Mapping Vocabulary Research

Paul Meara (p.m.meara@gmail.com)
Emeritus Professor, Swansea University
ORCID: 0000-0001-8863-5374

Abstract: This paper illustrates how a bibliometric method known as co-citation analysis can help novice researchers find their way around a large research field. The method is illustrated using bibliographical data from a set of 100 papers in the area of L2 vocabulary acquisition. This is an area of research which has mushroomed in the last 20 years. The paper shows how a fairly accurate picture of current priorities in this field can be built up from relatively small amounts of bibliographical data, and suggests how beginning researchers might be able to use the approach to inform their own work.

Keywords: bibliometric method, co-citation, vocabulary acquisition, research, analysis.

Introduction

I first started researching vocabulary acquisition in an L2 nearly forty years ago. At the time, I was running research methodology courses at MA level, and getting increasingly frustrated dealing with students who wanted to write dissertations on attitude and motivation. Most of them developed attitude and motivation questionnaires, and correlated the results of the questionnaires with performance on an achievement test, with entirely predictable results. I couldn't help thinking that there ought to be a better way of organising research at MA level. What I needed was a research area where good MA students could do some genuinely original research and produce findings that might actually change the way we looked at second language learning. I searched a number of areas to find a focus for research of this sort, and it did not take me long to realise that the big gaping hole in second language research at that time was vocabulary acquisition. Hardly any serious

research had been undertaken in this area, and the text books were full of statements implying that vocabulary acquisition was a non-problem. Hockett's claim that "there is no point in learning large numbers of (words) until one knows what to do with them ... The acquisition of new vocabulary hardly requires formal instruction (Hockett 1958, p. 266) succinctly sums up the prevailing attitude. Notwithstanding Hockett's belief, it soon became clear that vocabulary acquisition was an ideal topic for student projects: there was no well-established research methodology, there was no entrenched orthodoxy, and most importantly, the research literature was so small that a highly motivated student could read it all in a matter of weeks. I was able to trace only nineteen vocabulary focussed research papers published in 1971, eighteen in 1972, nineteen in 1973 and twenty-eight in 1974. With so little research in play, it was easy to grasp what the main themes of vocabulary research were, and which of these ideas were worth looking at in more detail. We were also able to make excursions outside of Applied Linguistics, and look at the types of vocabulary research that was being undertaken by psychologists and sociolinguists, and examine how these ideas might impact on our understanding of vocabulary acquisition in an L2.

The situation forty years later is very different. Increasingly large amounts of research are published every year, and thanks to the internet most of this work is easy to get hold of – not like the old days when what you read was limited by what you could find in your university library, and how big your Inter-Library Loans budget happened to be. At the time of writing, The VARGA database (http://www.lognostics.co.uk/varga/) - a fairly comprehensive and reliable source for vocabulary research - lists 235 reports and papers published in 2015 alone – the most recent year for which reliable data is available. The number of papers published in this single year is only slightly smaller than the number of papers published during the whole of the 1970s, and there is no sign of this flood diminishing.

This huge quantity of research presents a massive problem for researchers. How do I know which of the many published papers are relevant to my research? Which ones do I need to read, and which can I safely ignore? Which papers are treading new paths, and which are just re-hashing old ideas? Whose work is influential, whose work is falling out of favour and whose work has been unjustly neglected? The traditional way of resolving this problem, especially for beginning research students, is to turn to standard text-books, particularly large overview texts. For vocabulary acquisition, the main sources of this type are Nation's *Teaching and Learning Vocabulary* first published in 1990, or the follow-up text *Learning Vocabulary in Another Language* which appeared in 2001 and the more recent second edition of this text which appeared in 2013. Other, less comprehensive reviews also exist, but none of them has the extensive coverage provided by Nation's books, which are outstanding examples of this genre and should be compulsory reading for all beginning researchers in L2 vocabulary acquisition. Relying on standard text books does have its drawbacks, however. Nation's texts tend to prioritise pedagogical issues, and solving problems which are related to classroom research. And necessarily text-books oversimplify the research – they just do not have the space to critically assess each of the sources they cite. Nation's most recent volume, for example, cites work by nearly 900 researchers, many of whom have published extensively in this area. The book, which most readers would consider to be exceptionally comprehensive in its coverage, runs to about 600 pages of text, leaving us with somewhat less than half a page of discussion for each author's work – perhaps two hundred words or so. It is easy to see that it is really hard to do justice to any researcher's body of work in such a small space.

Clearly, then, vocabulary research has now become such a vast enterprise that it is very difficult for any researcher coming newly to the field to build up a clear map of this research inside their own head, and we need to develop alternative ways of mapping the work that is being reported. This paper explores how one bibliometric approach - co-citation mapping - can help to resolve this problem. Co-citation

mapping can identify cohesive clusters of research within vocabulary studies, and this allows us to single out influential sources in a research field. The approach is also useful in highlighting the bibliometric properties of the research field, drawing attention to under-researched areas and to topics whose importance varies over time. This makes co-citation maps a tool which is of considerable relevance to young researchers.

Co-citation mapping

Co-citation works on the idea that it is often possible to guess what a paper is about by looking at the citations which appear at the end of the paper. Suppose, for example, that we find a paper that cites **Nation, Laufer** and **Schmitt** in its bibliography. We can be pretty sure that this paper is going to say something about L2 vocabulary. If we also find **John Read** in that list, then it is very likely that the paper deals with L2 vocabulary testing. If the list includes **Beglar** and **Hunt**, then we can be almost certain that the paper deals with a variant of Paul Nation's *Levels Test*. In this way, the citations that appear in a paper serve as signposts to commonly held ideas that are important in shaping the field. Co-citation analysis takes this idea a bit further. First developed in the 1960s and 1970s by Price (e.g. Price 1965) and by Garfield and Small (e.g. Small 1973), co-citation analysis is a formal way of capturing intuitive insights of the kind illustrated above.

A co-citation analysis moves through a number of steps. First we identify a body of work that we think represents the field we are interested in. For instance, if we were interested in the strategies that learners use to acquire L2 vocabulary, then we might choose an arbitrary set of papers that all included the word STRATEGIES in their titles.

Next, we extract a list of cited authors whose names appear in these papers. For each paper, each author is paired with every other author to make a list of co-citations. So,

for example, in the paper described at the beginning of this section we would end up with a list of co-citations that includes:

Beglar~Hunt, Beglar~Laufer, Beglar~Nation, Beglar~Read, Beglar~Schmitt, Hunt~Laufer, Hunt~Nation, Hunt~Read, Hunt~Schmitt, Laufer~Nation, Laufer~Read, Laufer~Schmitt, Nation~Read, Nation~Schmitt, and *Read~Schmitt.* This process generates huge amounts of data: if we have a paper which cites fifty authors then we end up with 50*49/2 =1225 co-citations. A set of 100 papers with, say, an average of 75 cited authors each – reference lists as long as this are a common characteristic of more recent research papers - would generate a co-citation set of well over 100,000 data points. Fortunately, the data can be easily processed using standard computing methods.

Finally, we generate a map which illustrates how the co-citation analysis clusters the sources. Again, fortunately, a number of computer programs exist which can generate maps of this sort. One such program is *Gephi* (Bastian, Heymann and Jacomy, 2009) which was used to generate the maps discussed later in this paper. Basically, Gephi takes a large list of co-citations and performs a cluster analysis, grouping together clusters of sources which tend to occur together in the paper set that we are analysing. These clusters can then be mapped in a way which shows the strength of links between the main clusters.

The analysis

How does a co-citation analysis help us understand what are the main trends in vocabulary acquisition research? In order to answer this question, we first need to identify a corpus of papers which are typical of the field. This preliminary selection is actually more complicated than it looks at first sight because it involves some degree of subjectivity. In this paper, the approach I have adopted is to select a set of 100 papers published between 1980 and 2014, all of which cite **Paul Nation's** work in their bibliographies. These papers were randomly selected from the VARGA database, and a list of the papers will be found in Appendix 1.

Selecting papers to analyse using a single author as the principle inclusion criterion is actually rather unusual, and it immediately introduces a number of obvious biases into the analysis. The justification for using Nation as an inclusion criterion in this paper is that every bibliometric analysis I have carried out in this area contains a very large and very dense cluster of researchers who are co-cited with Nation, and in that respect, ***Nation*** acts as a short-hand identifier for a particular kind of vocabulary research. However, we need to be aware that work which does not cite Nation is not included in our data sample, and this will bias the analysis away from research on topics like mnemonics, or child bilingualism, or neurolinguistic studies of bilingual speakers, that tend to be less focussed on the pedagogical issues in vocabulary acquisition. Work that is published in languages other than English is also less likely to cite Nation, and this means that research traditions which are strong in languages other than English (*Action Psychology* and *Availability* are good examples of this problem) are likely to be ignored. A further concern is that using a wide time span (here 34 years) for our selection introduces yet another kind of bias: it gives additional prominence to established scholars, and diminishes the importance of people who have started to publish only recently. For illustrative purposes, however, we can afford to ignore these problems. Obviously, we need to be aware of biases of this sort when we undertake a co-citation study.

A preliminary analysis of the data identified a total of 2612 unique authors cited in the sample papers. However, not all these authors are equally cited. Many are cited only once, while a handful of authors are cited very many times. Nearly two-thirds of the 2612 authors are cited in only one of the 100 papers. Nation, of course, is cited in every paper. A small group of 14 authors are cited in twenty-five or more papers. The distribution is shown in Table 1.

Table 1. The number of people cited in N papers in the sample data set.

1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
1814	345	154	82	47	31	26	20	11	16	12	6	8
14	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25+**	**100**
5	5	3	2	2	4	2	2	0	0	0	14	1

Table 1 shows that just over 200 authors are cited in at least five of the papers, and this group of influential authors will be taken as the focus of the analysis presented in this paper. We can think of these authors, who are listed in Table 2, as the most significant influences in L2 vocabulary acquisition research. Note that not all of these names are people who are directly involved in L2 vocabulary research. Some - Nagy, for example - are principally involved in L1 reading research. They appear here because L1 reading research has had a big impact on L2 vocabulary research, and the L2 work has benefitted greatly from methodological advances made by people working in this area. Another example is Aitchison, whose work is cited in 15 papers in the data set. Aitchison is best known for her work on the psycholinguistics of vocabulary. However, her own work is mainly concerned with L1 speakers, and only tangentially touches on L2 vocabulary use. In spite of this, her work is a significant source of ideas for L2 research, and this shows up in the large number of citations recorded here.

Table 2. The 216 most-cited authors in the sample data set.

Nation 100, Meara 60, Laufer 59, Hulstijn 46, Nagy 41, N Schmitt and Paribakht 37, Wesche 36, Anderson and Coady 33, Read 31, Herman 30, Krashen 27, N Ellis 26, R Ellis 25, Gass and Haynes 21, Carter and Huckin 20, Craik, Greidanus, McCarthy and Richards 19, Cohen and Sternberg 18, Grabe and Jenkins 17, Bensoussan, McKeown, Oxford, Aitchison, Baddeley, Haastrup, Hollander, Lockhart and Waring 15, Day, Joe, Knight, Mondria, Olshtain 14, Beaton, Bialystok, Brown, Cobb, Hatch, Hirsh, Meister and Saragi 13, Henriksen, Kelly, Levenston, Pressley, Schouten-van Parreren, Stoller and West 12, Bauer, Chamot, de Bot, Francis, Hazenberg, Horst, Parry, D Schmitt, Swain, Wit-de Boer, Wysocki 11, Carrell, Clapham, Coxhead, Crow,

de Groot, G Jones, Kucera, Lorge, Qian, Schmidt, Sim, Stahl, Thorndike, Tulving, Xue and Zimmerman 10, Aphek, Beck, Bloch, Carroll, Dubin, Li, Long, Melka-Teichroew, O'Malley, Schoonen and Watanabe 9, Arnaud, Bogaards, Buxton, Carnine, Chapelle, Chun, Faerch, Kameenui, Kroll, Lewis, McDaniel, Newton, Pica, Rott, Segalowitz, Stein, Verhallen, Vermeer and Webb 8, Barcroft, Bernhardt, Channell, Elley, C Fraser, Harley, Hiramatsu, Hogben, Jern, Lawson, Liu, Luppescu, McLaughlin, J Milton, Omura, Paivio, Pearson, Perry, Plass, Prince, Sanaoui, Schreuder, Seibert, Sharwood Smith, Sökmen, and Tanaka 7, Adolphs, R Anderson, Baldwin, T Brown, Crookall, Dupuy, Fairbanks, Goulden, Grace, Gu, Hall, Jiang, Kennedy, Larsen-Freeman, Lee, Leech, Levelt, Levin, GA Miller, Nattinger, Palmnerg, Perfetti, Rubin, Schatz, Singleton, Takala, Weltens, Yamazaki, Yoshida and Zechmeister 6, Alderson, Bahns, Baker, Barnett, Brandsford, Chern, EV Clark, coyke, curtis, d'anna, de Glopper, dufon, Eyckmans, Gairns, Goldstein, Gougenheim, Graney, He, Hu, Hubbard, Jacobs, Johansson, Kasper, Kintsch, Koda, MacWhinney, Magoto, McCaslin, Mokhtari, Moon, Nienhuis, Pitts, Powell, Redman, Robinson, Scarcella, Selinker, Skehan, Stern, Stewart, Stewner-Manzanares, Sutarsyah, Swaffar, M Thomas, van Hout and Wallace 5.

A superficial account of this data would identify these 216 people as a sort of Vocabulary Hall of Fame. However, as we have seen, anyone familiar with the literature will recognise that many of the people in this list would probably not consider themselves as main-stream L2 vocabulary researchers. A co-citation analysis allows us to examine the relationships between these most-cited authors in a lot more detail.

The maps

This section describes how to carry out a co-citation analysis and presents the results of an analysis of the small data set of vocabulary research papers described above. The first step in this analysis is to list all the co-citations among the 216 most-cited authors, where "co-citation" means that source A is cited alongside source B in the

same paper. (Note that this does not necessarily imply that source A actually cites source B, only that source A and source B tend to be cited together by some authors.) Some of these co-citation links are very weak, and it is customary practice to eliminate these weaker links in order to simplify the analysis. Setting five citation occurrences as a threshold for inclusion in our analysis gives us a set of 1805 co-citation links to analyse. This figure is about 1% of the 180,290 co-citations in the complete data set. Figure 1 shows a preliminary mapping of this data.

Authors cited at least five times, each co-citation occurring at least five times.

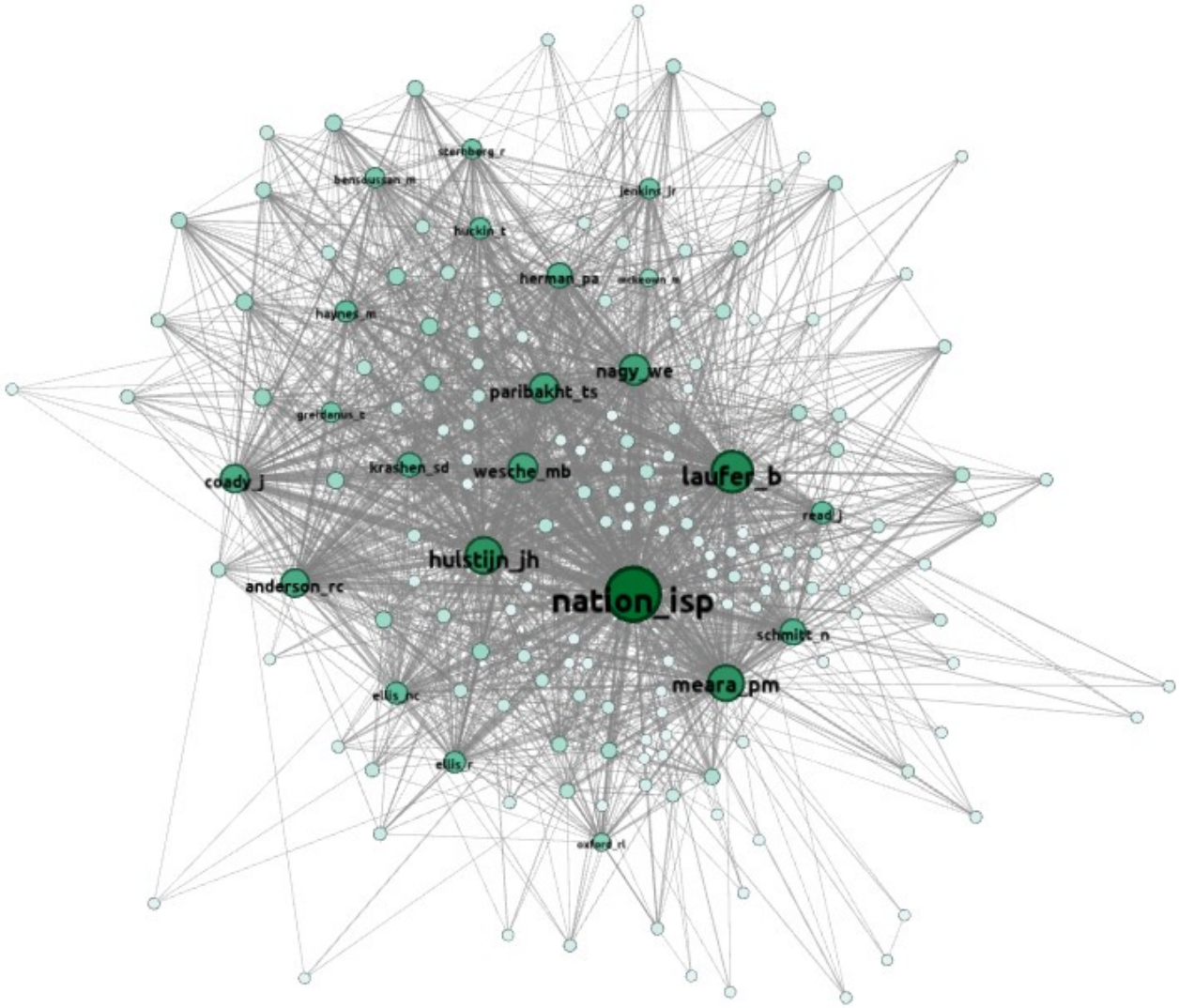

Figure 1. A preliminary mapping of the data set.

In this map, and the other mappings in this paper, the nodes represent the sources cited in the data set, and a link between two nodes indicates that they are co-cited in the bibliographies of some of the papers in the set. Thicker lines indicate a greater number of co-citations.

At first glance, this mapping appears to be fairly uninformative. What we have is a single large cluster, where everybody in the citation set is closely connected to everybody else. The map has a strongly inter-connected central core, and a periphery of nodes which connect to the core but have few connections with each other. We can infer from this that the vocabulary research mapped here is internally very consistent. This is a feature which has been identified in other bibliometric analyses of vocabulary research (e.g. Meara 2014), but one that appears to be rather unusual in other research fields, where more diffuse maps are the norm. It suggests that vocabulary research has some characteristics of orthodoxy, where almost everybody cites the same sources. This raises some interesting questions about whether this degree of consensus in the field is a good thing or not.

The mapping in Figure 1 shows that we have a very small group of authors who are central to the network. Nation, of course, is (by definition) co-cited along with everybody else in this map. However, we also have a small group of other researchers who are co-cited with nearly half of the other researchers in the map. This group includes Laufer (151 co-citations), Hulstijn (128 co-citations), Meara (124 co-citations), Nagy (98 co-citations), Paribakht (92 co-citations) and Wesche (91 co-citations). This central core accounts for about 45% of all the co-citation links in the map. They are also very tightly connected to each other, as we can see from Figure 2, with particularly strong co-citation links between Nation and Laufer, Nation and Hulstijn, and Nation and Meara. Nearly two thirds of the papers in this data set cite both Nation and Laufer, and both Nation and Meara; just under half cite Nation and Hulstijn. Laufer and Meara and Laufer and Hulstijn are cited together in about one third of the papers, and Hulstijn and Meara are co-cited in about a quarter of the

papers. A large number of the papers in the data set cite all of the authors in this central core.

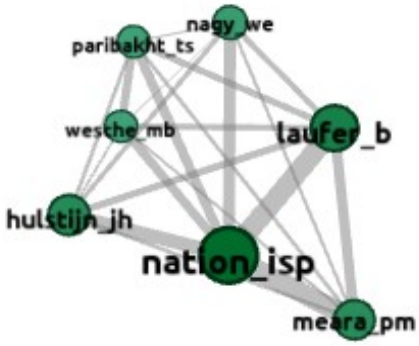

Figure 2: The Central core of the Vocabulary Research Network.

Paradoxically, because this central cluster of researchers plays such an important role in L2 vocabulary research, it does not actually provide us with very much information about the way the field is structured. By definition, everybody in this map is co-cited with Nation, so we could remove Nation's node from this map without losing any structural information. Similarly, almost everybody in the map is co-cited with Laufer and with Meara, so they too do not add much structural information to the mapping. This suggests that our map might look less monolithic and much more interesting if we remove the central core altogether. The results of carrying out this removal are shown in Figure 3. A small number of sources who appear in Figure 1 are co-cited only with members of the central core, so once the central core is removed from the map, these sources appear as isolated nodes with no connections. These isolated nodes have been removed from the map shown in Figure 3 for the sake of simplicity. The mapping program identifies four separate clusters of co-citations in the data. These clusters are identified in Figure 3 by distinct shading.

Co-citation map of the 100 papers with the Central Core removed.

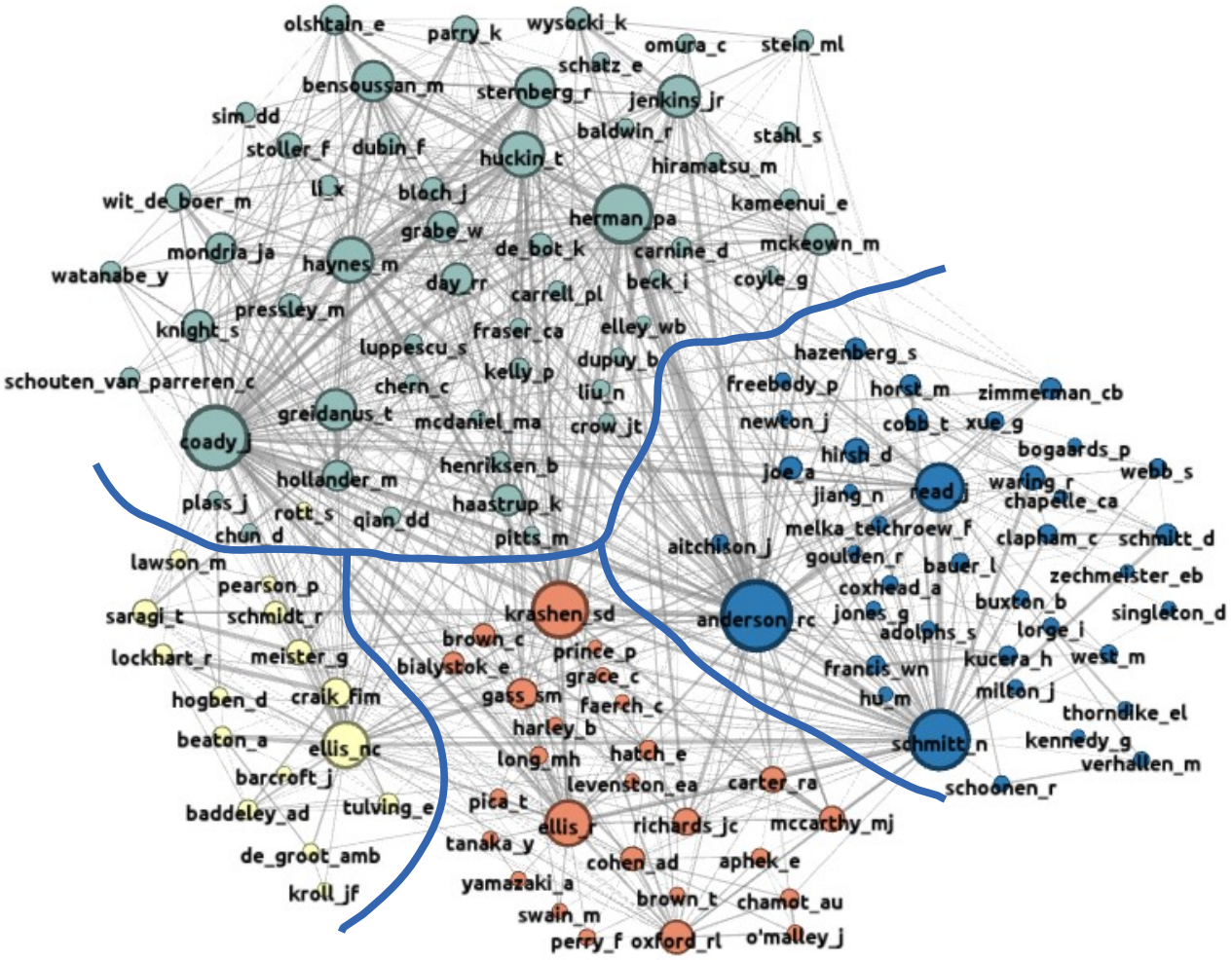

Figure 3. The donut map.

A number of interesting features emerge from this analysis. Firstly, anyone who is aware of the vocabulary research field as a whole will recognise that the clusters the program has identified make a kind of sense. **Cluster I** at the eastern edge of the map represents a group of people whose main research interest lies in vocabulary testing, and particularly the measurement of breadth and depth of vocabulary knowledge. This cluster, with Schmitt and Anderson as its most significant figures, contains a high proportion of people who are based in New Zealand, a feature which reflects the pre-eminence of Nation's research group at Wellington in this field. Most of the sources in this cluster are active L2 vocabulary researchers, but some other sources

are also included in the cluster. Thorndike and Lorge, Kucera and Francis and West, for example are the authors of important word lists that play a significant role in the development of vocabulary tests. In contrast, **Cluster II** at the SouthWestern corner of the map seems to consist principally of people who, while they are not themselves mainstream vocabulary researchers, have had a significant impact on the methodology and conceptualisation of L2 vocabulary research through their work in other areas, mostly psycholinguistics. The most notable source in this cluster is NC Ellis. The cluster also contains a group of psychologists whose main work is concerned with memory. De Groot and Kroll represent a group that works on formal models of bilingual lexicons. Only Barcroft is a significant L2 vocabulary researcher, suggesting that Barcroft's work might be rather different, and somewhat detached from mainstream L2 vocabulary research.

Cluster III at the northern edge of the map is the largest of the four clusters. The work of people in this cluster is mainly concerned with reading, and the map contains a significant group of L1 reading researchers. The significant influences in this cluster are Coady and Herman but we can also identify important sub-clusters of researchers concerned with lexical inferencing behaviour in L2 learners, glossing and mnemonics. **Cluster IV**, the final cluster at the southern edge of the map, dominated by Krashen and Rod Ellis, is perhaps the most difficult to pin down in terms of content. It contains people whose work is influential in L2 vocabulary acquisition research, without them being specifically interested in formal vocabulary testing. The cluster contains a number of people whose work concerns broader aspects of second language acquisition, strategies for vocabulary learning and negotiation of meaning. Overall, what is really surprising here is that a coherent and recognisable picture of the field clearly emerges from an analysis of only a handful of randomly selected source papers.

The second interesting feature which appears in this map is that there are different patterns of citation among the clusters. Cluster III shows very dense internal

connections: almost everyone in this cluster is co-cited with every-one else in the cluster, and only a few people positioned at the periphery of the cluster do not show this pattern. This cluster is also strongly linked to the other clusters: the co-citation link between Herman and Anderson is particularly strong, and Coady provides a strong bridge between Cluster III and Clusters II and IV. In contrast, the other three Clusters are somewhat less well-integrated into the map as a whole. Each of these clusters contains a small group of nodes which look to other clusters, and a larger group of nodes which only connect with other members of the cluster. Some of these cases arise for methodological reasons: word frequency is a big deal for vocabulary testers, for instance, so they frequently cite the Thorndike and Lorge, or the Kucera and Francis word counts, but this work seems to be of little importance to other strands of vocabulary research.

The third interesting feature is the key role of a small number of researchers in forging the links between the clusters. NC Ellis plays this role for Cluster II; Schmitt, Read and Anderson play the role for Cluster I; Herman and Coady play the role for Cluster III. All these people are strongly co-cited along with other sources outside their own cluster. In practical terms they are important because they represent bridges between what are often very diverse research traditions, and they facilitate the flow of information across the discipline. Cluster IV is different from the other clusters in that it does not seem to have a key figure of this sort. On the face of it, it looks as though Krashen is the principal linking node for this cluster. However, my gut feeling is that Krashen, being a well-known research figure, is often cited as a token vocabulary researcher by people who are not fully familiar with the field. (Krashen is frequently cited in the psychological literature, but he is cited much less frequently by the key sources in L2 vocabulary research.) My hunch is that the real heart of this cluster is R Ellis. It is worth noting, though, one characteristic of this cluster which makes it rather different from the other clusters in the mapping is that it contains a number of people who were active in the 1970s and 1980s. It may be

that this cluster is predominantly a historical one that is in the process of detaching from the larger map.

Repositioning the central core of sources - Nation, Laufer, Meara, Hulstijn, Paribakht, Wesche and Nagy - back into the donut map allows us to pick out some features which differentiate between its members. From Figure 1, we can see that Laufer and Meara appear to be most closely linked with the vocabulary testing cluster. Laufer, however, is often co-cited with members of Cluster III, whereas Meara is more often co-cited alongside the members of Cluster II. Hulstijn is only rarely co-cited with the testing cluster, but frequently co-cited alongside members of Cluster II and Cluster III. Paribakht, Wesche and Nagy are firmly placed within Cluster III, and are only rarely co-cited alongside the members of cluster II.

Discussion

The most surprising thing that emerges from this analysis is how much information can be extracted from a relatively small amount of source data. We began with a semi-random selection of only 100 papers, but the co-citation analysis has allowed us to identify the most significant trends in vocabulary research and the main figures whose work is cited within these research clusters. It looks as though relatively small amounts of bibliometric data can provide us with fairly accurate pictures of vocabulary research over long periods. But this suggests that it ought to be possible to construct a very accurate picture of current vocabulary research from a small set of papers chosen to represent a more condensed time span than the 34 years in our random sample. In fact, it would be relatively easy to construct a bibliometric history of vocabulary research by carrying out a number of analyses like the one described here. Some initial studies of this kind, covering in more detail the L2 vocabulary research published in 1982, 1983, 1984, and 1985 are to be found in Meara (2014, 2015, 2016, in press) respectively. These detailed annual surveys are intended to be the beginning of a series of larger maps, each covering a five year publishing window. From these larger maps, we might be able to begin constructing a

bibliometric history of vocabulary research. With extensive data of this sort we could easily establish who is influential at particular periods, who the rising stars of vocabulary research are, and how the focus of vocabulary research changes over time. A particularly important idea in bibliometrics is the idea of a "research front" – a cluster of ideas which come together and significantly change the direction of research in a field (Price 1965; Small 1973). A set of historical maps that allowed us to identify research fronts in vocabulary research and predict future trends in the field would be of considerable interest.

Two other applications of bibliometric analyses are also worth mentioning, as they might be particularly useful to beginning researchers.

One application is to use co-citation maps like the ones I have presented here as a way of positioning unfamiliar, new research. Experienced researchers make this kind of assessment every time they come across a new paper, but the way these assessments are made is often intuitive and implicit, and it is very difficult for inexperienced researchers to develop this skill. Using co-citation maps is one way of recognising how work you do not know well fits with what you know already, and whether it is worth looking at more closely.

Consider, by way of an example, a paper by Dotsenko and Leshchenko (2014) that crossed my desk recently. This is a short paper that cites only 16 sources. Four of these sources appear in our map – Nation, Meara, de Groot and Gass. The other twelve sources do not appear on the map. The four sources on the map all come from different clusters, mainly in the bottom half of the map, with an unusual co-citation linking Gass (Cluster IV) and de Groot (Cluster II). Links between these clusters are relatively scarce in Figure 3. Additionally, de Groot is very much on the edge of Cluster II, so this pattern of co-citations suggests that the paper might be working at the boundary between two rather different traditions. In fact, it uses word association data to make some points about the development of L2 word knowledge – research

of this kind is usually associated with Cluster I, the language testing cluster in Figure 3, and this suggests that Dotsenko and Leshchenko might represent a new angle on this question.

A particularly useful exercise for beginning researchers is to look at the authors they cite in their own work, and to position themselves on a co-citation map like Figure 3. A good example of this approach can be illustrated with a paper by Meara and Olmos Alcoy (2010). This paper has a relatively small citation list – Bell, Fitzpatrick, Laufer, Lincoln, Melka Teichroew, Malvern, Miralpeix, Nation, Petersen, Seber, Slagter and Zipf. Hardly any of these names appear in our map, but those that do (Laufer, Nation, Melka Teichroew) clearly place this paper close to Cluster I, the vocabulary testing cluster. The paper is actually an attempt to explore a radically innovative approach to measuring productive vocabulary that mimics the way ecologists count animal species in a natural environment, and this angle shows up in way the paper cites a cluster of ecologists – Lincoln, Petersen, Seber and Slagter - who do not figure in our map. In effect, what this paper is doing is drawing in new influences who do not figure in the current vocabulary research, but would do if research of this type takes off. Zipf is an important figure in lexico-statistics. His main work appeared in 1935, but it has recently started to re-appear as a reference in the L2 vocabulary research, and work that builds on Zipf's ideas is becoming increasingly important, particularly in vocabulary research that uses modelling approaches (e.g Edwards and Collins, 2011, 2013). Malvern, who is also cited in this list, exemplifies this type of approach in the context of measuring lexical diversity (Malvern, Richards, Chipere and Durán: 2004). Indeed, I would expect Zipf to emerge as a significant influence on L2 vocabulary acquisition research on a map of current research. Many people would consider papers like Meara and Olmos Alcoy (2010) to be "marginal" to vocabulary acquisition research. My own view is that really interesting research almost always takes place in the marginal areas of our map, usually at the edge of one of the major clusters, but sometimes in the space between two clusters. Things worth reading usually cite coherent groups of sources

which don't appear in the maps at all, but belong to different disciplines. In my experience, if your work only cites the central core of the map, then it is unlikely that you are bringing anything really original to the field.

A second useful exercise for beginning researchers is to use the co-citation maps as a way of generating interesting research questions. Figure 3 suggests that mainstream research in L2 vocabulary acquisition falls into four main clusters. But we could ask: what areas do NOT appear in this map? A number of obvious candidates spring to mind. Why does research on L1 vocabulary acquisition appear to have little influence on L2 research? Where is the L2 vocabulary research that uses neurolinguistics techniques? Why is there no obvious research strand that deals with the contexts of vocabulary acquisition? Why is the bilingual modelling research of de Groot and Kroll co-cited along the work of other psychologists, but only rarely co-cited alongside the work of mainstream vocabulary researchers? Where is the research that links L2 acquisition to language loss and aging research? Why do the sources in Cluster II appear to have very little influence on the work of vocabulary testing researchers? What might emerge if we brought these two strands of research closer together? And more generally, we might also fantasise about what kinds of research would emerge if we brought together people from opposite edges of the map, or injected ideas from completely unrelated areas of research.

Finally, we can use the co-citation maps to help us reflect on the broader characteristics of L2 vocabulary research. I mentioned earlier that mainstream research in this area is unusual in that almost all the research pays homage to a small group of researchers whose work is cited by almost everybody in the field. It is not obvious to me that this is a good thing. Much of this work is cited in an uncritical way, particularly by beginning researchers. It seems to me that a lot of the assumptions that form part of the work of the core group – and I am including my own work in this – need to be pulled apart and critically re-examined. A good example of this is the assumption that measuring vocabulary size must largely be

based on the frequency of words (or word families) in a native speaker corpus. This idea is one of the mainstays that characterises the work of both Nation, Meara, Laufer and Hulstijn. But we could ask why do we do it this way? What would happen if we simply agreed a Vocabulary Testing Word List, that reflected what learners actually know, one that lacked the theoretical foundation provided by corpus linguistics, but worked well in practice?

Hmmm! There is a lot of food for thought here. Maybe the main role of co-citation analysis is that it identifies orthodoxies for us to question, boundaries for us to push against, and uncharted territory waiting to be explored.

References

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. Icwsm, *8*, 361-362.Edwards, R and L Collins.

Lexical frequency profiles and Zipf's Law. *Language Learning* 61,1(2011), 1-30.

Edwards, R., & Collins, L. (2013). Modelling L2 vocabulary learning. *Vocabulary knowledge: Human ratings and automated measures*, 157-183. Amsterdam: John Benjamins.

Dotsenko, T., & Leshchenko, Y. L2 vocabulary acquisition: an experimental study of cross-linguistic interaction in adult bilingual mental lexicon. *ICT for Language Learning 5th Edition* 2014, 1-4.

Hockett, C. F. (1958). *A course in modern linguistics*. New York: Macmillan (1958).

Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development. Quantification and Assessment*. Houndmills, Basingstoke, Hampshire: Palgrave Macmillan.

Meara, P. M. (2014). Vocabulary research in the modern language journal: A bibliometric analysis. *Vocabulary Learning and Instruction*, *3*(1), 1-28.

Meara, P. (2014). Life before nation: Bibliometrics and L2 vocabulary studies in 1982. In *The functional perspective on language and discourse: applications and implications* (pp. 111-130). John Benjamins Publishing.

Meara, P. (2015). Vocabulary research in 1983: A bibliometric analysis. *Linguistics Beyond and Within*, *1*(1), 187-198.

Meara, P. (2016). Two steps backwards: A bibliometric analysis of L2 vocabulary research in 1984. *Linguistics Beyond and Within* 2, 139-152.

Meara, P. (2017). A new beginning? A bibliometric analysis of L2 vocabulary research in 1985 *Linguistics Beyond and Within* 3 (in press).

Meara, P. M., & Alcoy, J. C. O. (2010). Words as species: An alternative approach to estimating productive vocabulary size. *Reading in a Foreign Language*, *22*(1), 222-236.

Nation, ISP. (1990). *Teaching and learning vocabulary*. Rowley, Mass.: Newbury House.

Nation, ISP. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Nation, ISP. (2013). *Learning Vocabulary in Another Language*. Second Edition. Cambridge: Cambridge University Press.

Price, D. J. D. S. (1965). Networks of scientific papers. *Science*, 510-515.

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the Association for Information Science and Technology*, *24*(4), 265-269.

Zipf, GK. (1935). *The Psycho-Biology of Language. Boston, Ma.: Houghton Mifflin. 1935.*

Appendix 1: The 100 source papers used in this analysis. The sources are listed in chronological order.

van Parreren, CF and MC Schouten-van Parreren Contextual guessing: a trainable reader strategy. *System* 9,3(1981), 235-241.

Wallace, M *Teaching Vocabulary.* London: Heinemann. 1982.

Ndomba, D Acquiring English vocabulary and structures: some procedures and problems. *English Teaching Forum* 21,2(1983), 18-24.

Carter, R Is there a core vocabulary? Some implications for language teaching. *Applied Linguistics* 8,2(1987), 178-193.

Palmberg, R P atterns of vocabulary development in foreign language learners. *Studies in Second Language Acquisition* 9(1987), 201-220.

Kern, RG Second language reading strategy instruction: its effects on comprehension and word inference ability. *Modern Language Journal* 73(1989), 235-149.

Krashen, S We acquire vocabulary and spelling by reading: additional evidence for the input hypothesis. *Modern Language Journal* 73(1989), 440-464.

Pitts, M, H White and S Krashen n Acquiring second language vocabulary through reading: a replication of the *Clockwork Orange* study using second language acquirers. *Reading in a Second Language* 5,2(1989), 271-275.

Abe, H How ESL students acquire the meanings and use-potentials of communicative vocabulary. *Reading in a Foreign Language* 10(1991), 9-27.

Laufer, B The development of lexis in the expression of the advanced learner. *Modern Language Journal* 75,4(1991), 440-448.

Clarke, M Vocabulary learning with and without computers: some thoughts on a way forward. *CALL Journal* 5(1992), 139-146.

Laufer, B How much lexis is necessary for reading comprehension? In: **P Arnaud and H Béjoint** (eds.) *Vocabulary and Applied Linguistics*. London: Macmillan. 1992.

Hulstijn, JH Retention of inferred and given word meanings: experiments in incidental vocabulary learning. In: **P Arnaud and H Béjoint** (eds.) *Vocabulary and Applied Linguistics*. London: Macmillan. 1992.

Goodfellow, R CALL for vocabulary: requirements, theory and design. *Computer Assisted Language Learning* 6(1993), 92-122.

Hulstijn, J When do foreign language readers look up the meaning of unfamiliar words? The influence of task and learner variables. *Modern Language Journal* 77(1993), 139-147.

Maiguashca RU Teaching and learning vocabulary in a second language. *Canadian Modern Languages Review* 50(1993), 83-100.

Knight, S Dictionary: the tool of last resort in foreign language reading? A new perspective. *Modern Language Journal* 78(1994), 285-299

Beaton, A, M Gruneberg and N Ellis Retention of foreign vocabulary learned using the key-word method: a ten year follow-up. *Second Language Research* 11,2(1995),

Ellis, NC The psychology of foreign language vocabulary acquisition: implications for CALL. *Computer Assisted Languag Learning* 8,2-3(1995), 103-128.

Joe, A Text based tasks and incidental vocabulary learning. *Second Language Research* 11,2(1995), 95-111.

Sanaoui, R Adult learners' approaches to learning vocabulary in second languages. *Modern Language Journal* 79(1995), 15-28.

Schmitt, N An examination of the behaviour of four vocabulary tests. In: **D Allan** (ed.) *Entry Points*. Whitstable: IATEFL. 1995. 34-39.

Schmitt, N and D Rae Schmitt Vocabulary notebooks: theoretical underpinnings and practical suggestions. *English Language Teaching Journal* 49,2(1995), 133-143.

Chun, DM and JL Plass Effects of multimedia annotations on vocabulary acquisition. *Modern Language Journal* 80,2(1996), 183-198.

Hazenberg, S and JA Hulstijn Defining a minimal receptive second language vocabulary for non-native university students: an empirical investigation. *Applied Linguistics* 17,2(1996), 145-163.

Hulstijn, JH, M Hollander and T Greidanus Incidental vocabulary learning by advanced foreign language students: the influence of marginal glosses, dictionary use and the reoccurrence of unknown words. *Modern Language Journal* 80(1996), 327-339.

Prince, P Second language vocabulary learning: the role of context versus translations as a function of proficiency. *Modern Language Journal* 80,4(1996), 478-493.

de Bot, K, TS Paribakht and M Wesche

Toward a lexical processing model for the study of second language vocabulary acquisition. *Studies in Second Language Acquisition* 19(1997), 309-329.

Laufer, B and L Hadar Assessing the effectiveness of monolingual, bilingual and 'bilingualised' dictionaries. *Modern Language Journal* 81,2(1997), 189-196.

Schmitt, N and PM Meara Researching vocabulary through a word knowledge framework: word associations and verbal suffixes. *Studies in Second Language Acquisition* 19,1(1997), 17-36.

Waring, R A comparison of the receptive and productive vocabulary sizes of some second language learners. *Immaculata* 1(1997), 53-68. Notre Dame Seishin University.

Duquette, L, D Renie and M Laurier The evaluation of vocabulary acquisition when learning French as a second language in a multimedia environment. *Computer Assisted Language Learning* 11(1998), 3-34.

Grace, C Retention of word meanings inferred from context and sentence level translations: implications for the design of beginning level CALL software. *Modern Language Journal* 82(1998), 533-544.

Joe, A What effects do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied Linguistics* 19,3(1998), 357-377.

Laufer, B The development of passive and active vocabulary in a second language: same or different? *Applied Linguistics* 19,2(1998), 255-271.

Lomicka, LL To gloss or not to gloss: an investigation of reading comprehension online. *Language Learning and Technology* 1,2(1998), 41-51.

Scarcella, R and C Zimmerman ESL student performance on a test of academic lexicon. *Studies in Second Language Acquisition* 20,1(1998), 27-49.

Fraser, CA Lexical processing strategy use and vocabulary learning through reading. Studies in Second Language Acquisition 21(1999), 225-241.

Henriksen, B Three dimensions of vocabulary development. *Studies in Second Language Acquisition* 21(1999), 303-317.

Huckin, T and J Coady Incidental Vocabulary Acquisition in a second language: a review. *Studies in Second Language Acquisition* 21(1999), 181-193.

Koren, S Vocabulary instruction through hypertext: are there advantages over conventional methods of teaching? *TESL-EJ* 4,1(1999) A-2.

Mori, Y Beliefs about language learning and their relationship to the ability to integrate information from word parts and context in interpreting novel kanji words. *Modern Language Journal* 83(1999), 534-547.

Rott, S The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition* 21,4(1999), 589-620.

Paribakht, TS and M Wesche Reading and "incidental" L2 vocabulary acquisition. *Studies in Second Language Acquisition* 21(1999), 195-224.

Bogaards, P Testing vocabulary knowledge at a high level: the case of the Euralex French Tests. 21(2000), 490-416.

Jiang, N Lexical representation and development in a second language. *Applied Linguistics* 21,1(2000), 47-77.

Smith, R Vocabulary for academic writing: hit and miss? In: **D Brooks, R Long and J Robbins** (eds.) *On JALT 1999: Teacher belief, teacher action: connecting research and the classroom.* Tokyo: JALT. 2000. 172-177.

Wesche, M and TS Paribakht Reading-based exercises in second language vocabulary learning: an introspective study. *Modern Language Journal* 84(2000),196-213.

Bogaards, P Lexical units and the learning of foreign language vocabulary. *Studies in Second Language Acquisition* 23,3(2001), 321-344.

Greidanus, T and L Nienhuis Testing the quality of word knowledge in L2 by means of word associations: types of distracters and types of associations. *Modern Language Journal* 85(2001), 567-577.

Helms-Park, R Evidence of lexical tranfer in learner syntax: The acquisition of English causatives by speakers of Hindi-Urdu and Vietnamese. *Studies in Second Language Acquisition* 23(2001), 71-102.

Morin R and J Goebel Basic vocabulary instruction: teaching strategies or teaching words? *Foreign Language Annals* 34,1(2001), 8-17.

Wolter, B Comparing the L1 and the L2 mental lexicon. *Studies in Second Language Acquisition* 23(2001), 41-69.

de la Fuente, MJ Negotiation and oral acquisition of vocabulary. *Studies in Second Language Acquisition* 24(2002, 81-112).

Adolphs, S and N Schmitt Lexical coverage of spoken discourse. *Applied Linguistics* 24,4(2003), 425-438.

Barcroft, J Effects of questions about word meaning during L2 Spanish lexical learning. *Modern Language Journal* 87,4(2003), 546-561.

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

Fan, Y Frequency of use, perceived usefulness, and actual usefulness of second language vocabulary strategies: a study of Hong Kong learners. *Modern Language Journal* 87,2(2003), 222-241.

Frantzen, D Factors affecting how second language Spanish students derive meaning from context. *Modern Language Journal* 87,2(2003), 168-199.

Mondria, J-A The effects of inferring, verifying and memorizing on the retention of L2 word meanings: an experimental comparison of the "meaning-inferred method" and the "meaning-given method". *Studies in Second Language Acquisition* 25,4(2003), 473-500.

Mori, Y The roles of context and word morphology in learning new kanji words. *Modern Language Journal* 87,4(2003), 404-420.

Morin, R Derivational morphological analysis as a strategy for vocabulary acquisition in Spanish. *Modern Language Journal* 87,2(2003), 200-221.

Vidal, K Academic listening: a source of vocabulary acquisition? *Applied Linguistics* 24,1(2003) 56-89.

Gardner, D

Vocabulary input through extensive reading: a comparison of words found in children's narrative and expository reading materials.*Applied Linguistics* 25,1(2004), 1-37.

Jiang, N

Semantic transfer and its implications for vocabulary teaching in a second language. *Modern Language Journal* 88,3(2004), 416-432.

Smith, B Computer-mediated negotiated interaction and lexical acquisition. *Studies in Second Language Acquisition* 26,3(2004), 365-398.

Meara, PM Lexical frequency profiles: a Monte Carlo analysis. *Applied Linguistics* 26,1(2005), 320-347.

Greidanus, T, B Beks and R Wakely Testing the development of French word knowledge by advanced Dutch- and English-speaking learners and native speakers. *Modern Language Journal* 89(2005), 221-233.

Qian, DD Measuring lexical richness in business English writing: A study of Chinese learners. *The Hong Kong Linguist: Jubilee Version*25(2005), 36-42 .

Webb, SA Receptive and productive vocabulary: the effects of reading and writing on word knowledge. *Studies in Second Language Acquisition* 27,1(2005), 33-52.

Zareva, A, P Schwanenflugel and Y Nikolova Relationship between lexical competence and language proficiency: variable sensitivity. *Studies in Second Language Acquisition*27,4(2005), 567-596.

Aizawa, K Rethinking frequency markers for English-Japanese dictionaries. In: **M Murata, K Minamiide, Y Tono and S Ishikawa** (eds.)*English Lexicography in Japan*. Tokyo: Taishukan. 2006

Fukkink, RG, J Hulstijn and A Simis Does training in second language word recognition skills affect reading comprehension? An experimental study. *Modern Language Journal* 89,1(2005), 54-75.

Nassaji, H The relationship between depth of vocabulary knowledge and L2 learners' lexical inferencing strategy and use. *Modern Language Journal* 90,3(2006), 387-401.

Sagarra, N and M Alba The key is in the keyword: L2 vocabulary learning methods with beginning learners of Spanish. *Modern Language Journal* 90(2006), 228-243.

Tseng, W-T, Z Dörnyeei and N Schmitt A new approach to assessing strategic learning: the case of self-regulation in vocabulary acquisition. *Applied Linguistics* 27,1(2006), 78-102.

Gardner, D Validating the construct of WORD in applied corpus-based vocabulary research: a critical survey. *Applied Linguistics* 28,2(2007), 241-265.

Steinel, M, J Hulstijn and W Steinel Second language idiom learning in a paired associate paradigm. *Studies in Second Language Acquisition* 29(2007), 449-484.

Webb, S The effects of repetition on vocabulary knowledge. *Applied Linguistics* 28,1(2007), 46-65.

Hu, CF Rate of acquiring and processing L2 color words in relation to L1 phonological awareness. *Modern Language Journal* 92,1(2008), 39-52.

Jiménez Catalán, RM and M Terrazas Gallego The receptive vocabulary of English foreign language young learners. *Journal ofEnglish Studies* 5-6(2005-2008), 173-191.

Laufer, B and N Girsai Form Focussed Instruction in second language vocabulary learning: a case for contrastive analysis and translation. *Applied Linguistics* 29,4(2008), 694-716.

Lindstromberg, S and F Boers The mnemonic effect of noticing alliteration in lexical chunks. *Applied Linguistics* 29,2(2008), 200-222.

Vasiljevic, Z Teaching Vocabulary to Advanced Japanese Students: A Word Association Approach. *The East Asian Learner* 4,1(2008), 34-46.

Webb, S Receptive and productive vocabulary size of L2 learners. *Studies in Second Language Acquisition* 30,1(2008), 79-95.

Harrington, M and M Carey The on-line Yes/No test as a placement tool. *System* 37,4(2009), 614-626.

Pulido, D How involved are American L2 learners of Spanish in lexical input processing tasks during reading? *Studies in Second Language Acquisition* 31(2009), 31-58.

Staehr, LS Vocabulary knowledge and advanced listening comprehension in English as a Foreign Language. *Studies in Second Language Acquisition* 31,4(2009), 577-608.

Webb, S and MPH Rodgers The lexical coverage of movies. *Applied Linguistics* 30,3(2009), 407-427.

Barcroft, J and S Rott Partial word form learning in the written mode in L2 German and Spanish. *Applied Linguistics* 31,5(2010), 623-650.

Eldridge, J, S Neufeld and N Hancioglu Towards a lexical framework for CLIL *International CLIL Research Journal* 1,3(2010), 88-103.

Simpson_Vlach and NC Ellis An academic formulas list: new methods in phraseology research. *Applied Linguistics* 31,4(2010), 487-512.

Yu, G Lexical diversity in writing and speaking task performances. *Applied Linguistics* 31,2(2010), 236-259.

Webb, S and E Kagimoto Learning collocations: do the number of collocates, position of the node word and synonymy affect learning? *Applied Linguistics*32,3(2011), 259-276.

Bardel, C A Gudmundson and C Lindqvist Aspects of lexical sophistication in advanced learners' oral production. *Studies in Second Language Acquisition* 34(2012), 269-290.

Pichette, F, L De Serres and M Lafontaine Sentence reading and writing for second language vocabulary acquisition. *Applied Linguistics* 33,1(2012), 1-20.

Subon, F Vocabulary learning strategies employed by Form 6 students. *International Journal of Scientific and Research Publications*3,6(2013), 1-32.

Meara PM Life before Nation: Bibliometrics and L2 vocabulary studies in 1982. In: **M Gómez González, FJ Ruiz de Mendoza Ibañez, F Gonzálvez García and A**

Downing. *The Functional Perspective on Language and Discourse: Applications and implications*. Amsterdam: John Benjamins. 2014. 111-129.

Meara, PM Vocabulary research in *The Modern Language Journal*: A bibliometric analysis. *Vocabulary Learning and Instruction* 3(2014), 1-28.

Merikivi, R and P Pietilä Vocabulary in CLIL and in mainstream education. *Journal of Language Teaching and Research* 5,3(2014), 487-497.