the individual words in their vocabulary. Knowledge of individual words contributes to depth of knowledge, but the really interesting feature of vocabularies is the way that the individual words that make them up interact with each other. These interactions are what distinguish between a mere **vocabulary list** and a **vocabulary network**. The basic idea, one that has been widely taken up by writers on vocabulary acquisition, e.g Aitchison (1987) and McCarthy (1990), is that words in a vocabulary form some kind of linked network. Aitchison, for example, refers to a lexicon as 'a gigantic multi-dimensional cobweb' (p72), while McCarthy talks in very similar terms. Although these authors do not develop these metaphors in any detail, we believe that we can approach the question of vocabulary depth by characterising the properties of this network rather than by focussing on the properties of its separate components. The difference between this view of vocabulary depth and the more traditional view is summarised in Figure 2.

The left hand diagram in Figure 2 illustrates the way vocabulary breadth and vocabulary depth are currently conceptualised. Each word is shown as a bar. Words with more 'depth' are shown as longer bars, while words with less 'depth' are shown as shorter bars. Essentially, this is a list model. Adding new words (increasing breadth) has no implications for the other words in the list, and there is no intrinsic link between breadth and depth.

The right hand diagram shows a more complex, network metaphor. In this model, 'breadth', or size, corresponds to the number of nodes in the network. The second dimension of this feature is the number of connections between the nodes. For this model, adding a new node (increasing 'breadth')

Vocabulary breadth and depth



Vocabulary size and organisation

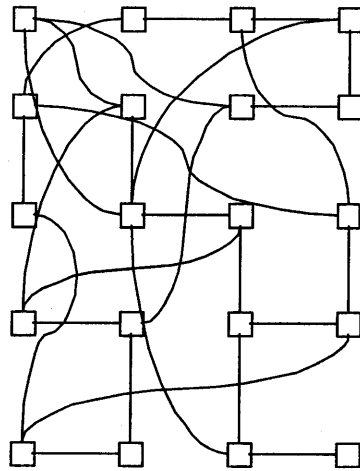


Figure 2. Two ways of looking at a vocabulary

does have implications for the rest of the network, depending on how the new node is linked to the existing ones. Adding new links (increasing 'depth') also has implications for the rest of the network.

The two metaphors are fundamentally different, and lead us to ask very different questions about the way 'breadth' and 'depth', or in our terms, **size** and **organisation**, interact. Basically, we think that the breadth/depth opposition is an unfortunate one, that leads in unhelpful directions. We believe that it makes more sense to talk about size and structure or size and organisation instead.

Our own research has been based on the idea that L2 lexicons are not as highly structured as the lexicons of L1 speakers. This seems like an intuitively plausible place to start: everyone agrees that L1 lexicons are highly developed and complex, while L2 lexicons are less well developed. In terms of our model, this should mean that L2 lexicons are smaller than L1 lexicons, and that the organisational links between the words that make up the L2 lexicon should be simpler than what we find in L1 lexicons.

The obvious way to investigate these ideas is to use word association data. In experiments of this sort, we give L2 speakers a series of single words, and we ask them to report the first L2 word that comes into their heads. We can then assume that the reported associations are linked in much the same way as the nodes in Figure 2 are linked. We might expect native speaker networks developed in this way to be denser and more highly organised than similar networks generated by L2 speakers, and this would suggest

that the complexity of the connections between words corresponds in some way to vocabulary 'depth'. Words which show a complex array of connections will tend to be more deeply known than words which are linked more tenuously to other words. This deceptively simple idea turns out to be much harder to work with than you would expect. Word associations generated by L2 speakers are quite different from those produced by L1 speakers (cf. Riegel and Zivian 1972), but the differences are very hard to pin down reliably in small scale experiments. This is largely because L2 speakers seem to produce a much wider range of associations than L1 speakers do, but it is also difficult to disentangle the effects of L1 interference in L2 word association tasks.

Most word association research relies on a methodology which requires test-takers to produce associations, and this tends to generate data which is particularly varied, and particularly difficult to work with. However, Wilks and Meara (2002) developed a sophisticated passive association recognition technique which allowed them to estimate the mean number of associational links between small sets of words. Their data showed that there were clear differences between native speakers and L2 speakers in this regard. In their approach, test-takers were provided with small sets of words and asked to decide whether any two words in each set were associated together. Not surprisingly, L1 speakers were more likely to find a link than L2 speakers were. Wilks and Meara computed the probability of a link being found for these sets, and then used a complex modelling method to estimate the complexity of the connections in their subjects' lexicons.

The work we report in the next section of this paper is basically a development of Wilks and Meara's methodology.

V_Links

The testing tools that we describe in this section are a preliminary attempt to develop a measure of lexical organisation for English. The test is known as V_Links, and its current version is version 2.00. The test consists of a set of 20 items. Each item consists of a selection of 10 words. The words all come from the first 1000 words in English. The test items were developed from a larger number of randomly selected word sets so that each set contains a number of obvious and some less obvious associational pairs. The test-takers are presented with each of these 20 items on a computer screen, and for each item they are given one minute to identify any association pairs that they can find. They do this by clicking on the words in the display. Each pair is confirmed when the test takers indicate how strong the association is by clicking on a four point scale at the bottom of the display. The display

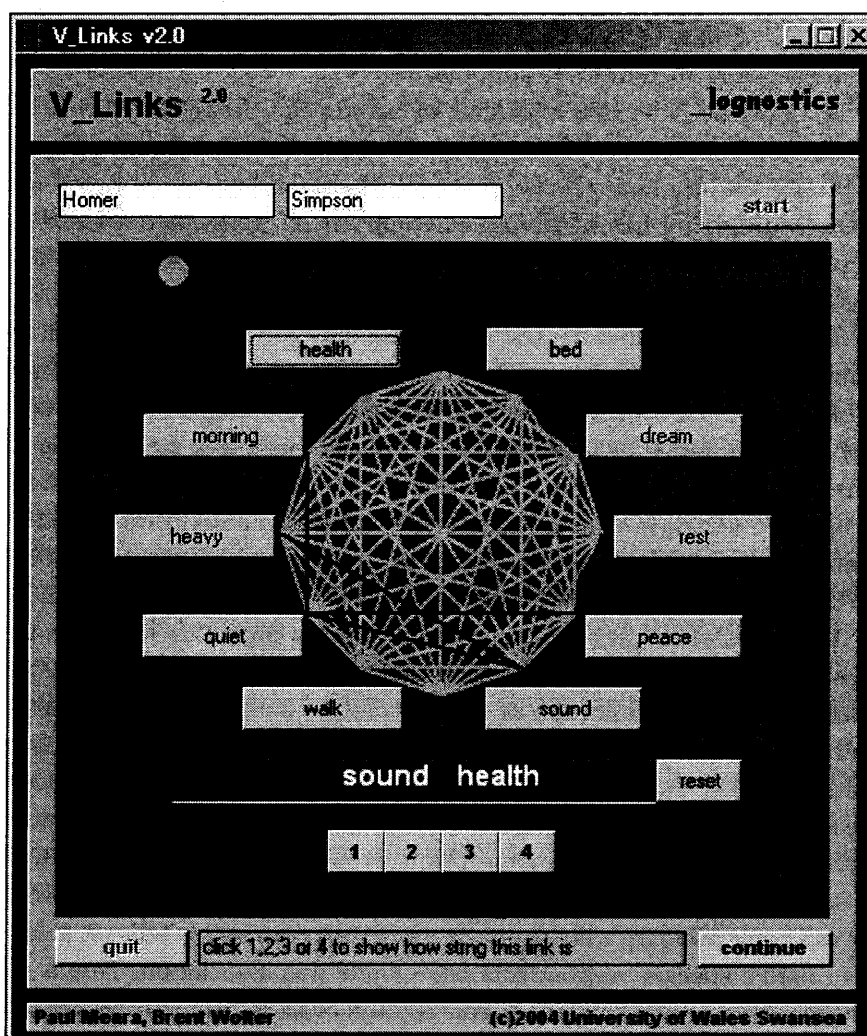


Figure 3. Screen shot from V_Links

then draws a link between the two members of the pair, with the strength of the display shown by differences in the line colour. (See Figure 3).

We have trialled this basic idea in a several different formats and the current version of our test works reasonably well. This version has a number of interesting features.

Firstly, it tests a large number of words in a relatively short space of time. Each of our 20 items contains 10 target words, so the whole test features a total of 200 words – one in five of the basic 1000 word core vocabulary. This figure is much larger than anything that could be attempted using an approach like VKS, and we think it gives us much greater insight into the

way a vocabulary is organised than a smaller test could. In spite of this, the test takes only 30 minutes to administer.

Secondly, each item has a possible 45 linked pairs, though in practice, the actual number of pairs identified is much smaller than this. Native speakers typically identify half a dozen word pairs as associational pairs for each item. Multiplying this up across all twenty items gives us a total of 120 providing us with a scale ranging from 0-120. This range seems to be large enough to clearly distinguish between native speakers and learners.

Thirdly, the fact that the test makes use only of words which lie in the first thousand frequency band for English means that the test in its current form can be used with test takers whose level of English varies over a considerable range of proficiency. Obviously, the test is not suitable for absolute beginners who have a very limited vocabulary, but it can be used with intermediate level learners, as well as advanced level learners, and the data we have collected so far suggests that the test may be sensitive enough to discriminate clearly between these cases.

There are, of course, a number of outstanding problems which we still need to address, and these form the object of our current work with the test format. The most important of these problems is that our L2 test takers persistently identify as associates word pairs which are never selected by native speaker test takers. We had originally hoped that these cases would be few, and that we would be able to ignore them, but this appears not to be the case. Our current approach to this problem has been to build up a database of the responses produced by a group of L1 speakers, and to accept as valid any response which appears more than once in this set – i.e. at least two native speaker respondents have made this association. This is not entirely satisfactory, as it fails to take account of L2 associations which arise as a result of specific local conditions – English loan-words used as trade names in Japan are a particular problem in this context – but in principle, the methodology could be adapted to take account of special cases such as these. Using a response database allows us to score the test automatically, and to provide instant feedback to test-takers.

The second problem is the question of association strength. In our earlier versions of V_Links, we asked test-takers to identify any associated pairs, but did not ask them to say how strong or how obvious the association was. This made the task easy for the test-takers, but it sometimes produced data which was difficult to interpret. Some test-takers, for example, would claim there was an association between a pair like COW and SNAIL, on the grounds that both were animals, or between LOOK and WRITE on the grounds that both were verbs. In our current version of V_Links, test-takers have to

indicate how strong they think each of their associations is, and we hope that this will allow us to weed out some of the more unsatisfactory associations in a principled way. Most people, for example, think that the association between DOG and CAT is stronger than the association between COW and SNAIL, and most people think that WRITE ~ PEN is a stronger association than WRITE ~ LOOK. However, this approach has thrown up other problems which we have not yet solved, notably a tendency for some test takers to claim that most associations are strong, while others appear to be very reluctant to use only the lower end of our four-point scale.

The third problem that we are still working on is the question of timing. Ideally, we would like to have a measure of vocabulary organisation which is independent of other factors, such as speed of word recognition and fluency. For this reason, some of our earlier versions of V_Links did not impose any time limit on the test-takers, and used an open-ended format instead. This worked well with some speakers, but others seemed to take a perverse delight in exploring all the possible combinations of words in each set, and finding obscure links between them. We have reintroduced a timer into the current version, with the time allowed for each test item being amply sufficient for test takers to identify the most obvious associations. It is possible that this makes the test harder for students whose reading speed is poor, but we do not think so. A more important factor seems to be how fluent test-takers are in using a mouse, and we think that this problem will disappear as more and more people are accustomed to this mode of working with a computer.

Does V_Links work?

The format we have described in this paper is the latest in a long series of trial versions which we have been working on for some time. V_Links clearly discriminates between native speakers and non-native speakers. In a large scale trial involving 147 L1-Japanese learners of English, the test showed a significant difference between these learners and a control group of native speakers, with the L2-speakers scoring about half the mean score for native speakers ($t = 3.25$, $p < .01$). We expect that our current version of V_Links will perform even better than this early trial version.

Data from the same group of subjects also suggests that there is only a very modest level of correlation between scores on the V_Links test and scores on a test of overall vocabulary size ($r < 0.3$), and this is exactly what we would expect if lexical organisation and size are more-or-less independent features of L2 lexicons. Clearly, further work on this is needed, and we will be carrying out more studies of this sort when we have finalised the current

version of V-Links.

Further work with V_Links

In the earlier sections of this paper, we argued that the size/organisation approach to L2 vocabularies was potentially more productive than the breadth/depth approach. In this section, we will explore this idea in more detail.

The size/organisation approach to vocabulary development is part of a multi-dimensional approach to L2 vocabularies that was first outlined in Meara (1996). Meara argued that both size and organisation were important characteristics which impacted on lexical behaviour. We have had reliable tests for measuring vocabulary size for some time. If we are right in thinking that V_Links is an effective way of assessing lexical organisation, then we now have tests for measuring these two basic dimensions in place, and this allows us to start asking some really interesting questions about the relationship between vocabulary size and vocabulary organisation. The basic question we can ask is whether organisation and size are correlated – i.e. whether the core vocabulary (the most frequent 1000 words) of a large lexicon is more structured than the same words are when they are part of a small lexicon. As we have seen, our preliminary results suggest that there is not a straightforward correlation between vocabulary size and vocabulary organisation. The question that then arises is just what is the relationship? Is it completely random or is it a complex non-linear relationship? The answer to this question is by no means obvious. There are, however, a number of plausible ways in which size and organisation might be related in a non-linear fashion.

One possibility is that people with similar sized vocabularies might differ in respect of how organised they are, i.e. we might find learners with similar vocabulary sizes, but very different degrees of organisation in their lexicons. If this turned out to be the case, then we might begin to ask how the different learner types identified by the dimensional approach differ in their language behaviour. We might expect learners with large, but weakly organised lexicons to behave differently from learners with similarly sized, but better organised lexicons – perhaps they would be less good at text comprehension, for example, or less good at understanding extended spoken input.

Another possibility is that lexical organisation may be an insignificant factor as long as the lexicons in question are below a critical size threshold, but that organisation becomes increasingly important once this critical size is reached. For example, it might be the case that small lexicons show a wide disparity in organisation, while large lexicons are always highly

organised. This idea in turn suggests that there might be a number of thresholds of this sort, and this would imply a complex relationship between size and organisation. Perhaps unstructured, or loosely structured lexicons can only grow to a limit, and cannot grow beyond this limit until they have restructured themselves. This would imply that lexicons might have growth phases and consolidation phases. We cannot think of any empirical work which supports this suggestion. However, it does fit well with some anecdotal accounts of vocabulary acquisition in L2 learners implying that learners feel their vocabulary reaches a sort of plateau from which it is difficult to make further progress.

V_Links should allow us to investigate these questions by carrying out detailed longitudinal studies designed to work out how vocabulary size and vocabulary organisation are related over time. Work of this sort would also indicate how far different learners follow the same trajectory in the space defined by our twin dimensions. At the moment, we have very little idea how much learners vary in the way their vocabularies are organised, and almost no idea how lexical organisation might facilitate further lexical growth, or how it might impact on other aspects of language performance. However, we expect to find considerable individual differences between learners in this respect, and if this turns out to be the case, then tests like V_Links will play an increasingly important role in vocabulary research.

Conclusion

In this paper, we have described our current thinking about lexical organisation, and shown how measures of vocabulary organisation offer a more interesting approach to the question of vocabulary development than the idea of 'vocabulary depth' does. We have described the current version of our tool, **V_Links**, and some of the preliminary investigations we have carried out using this tool. V_Links still has a way to go before it is fully functional, but we hope that this brief description of our current work will convince readers that the type of approach embodied in V_Links has the potential to open up some seriously interesting avenues in vocabulary research.

References

- af Trampe, P. (1983) 'Foreign Language Learning - a Criterion of Learning Achievement', in H. Ringbom (ed.), *Psycholinguistics and Foreign Language Learning*, Åbo Akademi: Åbo.
- Aitchison, J. (1987) *Words in the Mind*, Blackwell: Oxford.

- Blum-Kulka, S. (1981) 'Learning to Use Words: Acquiring Semantic Competence in a Second Language', in M. Nahir (ed.), *Hebrew Teaching and Applied Linguistics*, University Press of America: New York.
- Madden, J. F. (1980) 'Developing Pupils' Vocabulary Skills', *Guidelines* 3, 111-117.
- McNeill, A. (1996) 'Vocabulary Knowledge Profiles: Evidence from Chinese-Speaking ESL Teachers', *Hong Kong Journal of Applied Linguistics*, 1, 39-64.
- McCarthy, M. (1990) *Vocabulary*, Oxford University Press: Oxford.
- Meara, P.M. (1996) 'The Dimensions of Lexical Competence', in Brown, G., Malmkjaer, K. and J. Williams (eds.), *Performance and Competence in Second Language Acquisition*, 35-53, Cambridge University Press: Cambridge.
- Meara, P. M. and Milton, J.L. (2003) *X_Lex: the Swansea Vocabulary Levels Test*, Swansea: Lognostics.
- Nation, I. S. P. (1990) *Teaching and Learning Vocabulary*, Heinle and Heinle: Boston.
- Nation, I. S. P. (2001) *Learning Vocabulary in Another Language*, Cambridge University Press: Cambridge.
- Read, J. (1995) 'Refining the Word Associates Format as a Measure of Depth of Vocabulary Knowledge', *New Zealand Studies in Applied Linguistics*, 1, 1-17.
- Riegel, K. F. and Zivian, I. W. M. (1972) 'A Study of Inter- and Intralingual Associations in English and German', *Language Learning*, 22, 1, 51-63.
- Richards, J. C. (1976) 'The Role of Vocabulary Teaching', *TESOL Quarterly*, 10, 77-89.
- Schmitt, N. (1994) 'Vocabulary Testing: Questions for Test Development with Six Examples of Tests of Vocabulary Size and Depth', *Thai TESOL Bulletin*, 6, 2, 9-16.
- Schmitt, N. and Meara, P. M. (1997) 'Researching Vocabulary through a Word Knowledge Framework: Word Associations and Verbal Suffixes', *Studies in Second Language Acquisition*, 19, 1, 17-36.
- Schmitt, N., Schmitt, D. and Clapham, C. (2001) 'Developing and Exploring the Behaviour of Two New Versions of the Vocabulary Levels Test', *Language Testing*, 18, 1, 55-89.
- Wesche, M. and Paribakht, T. S. (1996) 'Assessing Vocabulary Knowledge: Depth vs. Breadth', *Canadian Modern Language Review*, 53, 1, 13-40.
- Wilks, C. and Meara, P. M. (2002) 'Untangling Word Webs: Graph Theory and the Notion of Density in Second Language Word Association Networks', *Second Language Research*, 18, 4, 303-324.
- Copies of V_Links v 2.00 are available from <http://www.swansea.ac.uk/cals/calsres/>