

# A time for consolidation: A bibliometric analysis of L2 vocabulary research 1985–1989

Paul Meara

*Swansea University, United Kingdom*

## Abstract

This paper uses a co-citation analysis to examine the research on L2 vocabulary acquisition that was published in 1989. Two analyses are presented. The first is a detailed account of the 1989 research on its own terms. The second analysis places this work in a larger context by looking at research published in a five-year window covering 1985–89. The analyses identify important themes in the research and significant sources who are influencing the way the research is developing at this time. The main features of this work are the substantial growth in dictionary and corpus research, and the emergence of Paul Nation as the Most Significant Source in 1989.

**Keywords:** L2 vocabulary acquisition, vocabulary research, bibliometric analysis

## 1. Introduction

This paper is the ninth in a series of studies in which I have been mapping out the way L2 vocabulary research has developed over the last 50 years. Beginning with the 1982 research, these papers have presented bibliometric maps for each year between 1982 and 1988 (Meara 2014, 2015, 2016, 2017, 2018, 2019, and 2020), with an exploratory foray into some more recent data from 2006 (Meara: 2012). This paper is a sequel to my earlier LingBaW papers, in that it covers the research output for 1989.

The paper falls into two parts. Part 1 reviews the 1989 research in its own terms. Part 2 puts this research into a wider context by summarising the main trends that appear in a five-year window covering 1985–89. Both parts use the author co-citation method, developed by Small (1973). Small's method is described in detail in Appendix 1 for readers who are not yet familiar with the approach used in these papers.

## 2. Part 1. The new research published in 1989

1989 seems to have been a good year for L2 vocabulary research. The previous year, 1988, saw a decline in the number of outputs, following a sustained rise from 1984. In 1989, the number of outputs is back to 1987 levels, and the VARGA data base (Meara n.d.) identifies 142 sources published in 1989. More importantly, the type of publication appearing in this year seems to be shifting, as the vocabulary research becomes more focussed and less haphazard. Notably, in 1989, we have several collections of papers and special issues of journals that deal with vocabulary issues. These include a special issue of the influential Dutch Journal *Toegepaste Taalwetenschap in Artikelen*, a special issue of the French Journal *Le français dans le monde*, and a special issue of the *AILA Review* edited by Ron Carter.

Several collections of edited papers also appeared in 1989: Tickoo edited a State of the Art collection on L2 dictionary use (Tickoo: 1989) and James edited a collection of papers on lexicography (James: 1989), a theme which is also taken up in volumes edited by Zagrebelsky (1989) and by Marelllo (1989). These last two sources are important, because they hint at a seam of vocabulary research in Italian, which has not previously appeared in our analyses. This year also saw the publication of a number of monographs dealing with L2 vocabulary issues. The most important of these are Francine Melka Teichroew's volume on receptive and productive vocabulary (Melka Teichroew: 1989) and a monograph by Elisabet Service which dealt with phonological coding and working memory as they relate to L2 vocabulary learning (Service: 1989). Another notable publication this year is Kirsten Haastrup's thesis (Haastrup: 1989), which deals with lexical inferencing by L2 readers. Other book length volumes also appeared in this year's outputs: Hammer and Giauque (1989) is a series of short chapters on cognates in L2. Helliwell (1989) appears to be mainly interested in lexical errors. The fourth edition of Hornby's *Oxford Advanced Learner's Dictionary of Current English* also appeared in 1989. Table 1 lists these works in more detail.

**Table 1:** Work published in 1989, but excluded from the analysis in this paper. For the reasons explained in my earlier papers, these items are not included in the main analysis in this section. Book length treatments tend to have citation patterns which differ from the patterns we find with the shorter research papers and chapters which form the bulk of the 1989 data set.

### BOOKS and MONOGRAPHS

- Hammer, P. and G. S. Giauque. 1989. *The role of cognates in the teaching of French*. New York.
- Helliwell, M. 1989. *Can I become a beefsteak? Trügerische Wörter zum Nachschlagen und Üben*. Berlin.
- Hornby, A. S. (ed.) 1989. *Oxford Advanced Learner's Dictionary of Current English*. Oxford. 4th edition.
- Marelllo, C. 1989. *Dizionari bilingui*. Bologna.
- Meara, P. M. 1989. *Beyond Words*. London.
- Melka Teichroew, F. J. 1989. *Les notions de réception et de production dans le domaine lexicale et sémantique*. Berne.
- Service, E. 1989. *Phonological coding in working memory and foreign language learning*. University of Helsinki: General Psychology Monographs, B9.
- Sim, DD and B Laufer-Dvorkin. 1989. *Vocabulary development*. Glasgow.

### THESES and DISSERTATIONS

- Haastrup, K. 1989. *Lexical inferencing procedures or talking about words*. PhD thesis: University of Copenhagen.
- Li Lan. 1989. *Extending vocabulary in English for specific purposes – with special reference to Chinese University students*. MPhil thesis, Exeter University,

## 2.1. The data sources

A total of 122 ordinary papers were identified as published in 1989. Most of these papers were regular journal articles, but a new feature of the field seems to be the large number of papers which appeared as book chapters. There were 35 publications of this type – more than a quarter of the total output for this year. Fifteen of these items appeared in the thematic collection of papers edited by Tickoo, two items appear in Zagrebelsky, and a further two items in the thematic volume edited by James. All three of these volumes deal with dictionaries. The other items of this type appear in edited volumes which do not have a thematic focus on vocabulary matters.

As usual, a small number of eligible papers turned out to be unobtainable. These items are listed in Table 2, and they are not included in the analyses reported later in this report. It is unlikely that the absence of these items would make a significant difference to the analysis reported later. However, it is worthwhile pointing out that two of these sources are in German, and the other two sources are in Japanese, and Dutch respectively. Publication of German language resources in 1989 seems to have been seriously disrupted by the political events surrounding the fall of the Berlin wall, and the subsequent regime change in East Germany. However, Scherfer does appear as an author of other papers in the data set. Vermeer, too, has other outputs in the 1989 data set.

**Table 2:** *Items that appeared in 1989 but were untraceable.*

|                                                                                                                                                                |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Harado, H. 1989. Kanji no shidōhō: kanji-kei. In: Katō, A (ed.) <i>Kōza nihongo kyōiku 9, nihongo no moji to hyōki</i> . Meiji Shoin: 1989. 265–289.           |
| Mrazovic, P. and W. Stötting-Richert. 1989. Die Wortschatzbeherrschung zweisprachiger jugoslawischer Schüler in Niedersachsen. <i>Deutsch Lernen</i> 4: 28–55. |
| Scherfer, P. 1989. Vokabellernen. <i>Der Fremdsprachen Unterricht</i> 98: 4–10.                                                                                |
| Vermeer, A. 1989. Woordenschat en Nederlands als tweede taal: een experiment. <i>Maggezien</i> 14,3: 16–21.                                                    |

This leaves us with a set of 118 papers that make up the data set to be analysed here. This is a significant uplift on the 1988 data set, which included only 81 papers. For reasons of space, the items included in the data set are not listed here, but interested readers can identify them in the Vocabulary Acquisition Research Group Archive <https://www.lognostics.co.uk/varga/>

Typing ## 1989 into the search box, will return a complete list of these papers.

As usual, we begin with a superficial analysis of the data set. 134 authors contribute to the 1989 data set, and once again the vast majority of these authors contribute only one paper: 109 of the 134 authors fall into this category – a considerable increase on the 1988 figure of 83. Proportionately, however, the number of single paper authors has decreased from 90% in 1988 to 81% in 1989.

There is a considerable turnover in the list of authors who contribute to more than one paper. Several of last year's prolific authors no longer maintain their position in 1989, notably McCarthy, Johns, Tono, and Summers. Almost all of the authors of multiple papers are new in 1989: only Carter, Meara, Nation and Robinson appeared in both the 1988 prolific author list and in this year's list. Overall, the number of authors contributing a large number of outputs in 1989 is down. The outstanding contributor in 1989 is Zimmerman with four outputs. Four

authors (Feldman, Katz, Meara and Robinson) contribute three papers each – a small increase on the 1988 figure. The main growth appears to come from authors who contribute to two outputs in 1989. Eighteen authors fall into this group (Lukatela, Turvey and Carello, Hafiz and Tudor, Carter, HC Chen, Harley, Hartmann, Hulstijn, Krashen, Laufer, Mondria, Nation, Scherfer, Stein, Tickoo and Weltens) – almost five times as many sources contributed to two outputs in 1989 as was the case in 1988. (See Table 3).

The overall result of these shifts is that the distribution of authors in the 1989 data set is much more normal than it was in previous years. “Normal” here means that the pattern of authorship conforms to a logarithmic distribution described by Lotka (1926), and commonly known as Lotka’s Law. (cf. Appendix 2). Lotka’s model estimates the number of authors we might expect to be making

**Table 3:** *The number of authors contributing to N outputs in the 1989 data set.*

| Contributions | 6 | 5 | 4 | 3  | 2  | 1   |
|---------------|---|---|---|----|----|-----|
| 1989 data     |   |   | 1 | 4  | 18 | 109 |
| Lotka’s model | 3 | 4 | 7 | 12 | 27 | 109 |

multiple contributions to the data set given that we have 109 authors who contribute to only one paper. The relevant figures, shown in the bottom line of Table 3, suggest that we might expect a sizeable number of authors to be contributing to five or more papers in this data set. This does not appear to be the case in 1989, but the 1989 data is nonetheless a much closer match to Lotka’s model than was the case in 1988. The best-fitting exponent for the 1989 dataset is 2.73 whereas the equivalent figure for 1988 was 5.1. (Lotka suggests that this figure is normally close to 2).

The data summarised so far provides a fairly superficial summary of what was happening in L2 vocabulary research in 1989. Of course, the mere fact that a paper was published does not necessarily mean that it had any influence on the way researchers were thinking about vocabulary at the time. Indeed, many of the papers in the 1989 data set appear to have disappeared without trace, as they have no recorded citations in Google Scholar. An extreme example of this can be seen in Tickoo’s edited collection of papers: this volume accounts for nearly a tenth of the publications that appeared in 1989, but its long-term influence is relatively slight, and does not reflect its prominence in the 1989 data set.

## 2.2. *The analysis*

In order to get beyond these superficial characteristics of the 1989 data set, we need to look in more detail at the ideas that are influencing the published research. Our first step in this analysis is to identify the important sources whose work is being cited in the 1989 data set. To do this, the citation data for 1989 was extracted following the procedure described in my earlier papers, and summarised in Appendix 1. This analysis identified a total of 1911 sources – a considerable increase on the figure of the 1391 sources that we identified in the 1988 data. As usual, most of these sources are cited only once, but a small number of sources are cited much more than this. The data is summarised in Table 4.

**Table 4:** *The number of times sources are cited in the 1989 data set.*

|                  |           |           |           |           |           |           |           |           |           |           |           |           |
|------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| <b>frequency</b> | <b>24</b> | <b>23</b> | <b>22</b> | <b>21</b> | <b>20</b> | <b>19</b> | <b>18</b> | <b>17</b> | <b>16</b> | <b>15</b> | <b>14</b> | <b>13</b> |
| <b>cases</b>     |           |           |           | 1         |           |           |           |           |           |           | 4         | 3         |
| <b>frequency</b> | <b>12</b> | <b>11</b> | <b>10</b> | <b>9</b>  | <b>8</b>  | <b>7</b>  | <b>6</b>  | <b>5</b>  | <b>4</b>  | <b>3</b>  | <b>2</b>  | <b>1</b>  |
| <b>cases</b>     | 2         | 1         | 2         | 10        | 5         | 13        | 14        | 25        | 54        | 117       | 254       | 1406      |

The most cited sources in this data set are Nation (21 citations), Faerch, Ilson, Kasper and Sinclair (14), Carter, Hornby and Meara (13), and Laufer and Levenston (12). The main point to note here is the emergence of Nation as by far and away the most cited source. His 21 citations mean that his work is cited in 18% of the studies included in the data set. The next most cited sources manage to be cited in only a much smaller percentage of the data set (11%). It is worth pointing out here that only two *outputs* are recorded for Nation in the 1989 data set – another example of influence not being well reflected by output volume.

Once again, we have a lot of change in the list of most-cited sources. There are six new entrants in the 1989 Most Cited Sources List (Faerch, Ilson, Kasper, Hornby, Laufer and Levenston). These sources replace Channell, Krashen, Renouf, McCarthy and Richards – all major figures in the 1988 data set. Only four sources appear in both the 1988 and the 1989 lists: Carter, Meara, Nation and Sinclair.

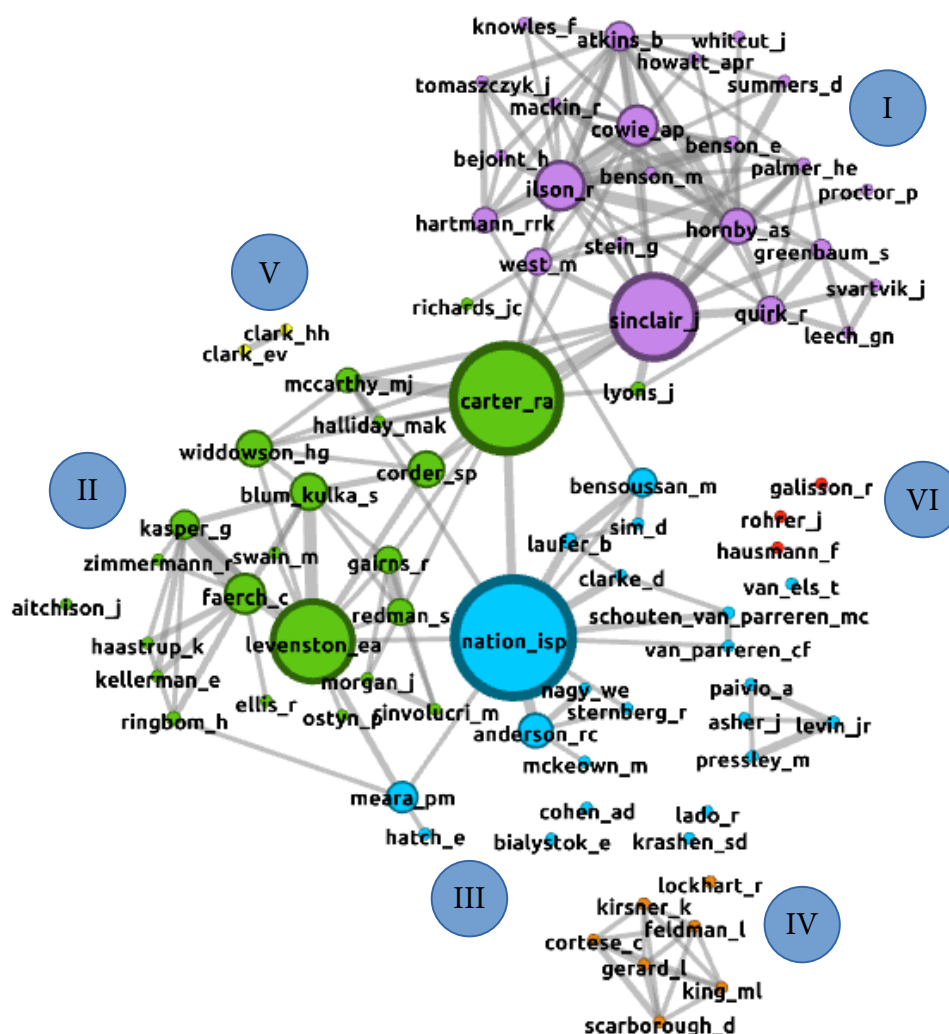
The next step in our analysis of the 1989 data set is to produce a co-citation map of the data set. To do this, we identify a large set of sources – conventionally about 100 – and compute the number of times these sources are cited together in the papers that make up the data set. Some arbitrary judgements are sometimes necessary here: in the 1989 data set, 134 sources are cited at least four times, but this figure is rather higher than we would like as it is a lot higher than the conventional figure of 100 sources. Only 80 sources are cited five times or more in the 1989 data set, somewhat fewer than we would like. However, given that the maps in last year's paper were based on the 82 most cited sources with an inclusion threshold of four citations, the best choice here seems to be to adopt an inclusion threshold of five citations. With this threshold, a total of 80 cases qualify for inclusion – a figure which is very close to number of sources that appeared in our map of the 1988 research. In percentage terms, the inclusion threshold of 5 citations is almost exactly the same as last year's inclusion threshold of four citations: in both cases, sources are included in the analysis if they are cited in just over 4% of all the papers included in the data set.

The co-citation data for the 80 most cited sources in the 1989 data set were extracted from the complete data set (all 1911 sources cited in 1989), and the results analysed using the Gephi software package (Bastian, Heymann and Jacomy: 2009). The results of this analysis are reported in Figure 1. Figure 1 looks a little different from the maps I presented in my earlier papers, as I have slightly changed my way of working in order to simplify the additional complexity that arises as the maps get bigger. The clustering in Figure 1 is based on all the co-citation links between the 80 nodes that appear at least twice in the data set. All this data is used to identify the co-citation clusters between the sources. When it comes to drawing the maps, however, I have ignored all the co-citation links which appear fewer than four times in the data set in the interests of clarity. Each co-citation link in the map thus occurs at least four times in the data set. This process loses some data (weak co-citation links that appear only rarely in the

data set), but it results in a clear set of clusters, with only a small number of detached nodes, and each of these detached nodes can be associated with a specific cluster in the map.

The 1989 map is a fairly easy to interpret.

The largest cluster, **Cluster I**, at the top of the map, is mainly composed of British linguists who work on dictionaries, corpora, word lists, and similar tools which inform the research on vocabulary. The outstanding source here is Sinclair, already identified as a significant source in the 1988 map, but here consolidating his position. The co-citation links in this cluster are surprisingly dense, but the cluster is only weakly linked to the rest of the map, mainly through Cluster II. Most of the sources in this cluster are descriptive linguists, but the cluster also includes a group of sources whose main interest is how L2 learners use dictionaries.



**Figure 1:** A co-citation analysis of the 1989 data set. There are 80 nodes in this map; links appearing fewer than 4 times in the data set have been eliminated. Nodes are sized according to their betweenness centrality. See the discussion in the text.

**Cluster II**, on the western edge of the map, is rather more difficult to describe succinctly. It consists of several sub-clusters. Carter and McCarthy need to be considered in conjunction with the corpus and dictionary sources we found in Cluster I. Halliday, Widdowson and Corder are a group of influential British applied linguists, who also have close co-citation links with the

sources in Cluster I, but are not themselves L2 vocabulary researchers. Gairns and Redman, Morgan and Rinvulcri and Ostyn are all authors of vocabulary teaching materials. Ellis and Aitchison are textbook sources: Ellis is the author of the main textbook on Second Language Acquisition at this time (Ellis 1985), while Aitchison (1987) was influential in popularising some psychological ideas. Swain is mainly cited by North American researchers in connection with an important paper on communicative competence (Canale and Swain: 1980). The remaining sources in this cluster are representative of an important strand of empirical research in L2 vocabulary acquisition, as opposed to the corpus based approaches that we identified in Cluster I.

**Cluster III**, at the centre of the map is also difficult to describe succinctly. The key source in this cluster is Nation: the map identifies him as the most central node in this year's data set. Most of the sources in this cluster are concerned with reading in an L2 and the acquisition of L2 words through reading. Sternberg, Nagy, McKeown and Anderson are L1 reading researchers. The cluster also contains a detached sub-cluster consisting of four sources: Asher, Paivio, Pressley and Levin. Asher is best known for his work on the Total Physical Response approach to vocabulary learning. This work appeared in the 1960s (e.g. Asher: 1969), and it is surprising to find it resurfacing here. Pressley and Levin are the main researchers behind the keyword mnemonic method. This work too appeared much earlier than 1989 (e.g. Pressley: 1977), so it is possible that this sub-cluster is mainly a historical one. Paivio's work on memory, dual-coding and L2 vocabulary acquisition is current at this time, however. (e.g. Paivio, Clark and Lambert: 1988).

These three clusters make up the main core of the 1989 map, but the analysis also identifies three small, detached clusters.

**Cluster IV**, at the bottom of the map, is a group of psychologists who mainly work on word recognition in bilinguals. This cluster seems to be remains of the very large psycholinguistics clusters that we identified in earlier maps.

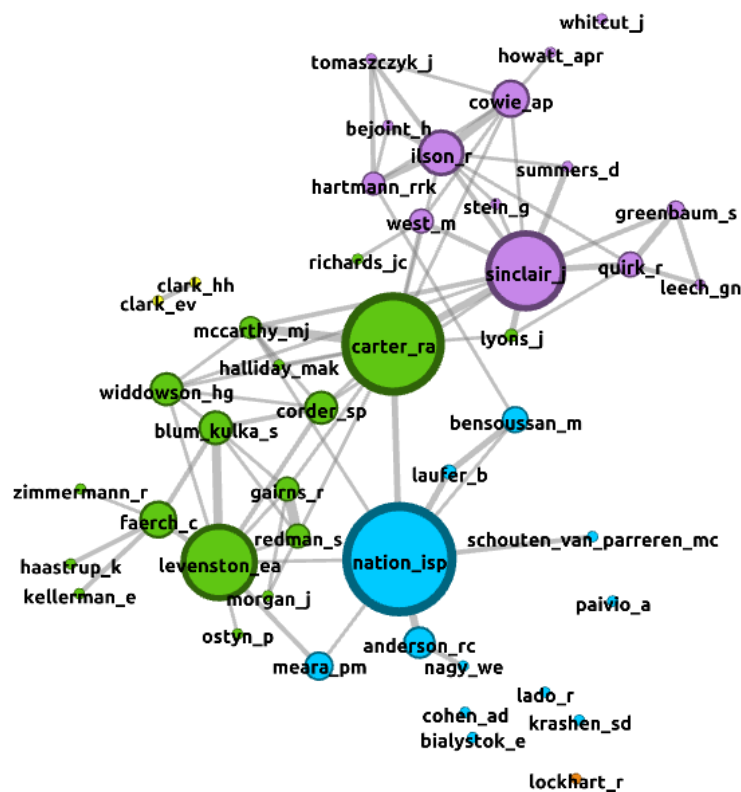
The small **Cluster V**, on the western edge of the map, contains only two sources, Eve Clark and Herb Clark, both child language development researchers. These two sources will be familiar from our earlier maps.

The three sources who make up **Cluster VI** seem to be linked primarily because all three are cited frequently in the special issue of the French language journal *Le français dans le monde*.

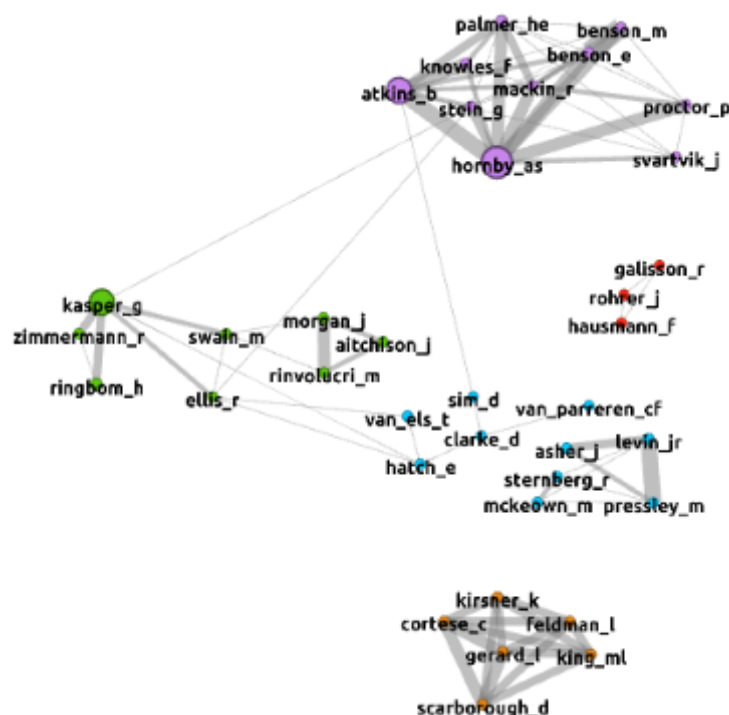
The central sources in this map are Nation, Carter and Sinclair and Levenston. Sinclair and Nation were both identified as Significant Influences in 1988, but they have consolidated their position here, clearly dominating the 1989 map. Carter and Levenston had both appeared in the 1988 map, but here they play much more important roles. Two Significant Influences from 1988 no longer have a central role in 1989: Renouf does not figure at all in the 1989 map, while Meara plays a much reduced role in Cluster III.

As usual, there is quite a lot of turn-over in the sources, and this feature is summarised in Figure 2 and Figure 3, which show the "survivor" sources from 1988 and the new sources in 1989 respectively.

Figure 2 shows that fewer than half of the 80 sources in the 1989 data set also appeared in the 1988 data set. The main outline of the 1988 map is clearly still visible in the 1989 data, but some consolidation seems to be taking place.



**Figure 2:** The “survivors”: sources who appear in both the 1988 and 1989 maps. Links which occur fewer than 4 times have been removed in the interests of simplification.



**Figure 3:** The new entrants in the 1989 map.

Figure 3, which identifies the new entrants in the 1989 map, shows that consolidation is not the only thing that is going on: the changes noted in 1989 are not evenly distributed. There are 38 new sources, slightly fewer than the new sources in 1988. A small handful of these sources

have appeared in earlier analyses, but most of them are genuinely new. The figure suggests that new sources are appearing in all of the clusters. However, most of the consolidation is taking place in Cluster I, the dictionaries and corpus linguistics cluster, where there has been a very large increase in the number of new sources being co-cited, and a significant strengthening of the existing co-citation links. The new entrants in the other clusters may be an indication that their research focuses are shifting a little. The new entries in Cluster III – Clark, Sternberg, McKeown and van Parreren – focus this cluster more obviously on L2 reading and inferencing behaviour. Cluster IV, which also formed part of the main L2 cluster in 1989, now appears as a group of sources who are mainly concerned with vocabulary theory and empirical studies of L2 acquisition. The Scandinavian research group that we have identified in our previous maps – Faerch, Kasper, Haastrup, Ringbom – form an important sub-cluster here. However, more than half of the sources in this cluster are new, making it the most volatile of the larger clusters in the 1989 map. Cluster V and Cluster VI are almost entirely made up of new entries.

### 3. Part 2. A wider perspective 1985–89

In this section, we will place the 1989 data into a larger context by looking at a five-year window covering all the research published between 1985 and 1989. This larger window smooths out some of the ad hoc fluctuations in the annual data, and it allows us to identify important sources whose influence is more than just a transitory one.

Table 5 summarises the main characteristics of the 1984–1988 window that we analysed in Meara (2020a). Table 6 summarises the new window covering 1985–89. Obviously, there is a lot of overlap between the two data sets, since both data sets include works published between 1985 and 1988.

**Table 5:** *The main characteristics of the 1984–88 data set.*

|                                                                       |                         |
|-----------------------------------------------------------------------|-------------------------|
| Number of papers in the data set:                                     | 403                     |
| Number of authors contributing to the data set:                       | 375                     |
| Number of sources cited in the data set:                              | 4080                    |
| Inclusion threshold for this data set                                 | 12 citations            |
| Number of cited sources meeting the inclusion threshold               | 100                     |
| <b>Identifiable co-citation clusters</b>                              | 7+3 detached singletons |
| I: vocabulary acquisition (31)                                        |                         |
| II: word recognition in an L2, performance of bilingual speakers (26) |                         |
| III: meaning, corpus analysis (22)                                    |                         |
| IV: word frequency counts (5)                                         |                         |
| V: Français Fondamental (5)                                           |                         |
| VI: Dictionary research (5)                                           |                         |
| VII: Dutch research (3)                                               |                         |

In Table 6, however, the 1984 data has dropped out of our five year window, while the newer 1989 data has been added to the window, and because 1984 was a fairly lean year for vocabulary research while 1989 was a very productive year, the figures in Table 6 are quite different from those in Table reported in Table 5.

The number of papers in the data set has increased by about 18%. The number of authors contributing to the data set has increased by 27%. The number of cited sources in the data set has increased by 13%. These figures all signal a substantial and steady increase in the amount of research available. This increase is also reflected in the increased threshold that we are using to identify the most significant sources in the data set. This figure too is steadily creeping upwards.

**Table 6:** *The main characteristics of the 1985–89 data set.*

|                                                                       |                        |
|-----------------------------------------------------------------------|------------------------|
| Number of papers in the data set:                                     | 477                    |
| Number of authors contributing to the data set:                       | 475                    |
| Number of sources cited in the data set:                              | 4616                   |
| Inclusion threshold for this data set                                 | 14 citations           |
| Number of cited sources meeting the inclusion threshold               | 103                    |
| <b>Identifiable co-citation clusters</b>                              | 6+1 detached singleton |
| I: vocabulary acquisition (26)                                        |                        |
| II: dictionaries and corpus analysis (25)                             |                        |
| III: reading, meaning and inferencing (23)                            |                        |
| IV: word recognition in an L2, performance of bilingual speakers (19) |                        |
| V: mental imagery (5)                                                 |                        |
| VI: Français Fondamental (5)                                          |                        |

The main driver here is the number of papers being published: in both Tables, the threshold for inclusion is just under 3% of all the papers published in their respective windows.

**Table 7:** *The number of authors contributing to N papers in the 1985–89 data set, and the expected number of authors based on Lotka's Law.*

| Papers  | 16 | 15 | 14 | 13 | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5  | 4  | 3  | 2  | 1   |
|---------|----|----|----|----|----|----|----|---|---|---|---|----|----|----|----|-----|
| Authors | 1  |    |    |    |    | 1  | 1  | 1 | 2 | 2 | 1 | 5  | 11 | 21 | 55 | 314 |
| Lotka:  | 1  | 1  | 1  | 2  | 2  | 3  | 3  | 4 | 5 | 6 | 7 | 13 | 20 | 34 | 72 | 314 |

Table 7 shows the distribution of authors contributions to this new data set. The number of authors contributing five or more papers to the data set (i.e. at least one paper per year) has increased slightly, moving from 10 in 1984–88 to 13 in the 1985–89 data set. However, the list of prolific authors has not changed substantially. Meara (16 papers), Zimmerman (11), Palmberg (10), Carter (9), Broeder and Laufer (8), Extra and van Hout (7), and Beheydt (5) all appeared in the 1984–88 prolific authors list. Nation (6 papers), Alfes, Cohen, McCarthy and Robinson (5 papers each) are all new entrants to this list. Arnaud (with 4 relevant papers) drops out of the list – though it is worth noting that a contribution of four papers in a five year window would have counted as a major achievement only a few years earlier.

The overall increase in outputs is reflected in the number of authors who contribute to just a single paper. In the 1985–89 window, 314 authors fell into this category, an increase of 9%. We have noted

in our earlier analyses that the field as a whole is characterised by an unusually large number of authors who make just one contribution to the data set, and this continues to be the case in 1985–89. The bottom line of Table 7 shows how far the data remains abnormally

skeweds. This data line shows the number of authors we might expect to be making *N* contributions to the data set given that the number of authors making a single contribution is 314. The model developed by Lotka (1926) suggests that we would expect to get substantially more authors making two, three, four and more contributions to the data set – so for example, Lotka’s model suggests that we could expect to find 20 authors contributing 4 papers, whereas in fact we have only 11 in this data set.

As always, we need to note that publishing a paper is not a sufficient condition for its author to become a significant source in the data set. The real test of this comes from the extent to which papers are actually cited, and we turn to this feature in Table 8. This table summarises the number of times each source is cited in the 1985–89 data set. As usual, most sources are cited only once in the complete set of 477 papers – 2957 sources fall into this category, 64% of the total. But a number of sources are cited much more frequently than this. The most frequently cited sources in the 1985–89 data set are Meara (cited in 61 papers), Nation (55), Krashen (49), Levenston (47), JC Richards

**Table 8:** *The number of times sources are cited in the 1989 data set.*

| FREQUENCY | 60+ | 59 | 58 | 57 | 56 | 55 | 54 | 53 | 52 | 51 | 50  | 49  | 48  | 47  | 46   |
|-----------|-----|----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|------|
| CASES     | 1   |    |    |    |    | 1  |    |    |    |    |     | 1   |     | 1   | 1    |
| FREQUENCY | 45  | 44 | 43 | 42 | 41 | 40 | 39 | 38 | 37 | 36 | 35  | 34  | 33  | 32  | 31   |
| CASES     | 1   |    |    |    |    |    |    |    |    |    |     | 4   | 1   |     | 3    |
| FREQUENCY | 30  | 29 | 28 | 27 | 26 | 25 | 24 | 23 | 22 | 21 | 20  | 19  | 18  | 17  | 16   |
| CASES     |     | 3  | 2  | 2  | 3  | 3  | 4  | 2  | 4  | 2  | 4   | 5   | 8   | 13  | 8    |
| FREQUENCY | 15  | 14 | 13 | 12 | 11 | 10 | 9  | 8  | 7  | 6  | 5   | 4   | 3   | 2   | 1    |
| CASES     | 12  | 14 | 12 | 15 | 28 | 27 | 33 | 31 | 62 | 73 | 112 | 186 | 321 | 656 | 2957 |

(46), Faerch (45), AD Cohen and Kellerman (both with 34 citations), West (33), Blum-Kulka, Corder and Schouten-van Parreren (each cited in 31 papers). Again, this list is almost identical to the equivalent 1984–88 list: West, Schouten-van Parreren and Blum-Kulka are new entries. Nation, Levenston and Faerch have all substantially increased their citation counts.

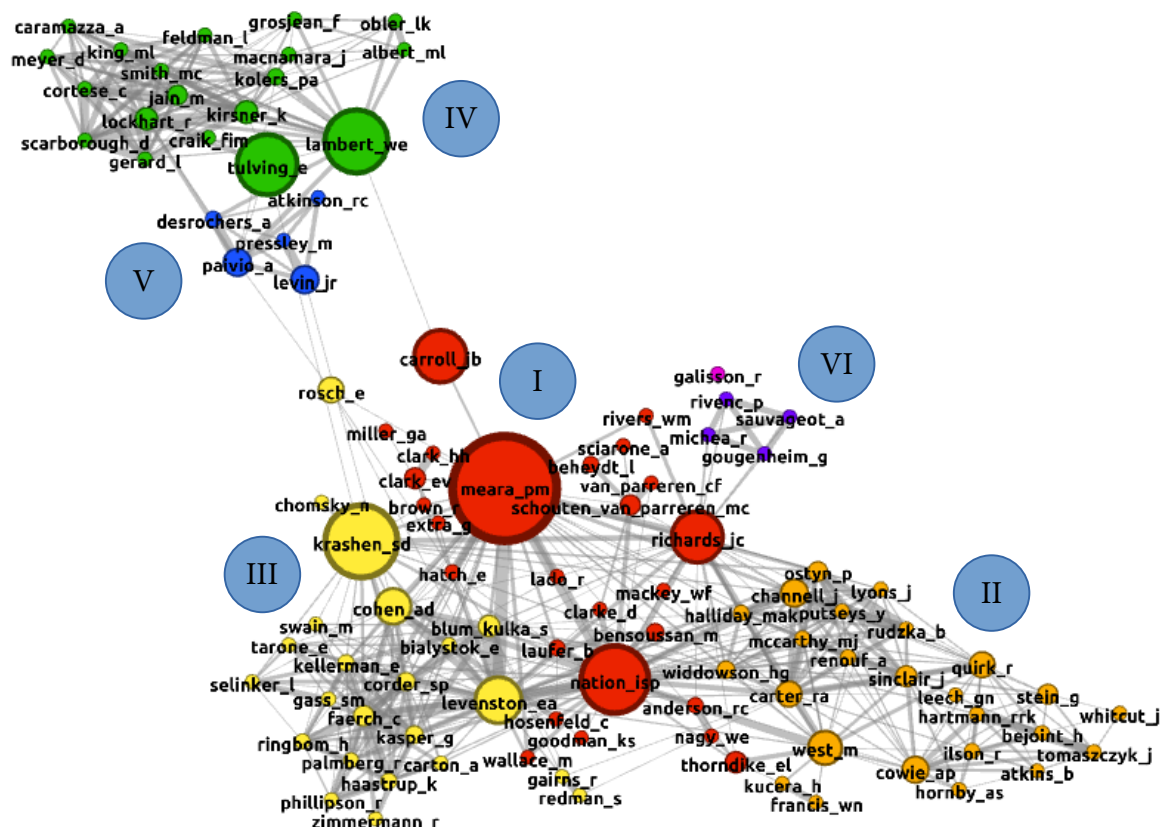
This distribution suggests that the 1985–89 maps should be based on the 103 sources who are cited fourteen times or more. This figure is very close to last year’s figure of 100 sources cited at least 12 times. Accordingly, we can set a threshold of 14 citations for a source to be included in the analysis that follows.

Figure 4 shows the basic map for the 1985–89 data set. Gephi identifies six clusters in this data set, and one detached singleton (Galisson).

**Cluster I**, in the centre of the map, is dominated by Meara, Nation and Richards, and is clearly identifiable as the mainstream vocabulary research cluster. Sources in this cluster are strongly co-cited with sources in Cluster II and Cluster III. Cluster I seems to have absorbed a small Dutch research cluster that we identified in 1984–88, suggesting that the Dutch research is becoming better integrated into a wider research program than it had been.

**Cluster II** at the Southeast corner of the map is a dictionaries and corpora cluster. This cluster was also present in the 1984–88 map, but here it has quadrupled in size. The cluster has a dense internal structure, but only a small number of its members are co-cited outside of the cluster.

Cluster III lies at the Southwest corner of the map. I think the defining characteristic of this cluster is a concern with the way L2 speakers acquire the meaning of new words (mainly through reading). The cluster is dominated by Krashen, but it also contains a large sub-cluster of Scandinavian vocabulary researchers whose main focus is inferencing. Transfer between L1 and L2 is also a theme here.



**Figure 4:** Co-citation analysis of the 1985–89 data set. Nodes are cited at least fourteen times in the data set. Links that appear fewer than 6 times have been eliminated in the interests of simplicity.

**Cluster IV**, at the Northwest corner of the map is clearly a psycholinguistics cluster, with a focus on word handling by bilinguals. This cluster has few co-citation links with the other clusters in the map. The weak links between Lambert and Carroll, and Tulving and Krashen contribute strongly to the high betweenness centrality scores recorded for these sources. This cluster also had a presence in the 1984–88 map, but here it appears to have shrunk considerably. Whether this indicates a general decline in psycholinguistic studies of L2 word handling is unclear. It may be an artefact of the unexpectedly large growth in dictionary research (Cluster II), which typically does not cite much psychological research. The dense co-citation links in Cluster IV are largely a function of authoring practice in psychology papers: these tend to have multiple authors who contribute to two or three papers in the period under analysis. This practice is rather different from what happens in the applied linguistics tradition, and may cause the real importance of these sources to be overstated.

**Cluster V**, just below cluster IV, is another psycholinguistics cluster. This one is focusses on mental imagery and mnemonic approaches to vocabulary acquisition, and is generally more

concerned with practical applications of the research for vocabulary instruction. The striking thing here is that this cluster has a few co-citation links with Cluster IV, but only very tenuous links with Cluster III, and no links at all with Cluster I or Cluster II. Mental imagery did appear in the 1984–88 map, but there it was rather more closely integrated with the mainstream L2 vocabulary research. Here, it looks to be in the process of becoming more detached from the L2 vocabulary sources: basically, L2 vocabulary research by applied linguistics obstinately refuses to pay any attention to the interesting work that psychologists are doing in this area.

**Cluster VI** is the familiar *Français Fondamental* cluster that we have identified in our earlier maps, with Galisson as a detached source. Like Cluster V, this work is only rarely co-cited along-side other sources in the map. The co-citation links with Richards prevent this group of sources from becoming a detached cluster, and it is clear that this work is not well integrated into the rest of the network.

Overall, this map shows that the main changes in this new five-year window are changes in emphasis, rather than changes of direction. Of the 103 sources in this map, 85% were also identified as important sources in the 1984–88 map. There are only 19 new sources. Six new sources are found in cluster I (Clarke, Goodman, Hatch, Nagy, van Parreren and Wallace), Goodman and Nagy are important L1 reading sources. Cluster II merges two separate cluster identified in the 1984–88 map, and also contains six new entrants (Atkins, Cowie, Hornby, Leech, Stein and Whitcut), indicating that the dictionary research that we noted last year is becoming significantly more important. Cluster III hosts four new entrants: Gairns and Redman authored a series of vocabulary textbooks; Swain is mainly cited for her early work on *communicative competence*. Zimmerman, one of this year's prolific authors, is the first German source to appear in our five-year window maps, and this may indicate that Cluster III reflects an upsurge in European vocabulary research. Cluster IV has one new source, King, a co-author of Kirsner's. RC Atkinson and JR Levin both strengthen the mnemonics and image cluster, cluster V. Cluster VI has no new members.

#### 4. Discussion

Four themes seem to emerge from this year's maps.

The first theme is that 1989 is mainly about consolidation. There are some changes from the 1988 maps, but most of these changes are marginal – a few new additions to the list of significant sources, and some changes in the relative importance of individual sources. On the whole, however, we have a remarkably stable structure in the pattern of co-citations, and this is all the more striking when we consider how volatile the maps were only few years previously. Of course, we are still a long way away from the *first paradigm* scenario described in Meara (2020b), but many of the components of this first paradigm are already in place in the 1989 maps.

The second theme that characterises the 1989 data is the emergence of Nation as Most Significant Source in the 1989 map. Conventional wisdom has it that “modern” L2 vocabulary research really takes off around 1992, shortly after the first publication in 1990 Nation's seminal textbook *Teaching and Learning Vocabulary* (Nation: 1990). Early drafts of Nation's book had

already been circulating among researchers for some time (cf. Nation: 1983), but were not really well-known, and they are only rarely cited in the research literature. However, Nation's prominence in the 1989 map suggests that his influence on vocabulary research has already become pervasive before the publication of the 1990 volume. So far, however, citations of Nation's work are selective: he has extensive co-citations within Cluster I, Cluster II and Cluster III, but he is only rarely co-cited with sources in the dictionary research cluster, and never co-cited with sources in the psycholinguistics cluster.

At first glance, the 1985–1989 map suggests that Meara is the Most Significant Source in the 1984–1989 window, but I think that this is a misleading conclusion. Meara *was* important in the 1988 data set, but by 1989 he plays only a minor role in Cluster III. The reason he appears to be important in the 1985–89 map is that he is one of the few vocabulary researchers who is co-cited along with the sources in psycholinguistics cluster, Cluster IV. Meara is co-cited with psycholinguists because he had been arguing for a long time that applied linguists researching vocabulary seemed to be unaware of the importance of this research strand for their own research. For example, he frequently cites the keyword mnemonic research of Pressley and Levin as research that applied linguists ought to pay attention to, and this generates many co-citation links between Pressley, Levin and their co-workers and his own work. However, this is not a reciprocal relationship: Meara cites Pressley and Levin alongside his own work, but Pressley and Levin don't cite him at all. The maps don't show this bias in the co-citations. Clearly, Meara is not as central a source as the map suggests.

So far in these maps, I have been using betweenness centrality as way of teasing out the importance of individual sources. However, a number of correspondents have pointed out to me that it might not be appropriate to use this measure in the context of vocabulary research. The argument seems to be that betweenness centrality is an important feature of networks in which things actually move between nodes, but it is not obvious that this metaphor works where there is no flow involved. For example, in a network that describes traffic flow, a node which has a high betweenness centrality value is critical to the way the network functions: we are talking real physical properties here. But when, in Figure 8, we identify JB Carroll as the only node that connects Cluster IV to the rest of the network, it is not at all clear what this actually means. A co-citation network is more of an abstract metaphor than a physical description. The individual sources do not depend on each other in quite the same way as intersections in a road network do. The betweenness centrality measure captures the fact that dictionary researchers do not routinely cite psycholinguists and vice-versa, but we could argue that this does not assign any real significance to JB Carroll. The important point here is that dictionary writers only rarely see the need for psycholinguistics, not that JB Carroll provides a link when they do. Clearly, this use of the betweenness centrality feature needs some further thought, and I will be reviewing this practice in future analyses.

The final feature worth commenting on is the role of regional research networks in the 1989 vocabulary research. In 1988, I noted the increased importance of dictionaries and corpora in the vocabulary research, and I drew attention to the fact that this was largely a British initiative. Dictionary research continues to be important into 1989, and it continues to be largely a British interest. Cluster I in Figure 1 has doubled in size from 1988. However, all but one of the new entries in this cluster are British researchers – and the exception, Jan Svartvik,

professor of English at Lund University in Sweden – is closely associated with Quirk's research group at University College, London. In contrast, about half of the sources in Cluster II are British researchers, but most of the empirical work here is done by a group of northern European researchers. Cluster III contains a single British researcher. The remaining clusters are mainly focussed on North America. However, change is looming on the horizon. By 1989 researchers were beginning to have access to early versions of the World Wide Web, and we can expect these technological developments to have a marked impact on L2 vocabulary research in the years to come.

## 5. Conclusion

The analyses in this paper have suggested that the L2 vocabulary research that appeared between 1985 and 1989 continues to show a fairly stable structure, but there are signs that some large-scale changes may be in the offing. In particular, the emergence of Paul Nation as the central source in the 1989 map marks the beginning of an important shift in L2 vocabulary research, and we can expect Nation's influence to grow in the immediate future. Some historically important figures still appear as influences in the five-year window map, but their influence in the 1989 map is greatly reduced, hinting that the vocabulary research is perhaps becoming more independent, and more self-sufficient in its use of sources. The next paper in this series will examine the research published in 1990 in the context of a five-year window covering 1986–90. 1990 is an important year for L2 vocabulary research in that it sees the publication of Nation's seminal book (1990) a critical milestone in L2 vocabulary research. It will probably be too soon for the influence of this work to make its presence felt in the bibliometric maps, but it is difficult to avoid the feeling that in 1990 some major new developments lie just around the corner.

## References

- Aitchison, J. 1987. *Words in the mind: An introduction to the mental lexicon*. Oxford.
- Asher, J. J. 1969. The Total Physical Response approach to second language learning. *Modern Language Journal* 53: 3–17.
- Bastian, M., Heymann, S., and Jacomy, M. 2009. Gephi: An open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media.
- Canale, M., and M. Swain. 1980. Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* 1: 1–47.
- da Solla Price, D. 1965. Networks of scientific papers. *Science* 49(3683): 510–515.
- Ellis, R. 1985. *Understanding second language acquisition*. Oxford.
- Haastrup, K. 1989. *Lexical inferencing procedures or talking about words*. PhD thesis: University of Copenhagen.
- Hammer, P., and G. S. Giauque. 1989. *The role of cognates in the teaching of French*. New York.
- Helliwell, M. 1989. *Can I become a beefsteak? Trügerische Wörter zum Nachschlagen und Üben*. Berlin.
- Hornby, A. S. (ed.) 1989. *Oxford Advanced Learner's Dictionary of Current English*. Oxford. 4<sup>th</sup> edition.
- James, G. (ed.) 1989. *Lexicographers and works*. *Exeter Linguistic Studies* 14. Exeter.
- Lotka, A. J. 1926. The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences* 16(12): 317–324.
- Marello, C. 1989. *Dizionari bilingui*. Bologna: Zanichelli.
- Meara, P. M. 2012. The bibliometrics of vocabulary acquisition: An exploratory study. *RELC Journal* 43(1): 7–22.

- Meara, P. M. 2014. Life before Nation: Bibliometrics and L2 vocabulary studies in 1982. In: M. Gómez González, F. J. Ruiz de Mendoza Ibañez, F. González García and A. Downing (eds.), *The Functional Perspective on Language and Discourse: Applications and implications*, 111–129. Amsterdam..
- Meara, P. M. 2015. Vocabulary research in 1983: A bibliometric analysis. *Linguistics Beyond and Within* 1: 187–198.
- Meara, P. M. 2016. Two steps backwards: A bibliometric analysis of L2 vocabulary research in 1984. *Linguistics Beyond and Within* 2: 139–152.
- Meara, P. M. 2017. A new beginning? A bibliometric analysis of L2 vocabulary research in 1985. *Linguistics Beyond and Within* 3: 136–154.
- Meara, P. M. 2018. Laying the foundations. A bibliometric analysis of vocabulary research in 1982–86. *Linguistics Beyond and Within* 4: 108–128.
- Meara, P. M. 2019. Steady as she goes? A bibliometric analysis of L2 vocabulary research in 1987. *Linguistics Beyond and Within* 5: 103–119.
- Meara, P. M. 2020a. The British are coming...! A bibliometric analysis of L2 vocabulary research in 1989. *Linguistics Beyond And Within* 6: 111–129
- Meara, P. M. 2020b. The emergence of a *First Paradigm* in vocabulary research: The bibliometrics of *System. Vocabulary Learning and Instruction* 9(1): 1–32.
- Meara, P. M. (no date) VARGA: The Vocabulary Acquisition Research Group Archive. Available at: <https://www.lognostics.co.uk/varga/> (accessed 25th October 2021).
- Melka Teichroew, F. J. 1989. *Les notions de réception et de production dans le domaine lexicale et sémantique*. Berne.
- Nation, I. S. P. 1983. *Teaching and Learning Vocabulary*. Wellington, New Zealand.
- Nation, I. S. P. 1990. *Teaching and Learning Vocabulary*. Rowley, Mass.
- Paivio, A., J. Clark, and W. E. Lambert. 1988. Bilingual dual-coding theory and semantic repetition effects. *Journal of Experimental Psychology: Human Memory and Cognition* 14: 163–172.
- Pressley, M. 1977. Children's use of the keyword method to learn simple Spanish vocabulary words. *Journal of Educational Psychology* 69(5): 465–472.
- Service, E. 1989. *Phonological coding in working memory and foreign language learning*. University of Helsinki: General Psychology Monographs, B9.
- Small, H. 1973. Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science* 24: 265–269.
- Tickoo, M. L. (ed.) 1989. *Learners' dictionaries. State of the Art*. Singapore, SEAMEO Regional Language Centre.
- White, H. D., and B. C. Griffith. 1981. Author co-citation: A literature measure of intellectual structure. *Journal of the American Society for Information Science* 32: 163–171.
- Zagrebelsky, P. (ed.) 1989. *Dal Dizionario ai Dizionari. Orientamento e Ghida all'uso per Studenti di Lingua Inglese*. Turin, Tirrenia Stampatori.

## Acknowledgements

Thanks to Marlene Schwarz for help with sourcing the German literature, and to Anne Vermeer who kindly supplied me with copies of his Dutch language papers.

Thanks to Julia Mitchell for help with transcribing the citation data.

Special thanks to John Read for his insightful comments on the earlier papers in this series.

## Appendix 1: Co-citation analysis: The methodology

The co-citation method used in this paper was developed by Small in a number of papers published in the 1970s (e.g. Small: 1973). This approach, which was actually built on earlier bibliometric work by da Solla Price (1965), has been extensively used to analyse research in the

natural sciences (e.g. White and Griffith 1981) but does not seem to have been adopted as a standard tool by researchers in the Humanities.

The raw data for a co-citation analysis consists of a list of all the authors cited in the set of papers to be analysed. For each paper in the data set, we make a list of every author that the paper cites; for each paper, each cited author counts only once, regardless of how many times they are cited in the paper; and for a cited paper with multiple authors, each of the contributors is added to the author list. Self-citations, where an author cites their own work, are treated in the same way as any other citation, on the grounds that authors only rarely fail to cite their own work. This raw data is then used to construct a large matrix showing which authors are cited together in each of the papers in the data set. The matrix can then be analysed using a program such as Gephi (Bastian, Heymann and Jacomy 2009). Gephi performs a cluster analysis on the data, groups together authors who tend to be cited alongside each other in a number of papers, and outputs a map which shows the composition of the clusters and the relationship between them. The clusters are generally taken to represent “invisible colleges” in the data – i.e. groups of researchers who share similar reference points and a common research focus.

## Appendix 2: Lotka’s model

Lotka (1926) suggested that there might be a straightforward relationship between the number of authors who contribute a single paper to a field and the number of authors who make multiple contributions to the field. Suppose, for example, that we have 250 authors who make a single contribution to a data set, then it would be unusual to find only a single author making two contributions, and it would likewise be very unusual to find that a single author makes twenty contributions, while no other authors make more than one contribution to the data set. Lotka suggested that the expected relationship could be described as a power law:

$$E_N = T / N^x$$

where  $T$  is the total number of authors who contribute a single paper to the data set,

$N$  indicates 2,3,4,5... outputs,

and  $E_N$  is the expected number of authors contributing to  $N$  outputs.

In practice, the value of  $x$  (the exponent in Lotka’s formula) is usually around 2 – that is, a value of 2 for this exponent gives a fair approximation of what happens in real life. So, for a data set in which 250 authors contribute to just one paper in the data set Lotka’s model predicts that we can expect  $250/2^2 = 63$  authors who contribute to two papers in the data set,  $250/3^2 = 28$  authors who contribute three papers to the data set,  $250/4^2 = 16$  authors making four contributions to the data set, and so on as shown in the table below.

|                                  |    |   |   |   |   |    |    |    |    |     |
|----------------------------------|----|---|---|---|---|----|----|----|----|-----|
| <b>contributions</b>             | 10 | 9 | 8 | 7 | 6 | 5  | 4  | 3  | 2  | 1   |
| <b>Expected <math>E^N</math></b> | 2  | 3 | 4 | 5 | 7 | 10 | 16 | 28 | 63 | 250 |

Clearly, this model predicts that the number of papers an active researcher might be expected to produce falls off rather quickly. Empirical tests of what has become known as “Lotka’s Law”

do seem to work well. However, the model works best when we are dealing with well-established fields, and very large data sets. The single year data sets that I have discussed in this series of papers are not a close match to Lotka's expectations, but the larger 5-year data sets are generally a better fit to the power law model. In both cases, however, we get a much better fit when the value of  $N^x$  is raised above 2. For example, we get the best fit for the 1989 data when  $x = 2.73$ , though this figure needs to be treated with some caution because the data set is relatively small. Higher values of  $x$  seem to be typical of immature, highly volatile fields. Generally speaking, the exponent values we find for the L2 vocabulary research literature are higher than we would normally expect. I do not yet fully understand the implications of this.