

Steady as she goes? A bibliometric analysis of L2 vocabulary research in 1987

Paul Meara

Swansea University, the United Kingdom

Abstract

This paper uses a co-citation analysis to examine the research on L2 vocabulary acquisition that was published in 1987. This year is surprisingly volatile compared with the previous year, 1986, with a very large number of new sources appearing in the maps, and many sources identified in previous years losing their influence in the research papers that make up the 1987 data set. The paper also reports a larger analysis of all the work that appeared in a five year window from 1983–1987. This larger data set is not quite as volatile as the smaller 1987 data set, but it suggests that some sources who dominated the co-citation maps in earlier years are losing their influence.

Keywords: L2 vocabulary acquisition, vocabulary research, bibliometric analysis

1. Introduction

This paper is the seventh in a series of studies that attempt to plot the way research in L2 vocabulary acquisition has progressed over the last fifty years. Earlier papers in this series have analysed the research outputs in 1982, 1983, 1984, 1985, 1986 and 2006 (Meara 2012, 2014, 2015, 2016, 2017 and 2018). This paper is a sequel to my earlier *LingBaW* papers, in that it covers the research output of 1987. The paper also includes an analysis of research published in the five-year window of 1983–87.

Accordingly, this paper falls into two parts. Part I reviews the new research that appeared in 1987 in its own terms. Part 2 provides a wider context for this research, by summarising the main trends that appear in a five-year window covering 1983–1987. Both sections use the co-citation methodology which readers of *LingBaW* will now be familiar with. For new readers, the methodology is summarised in Appendix 1. The larger five-year window has a number of advantages compared with data from a single year. The most important advantage is that it smooths out low-level fluctuations in the data. For example, some sources do not always appear consistently in the annual accounts because their publication record is spasmodic and irregular. Sometimes this makes them appear as new sources, which, strictly speaking, they are not. Similarly, some sources garner a lot of co-citations as a result of a thematic issue of a journal, or

an edited volume consisting of a collection of linked papers, but this may not really reflect their true importance in the larger scale of things. Using a longer time-frame for a bibliometric analysis avoids both of these problems. However, it does introduce some problems of its own. The most important of these problems is that when the data is collected in a larger window of time, it becomes more difficult for a new source to be recognised. In this way, larger windows tend to emphasise the stable, conservative features of the field, rather than the innovations which are taking place. The point here is that a five-year window is not exactly comparable with a series of five single year windows, and readers need to take this into account when interpreting the maps that I will be presenting.

2. Part 1. The new research published in 1987

My analysis of the 1986 data (Meara 2018) suggested there were some marked differences between the 1986 research map and what we would expect to find in a bibliometric map covering a more recent period – any year since 2000, for example. The 1987 map (discussed below) exhibits broadly similar characteristics to the 1986 map, though a number of trends which are features of the post 2000 research are beginning to emerge. The 1987 data set shows a substantial increase in the number of publications dealing with L2 vocabulary acquisition, and this suggests that some sort of shift is taking place. The bibliometric maps should help us to understand what these shifts consist of.

2.1. The data sources

First, however, we should note the raw statistics of the 1987 data set. The VARGA database¹ (Meara n.d. accessed November 2019) identifies a total of 119 relevant sources – an increase of

¹ A number of readers have asked for more information about the **Vocabulary Acquisition Research Group Archive** database (VARGA). The data base is an open access resource which attempts to provide a comprehensive record of all the research on second language vocabulary acquisition. It currently contains about 10,000 entries.

The database was originally compiled from a series of annotated bibliographies that I published in 1982, 1987 and 1992. The development of the World Wide Web made it possible for all this data to be placed on-line and made searchable. From these small beginnings, the site has grown into a major resource, which currently logs about 1000 new entries each year.

The criteria for inclusion are not specific. I log any paper crossing my desk that appears to deal with how second language speakers learn/acquire/use/forget vocabulary, as long as it makes a theoretical claim about these processes. So for example, a paper suggesting that gap-filling exercises are a good way to practice new vocabulary, would probably not be included, but a paper that showed gap-filling exercises were a more effective way of learning vocabulary than a mnemonic method would be included. VARGA attempts to be comprehensive: I run a number of alerts on Google Scholar, and add to these as new terms become important. I systematically add new sources which are cited in the literature that crosses my desk. Each year, I scour the main research journals for work on L2 vocabulary, and supplement this with searches focussed on individual authors. Nonetheless, a number of features make this process less objective than I would like it to be. Not all the relevant papers are clearly marked as vocabulary research, and this makes them harder, though not impossible to locate. Early researchers, for example, tended to give their papers whimsical titles, whose meaning is not immediately apparent. Other vocabulary research is hidden inside papers which look at a number of related

some 20% over the total output in 1986. Particularly noteworthy in the 1987 output are four monographs (Carter, Koster, Ringbom and Wojtak & Herrmann), Meara's bibliography, and four collections of papers (Cowie, Carter & McCarthy, Diller et al. and a thematic issue of *Studies in Second Language Acquisition*).

A total of 110 eligible papers were identified for this analysis. (Monographs, theses and unpublished reports are conventionally excluded from bibliometric analyses, as their use of citations is rather different from what we find in normal research papers.) Eleven unobtainable items were omitted from the data set. These eleven items are listed in Appendix 2. The remaining 99 papers are referred to here as **the 1987 data set**. They are not listed here in full, but interested readers will be able to locate them in the VARGA database (<http://www.lognostics.co.uk>). Entering **1987 ##** into the search box will return a complete list.

These 99 papers involve a total of 111 authors. (Where a paper has more than one author, all the authors are counted as contributors.) As we have noted in previous reports, the list of authors is largely made up of authors who contribute to just a single paper (cf. Table 1). For 1987, there are 99 authors in this category – an increase of about 22% on 1986. Only 12 authors contribute to more than one paper in the data set. This figure is slightly lower than the equivalent figure for 1986. (We will refer to this group as the *1987 Prolific Authors*, though “prolific” is perhaps stretching a point here, and the term should be taken as a convenient shorthand label, rather than a literal description.) Meara and Zimmerman contribute to four papers each. Scholfield contributes to three papers. Nine authors contributed to two papers (Beauvillain, Beheydt, Carter, Gass, Grainger, Hausmann, Levenston, Palmberg and Singleton). It is worth noting that some of these authors also produced book length treatments in 1987, which perhaps suggests that their real contribution is being underestimated by the raw paper count. It is also worth noting that the list contains a number of new names – in fact, only Meara and Zimmerman appear in both the 1986 and 1987 prolific author lists: all the other prolific authors are new.

As in previous years, the data shows that there are fewer prolific authors than we might expect. Table 1 shows the pattern of contributions that we can expect when the number of authors of singleton papers is 99.

variables, of which vocabulary is just one. Sometimes, I have had to make a judgement call about whether a particular strand of research should be included. Should papers that deal with the writing of dictionaries for L2 learners be included? Yes, I think so. Should vocabulary acquisition in young bilingual children be included? Definitely. Should research on ESP word lists be included? Sometimes, but not automatically. The main strategic decision of this type that affects the content of the database is that I do not routinely include papers which deal with multi-word sequences or formulaic sequences. As far as research in the 1980s goes, I think this oversight can be excused, as very little work of this type was being undertaken. However, as more work on formulaic sequences is published in recent years, this constraint is becoming more of an issue.

A really serious problem concerns publications which appear in journals whose language of publication is not English. VARGA attempts to capture as much of this material as possible, and it has good coverage of vocabulary research published in Dutch, German and Spanish. Its coverage of Polish, French and Japanese research is much less satisfactory. I have correspondents who keep me informed of relevant research in these languages. Since 2015, there has been a huge surge in output from China, Korea and Iran that is not published in English. We are currently logging as much of this material as we can. Anyone who thinks they can contribute to this effort is welcome to get in touch with me.

Table 1: *The number of authors contributing N papers to the data set*

Number of papers (N)	5	4	3	2	1
<i>Actual data</i>		2	1	9	99
<i>Lotka's model N = 99</i>	4	6	11	25	99

The data for 1987 is highly divergent from the predictions made by Lotka's Law (Lotka 1926). Lotka's Law states that the number of prolific authors we would expect to find in a data set can be estimated from the number of authors contributing to a single paper. This expectation is normally expressed as:

$$E_N = T / N^2$$

where T is the number of authors contributing to one paper,

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

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

Thus, with 99 authors of just one paper, we would expect to find $99/2^2 = 25$ authors who contribute to 2 papers. The actual data falls well short of this prediction. (The data suggests that the distribution of the authors in the 1987 data set is better described as:

$$\text{authors of a single paper} / N^{4.38}$$

This figure is consistent with the data from earlier years, but its implications remain unclear.)

2.2. The analysis

The citation data from each of the papers in the 1987 data set were extracted in the usual way (see Appendix 1). This analysis identified 1587 unique sources – again a substantial rise on the 1986 figure, indicating that the range of sources is becoming more diverse as the field expands. The number of times each of these sources is cited in the data set is summarised in Table 2.

Table 2: *The number of times sources are cited in the 1987 data set.*

frequency	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
cases	1		1		2	1	2	6	5	6	5	28	42	86	240	1163

The most cited sources in this data set are Meara (16), Faerch (14), Kasper and Kellerman (12), Levenston (11), Hartmann and Nation (10) and six sources who are cited nine times each (Blum-Kulka, Carroll, Cowie, Haastrup, Lockhart and JC Richards). Compared with the 1986 data, the number of times these influential sources are being cited has increased. In 1986, for example, half of the top ten sources were cited only four times. In this data set, a source would need to be cited at least 9 times for them to be included in the top ten sources. Clearly, for a source to be counted as “significant” takes more in 1987 than it did in 1986.

Despite this “grade inflation”, the resulting distribution is relatively easy to work with. In the 1987 data set, 99 sources were cited at least four times, and this figure is very close to the conventional number of 100 often used in bibliometric studies of this kind. However, this inclusion threshold is slightly stricter than the threshold of three citations that we used in our

earlier studies. The fact that both the number of papers in the data set and the inclusion threshold have increased relative to 1986 means that sources going forward to the analysis have to be cited in about 4% of the publications in the data set. This is the same proportion that we used in the 1986 report. Despite the differences, then, the 1986 and 1987 data sets remain essentially comparable.

The citation data for the 99 sources were analysed using the Gephi software package (Bastian, Heymann and Jacomy 2009), and the results of this analysis are shown in Figure 1.

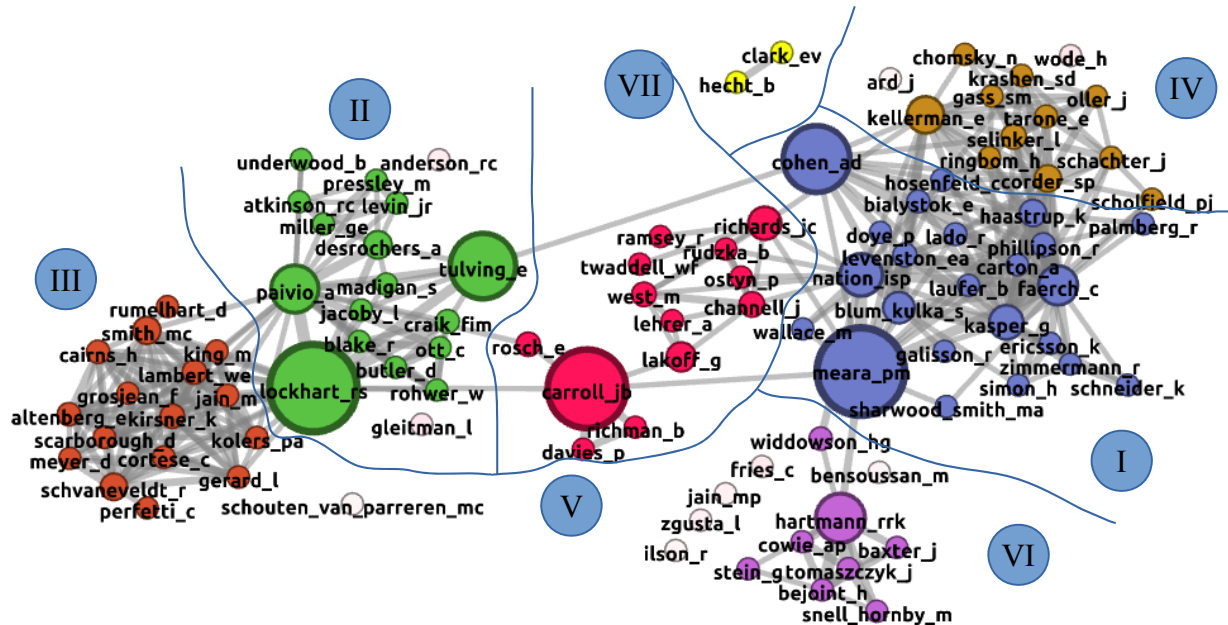


Figure 1: Co-citation analysis of the 1987 data. Each source is cited at least four times in the data set. Weaker edges that appear less than three times are eliminated in the interest of simplicity. Nodes are sized according to their betweenness centrality.

Gephi identifies six main clusters in this map.

The largest cluster, **Cluster I**, and **Cluster IV** on the Eastern edge of the map, seem to be the main-stream of L2 vocabulary research at this time. Cluster I is largely made up of European sources, with strong sub-clusters of Danish research (Faerch, Kasper, Haastrup), and Israeli research (Laufer, Levenston, Blum-Kulka, and Cohen). French language research is represented by Galisson, and German language research by Wode, Zimmerman and Schneider. In contrast, Cluster IV contains a high proportion of North American sources. Thematically, this cluster seems to be strongly oriented towards transfer research (Ringbom, Kellerman) and error analysis (Corder, Selinker and Scholfield). Co-citation links within these two clusters are strong, but there are few connections between these two clusters and other clusters in the map. The main structural feature that distinguishes the two clusters is that sources in cluster I are strongly co-cited with Meara and Nation, whereas the sources in Cluster IV generally are not co-cited in this way.

Cluster II and **Cluster III**, at the Western edge of the map, are mostly made up of psychologists rather than linguists – a feature that we have noted in our earlier maps. The sources in cluster II mostly work in the area of imagery and mnemonics. The sources in cluster III are mainly psychologists with an interest in bilingualism.

Cluster V, in the centre of the map, is a semantics and meaning group.

Cluster VI, at the bottom centre of the map, is a dictionaries and dictionary usage group.

Gephi identifies one very small cluster (Cluster VII) containing only two sources – Clark and Hecht co-authored a review paper that summarises recent work on children’s first language development.

Gephi also identifies nine detached singletons. These sources are frequently cited in the 1987 data set, but their co-citation links do not reach the threshold for inclusion in the map. Five of these singletons are associated with Cluster VI, suggesting that this cluster may actually be more important than it appears to be in figure 1.

The strongest links in this data set are to be found in Cluster I – the group consisting of Faerch, Kasper, Phillipson and Haastруп contains the strongest links in the entire data set.

Cluster VI has a few links with cluster 1, but no links with any other cluster.

Cluster III is strongly linked with cluster II, but has no links with other clusters.

Cluster II has loose connections with cluster V and cluster I, but members of this cluster are not co-cited with members of cluster IV or cluster VI.

The key sources in the map are Lockhart (linking cluster II and cluster III), Carroll (linking cluster II, cluster V and cluster I), Meara (linking cluster I and cluster VI), and Tulving and Cohen (providing a strong link between cluster I and cluster II).

The map exhibits some interesting, and perhaps unexpected features. It is tempting to see the map as a fairly straightforward development of the 1986 map, but the simple picture which emerges in 1987 hides a surprising amount of complexity. It is not immediately obvious that there has been a huge turn-over of sources between 1986 and 1987. However, all the sources with high betweenness centrality scores in the 1986 map (the larger nodes in the 1986 map) no longer hold this privileged position. Kucera and Francis, whose frequency counts provided an important link between the linguists and the psychologists have disappeared completely. Leech, whose work on corpus linguistics provided a similar bridge, has also disappeared. Krashen, the key figure in 1986, appears in 1987 as a relatively minor player in cluster II. Several new sources have emerged as pivotal sources who provide the co-citation links between the clusters: Meara, Kellerman, Tulving, Cohen, Lockhart and Carroll seem to play this role in the 1987 map. In addition, more than half of the significant sources who figured in the 1986 map fail to re-appear in the 1987 map, and some of these missing sources represent important losses. The psycholinguistic influences (clusters II and III) still make up a significant share of the sources being cited in 1987, but the structure of these clusters has undergone a major change. The single very dense psycholinguistics cluster that we identified in 1986 has dissolved into two separate clusters, with imagery and mnemonics sources slightly outnumbered by the remaining psycholinguists. More importantly, the majority of the psycholinguistics sources that appeared in the 1986 map are no longer found here. The European Science Foundation project, which we had flagged as a major development in 1986, is completely missing from the 1987 map. Also missing is any reference to the *Français Fondamental* project, which had been enjoying a revival in 1986. Suggestopedia is no longer a topic of interest. Sources who work on neurolinguistics and bilingualism seem to have disappeared. Two key sources in reading research (Goodman and F Smith) have also failed to make it into the list of significant influences. Fortunately, these losses are balanced by the emergence of some new centres of interest.

We can begin to judge the extent of this turnover by examining Figure 2 and Figure 3 which show the sources that appear in both the 1986 and the 1987 map (the survivors, for short) and the new entries to the 1987 map respectively. Overall, the survivors' map reflects two main features which we have consistently found in these studies: we have a group of psychologists whose work focusses rather narrowly on questions that can be addressed by laboratory studies, and a group of applied linguists whose work is methodologically more diverse. There are few strong links between these groupings, but very strong links within them. The survivors' map also allows us to identify sources whose influence is perhaps less important than it has been in previous years.

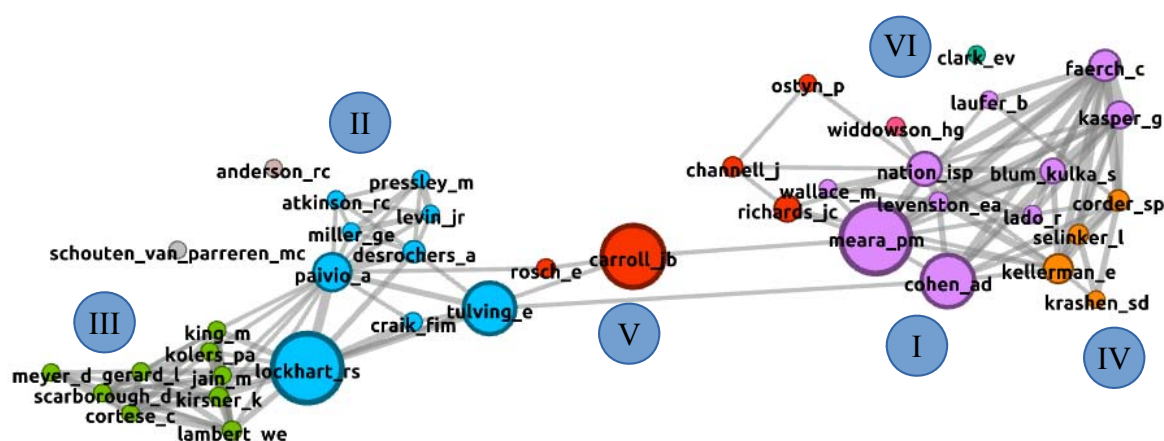


Figure 2: The 1987 Survivors: Significant Influences in 1986 who are also Significant Influences in 1987.

The number of survivors into 1987 is surprisingly small: only forty sources appear in both lists – about two thirds of the sources appearing in the 1986 map have been eliminated, and the survivors make up less than half of the tally of sources in the 1987 map. It is easy to identify some themes that appear in the survivor map. **Imagery and mnemonics** forms a coherent cluster of survivors (Paivio, Pressley, Miller, Levin, Desrochers), and as we have already noted, the cluster is strongly linked with the **performance of bilinguals** cluster. The key figures in this second cluster are Lambert and Kolars, who have appeared consistently in our maps from 1982 onwards though they now seem less influential than they were. Also important here are Kirsner and his colleagues, Jain and Lockhart, who form a cluster of influences whose work concerns **The Bilingual Lexicon** and how it is structured. The general feature that distinguishes this group of survivors is a methodological focus on experimental methods.

It is rather more difficult to identify coherent themes in the other three clusters in Figure 2. Taken together, these three clusters embody more of a pedagogical focus than we found in the clusters on the Western edge. **Transfer** (Kellerman) and **Error Analysis** (Corder, Selinker) appear to be strong themes here. Eve Clark reflects a continuing theme in L1 acquisition. There is a very strong, but rather eclectic set of Israeli sources (Cohen, Blum-Kulka, Levenston, Laufer), a Danish group (Faerch and Kasper) with a methodological interest in introspection. Ostyn and Channell represent an ongoing and growing interest in **semantics** and **meaning**. We also find a number of sources who are cited for their overview work (Nation, Richards, Wallace, Meara and Levenston).

Figure 3 shows the extent of new entries in the 1987 data – these sources make up more than half of the total entries in the 1987 map, and so it is not surprising that the main themes that we identified in Figure 1 also appear strongly here. Surprisingly, the new sources map appears to be relatively compartmentalised, and to a large extent they complement the clusters that we identified in the 1986 map. The new entries tend to have co-citation links within their own cluster, but not beyond it.

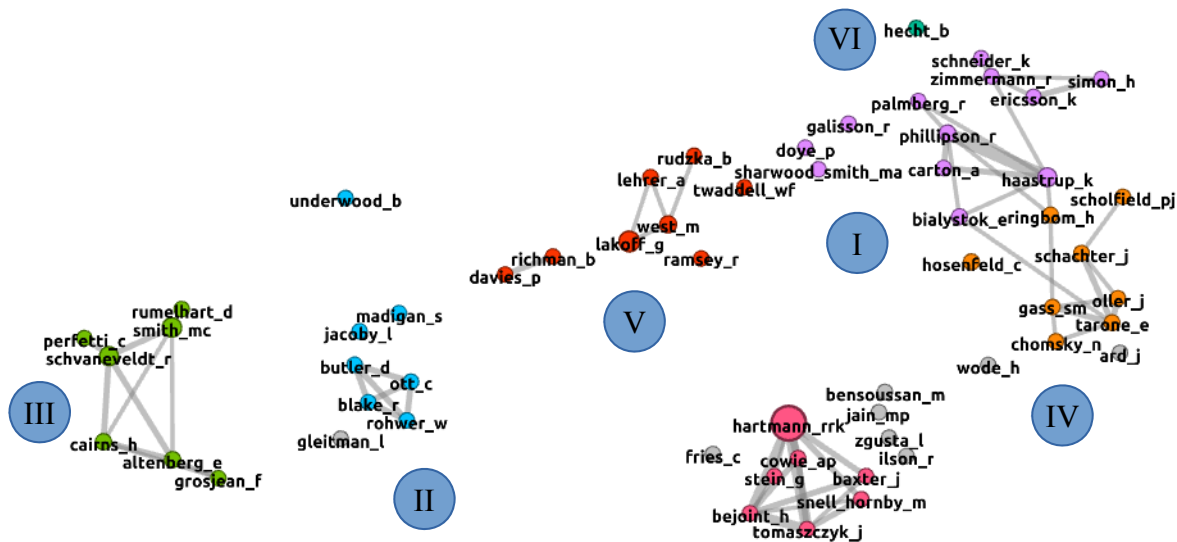


Figure 3: New sources in the 1987 data.

The two clusters at the western edge of the map clearly represent a consolidation of familiar themes. Ott, Butler and Blake are a new take on imagery and L2 vocabulary learning (cluster II). In the psycholinguistics cluster, (cluster III), Perfetti represents a set of influences who are concerned with L2 writing systems, particularly Chinese, while Altenberg and Cairns are a new focus on phonotactic constraints in two languages. The other sources here are mostly collaborators with Kirsner in his work on bilingual lexical models. The other clusters in Figure 3 all point to some significant shifts in the applied linguistics research.

The most striking feature here is the emergence of a new theme covering lexicography and dictionaries – the cluster focussed on Hartmann. This cluster contains a large and very well-connected set of new sources who are lexicographers, as well as some sources who are interested in how L2 learners use dictionaries. This set of sources seems to represent a new and significant research theme that has not had much of a presence in our earlier maps. It also illustrates the power of thematic collections of papers to radically change the overall structure of a small co-citation map. All the new sources in this sub-cluster published work in Cowie's edited collection, and cite each other extensively.

The loose central cluster shows some consolidation of the semantics theme that we noted in Figure 2. (Rudzka, Lehrer and Lakoff).

The largest cluster of new sources appears at the Eastern edge of the map, but the new sources here are less focussed than we noted for the central cluster. There is some evidence of consolidation for the Scandinavian vocabulary research that we have noted previously (Ringbom, Palmberg, Haastруп and Phillipson). Ericsson and Simon are a methodological source: they work on protocol analysis and are mainly cited by the Scandinavian vocabulary

researchers. We also have the appearance of group of North American sources (Tarone, Schachter, Oller, Ard, Gass – and even Chomsky!) who are not specifically vocabulary researchers, but do provide a theoretical framework in which vocabulary research develops. Significantly, perhaps, we can also note the emergence of a coherent group of German language researchers (Zimmermann, Schneider, Wode, Hecht and Doyé) that has not been evident in our earlier maps. Bialystok represents a new interest in bilingual children’s vocabulary acquisition. Scholfield looks to have the potential to forge links between the dictionaries/semantics cluster and error analysis, as his own work spans both these areas. For the moment, however, this potential remains unrealised.

To summarise, then, L2 vocabulary research published in 1987 continues to grow, and there is some evidence for new research themes becoming important in this data set, while other long established themes are in retreat. The psycholinguistic sources appear to be particularly volatile. National and regional groupings play an important role in the cluster structures.

3. Part 2. A wider perspective: 1983–1987

In this section, I will try to place the 1987 data into a larger picture by analysing data from a five-year window that covers 1983–1987. We begin by summarising in Table 3 the 1982–86 data set, which was reviewed in Meara (2018).

Table 3: *The main characteristics of the 1982–86 data set*

Number of papers in the data set:	317
Number of authors contributing to the data set:	309
Number of sources cited in the data set:	3480
Inclusion threshold for this data set	9 citations
Number of cited sources meeting the inclusion threshold	94
Number of clusters identified by Gephi	5+1
I: semantics, meaning, transfer, lexical errors	
II: brain functions and language, performance of bilingual speakers	
III: frequency counts and corpus linguistics	
IV: L2 reading skills	
V: applications of structural semantics to teaching materials	
VI: one disconnected singleton (Wilkins)	

The raw statistics for the 1983–87 data set are broadly similar to the earlier data set. This is not surprising, since the 1983–86 data is common to both data sets. However, the new 1987 data set is considerably larger than the 1982 data set that it replaces, and this changes the structural relationships between the sources, bringing new sources to prominence. The broad characteristics of the 1983–87 data set are summarised in Table 4.

Table 4: *The main characteristics of the 1983–87 data set*

Number of papers in the data set:	355
Number of authors contributing to the data set:	326
Number of sources cited in the data set:	3816
Inclusion threshold for this data set	10 citations
Number of cited sources meeting the inclusion threshold	92

Number of clusters identified by Gephi	5+2
I: overviews, lexical inferencing, transfer, lexical errors	
II: word recognition in an L2, performance of bilingual speakers	
III: word counts and dictionary use	
IV: semantics and meaning	
V: European vocabulary research (Netherlands and France)	
VI: two disconnected singletons (RC Anderson and Galisson)	

Clearly, it would be wrong to assume that removing the 1982 data and adding the 1987 data is just a cosmetic change that leaves the underlying structures largely intact.

The 1983–87 data set consists of 355 papers, making this data set about 12% larger than the earlier one. The individual papers are not listed here for reasons of space, but interested readers can access the list via the VARGA database (<http://www.lognostics.co.uk>). Set the search start date to 1983 and set the finish date to 1987. Enter ## as the search term, and the program will return a complete list of all the papers included in this data set.

326 unique authors contribute to at least one paper in the data set, a small increase on the 1982–86 figures. Table 5 shows the number of authors who make N contributions to the data set. The field as a whole contributes to be dominated by authors who make only a single contribution, although the proportion of these cases has fallen slightly to 71%. Table 5 also shows that the field as a whole has a smaller number of prolific authors than we might expect. The data here shows a significant shortfall compared with the numbers we would expect based on predictions made using Lotka's Law.

Table 5: *The number of authors making N contributions to the data set, and the number of authors expected to make N contributions, given 253 authors who make a single contribution.*

# contributions	10	9	8	7	6	5	4	3	2	1
1983–87 authors	1				5	5	2	15	45	253
Lotka's model	3	3	4	5	9	10	16	28	63	253

The most prolific authors in the 1983–87 window are Meara (10 papers), Broeder, Cohen, Palmberg, Laufer and Zimmerman (6 papers each), Beheydt, Extra, Nation, Ringbom, van Hout (5 papers each). These figures represent a very small improvement on the 1982–86 figures, but the shortfall in Big Hitters remains a striking feature of the data.

A preliminary analysis of the co-citation data from the 1983–87 data set showed that 3816 sources were cited in the data set – a small increase on the 1982–86 data set. However, the distribution of these citations remains very similar to the distribution of the earlier set. (See Table 6).

Table 6: *The number of authors cited in N papers in the combined 1983–87 data set.*

N cites	45	44	43	42	41	40	39	38	37	36	35	34	33	32	31
sources					1								1	2	1
N cites	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16
sources		1	1	1	1	2	2			1	2	3	4	7	6
N cites	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
sources	2	7	6	14	11	17	24	27	47	55	72	133	234	576	1584

The most cited sources here are Meara (41), Levenston(33), Krashen and Lambert (32), Corder (31), Richards (29) Faerch(28), Kellerman (27), Nation (26), Cohen and West (25), Eve Clark and Kolers (24) and Schouten van Parreren (21). Most of these sources were also in the most-cited list for 1982–86, but Lambert has slipped from top place, while Albert, Obler, H Clark and Macnamara have dropped out of the list. Four new entries – Faerch, Kellerman, Nation and West – are in place. 93 of these sources were cited at least ten times. The pattern of co-citations among the 93 sources was analysed using the Gephi software package. Figure 4 shows the basic map for 1983–87.

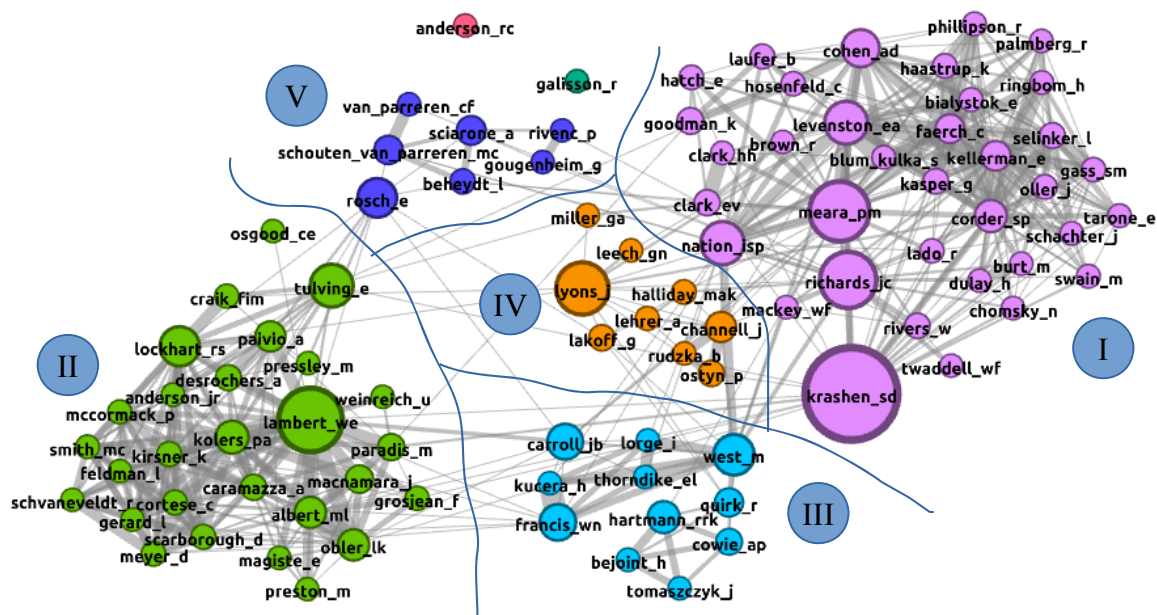


Figure 4: Patterns of co-citation among the 93 most cited sources in the 1983–1987 data set. Threshold for inclusion is 10 citations in the data set, with a minimum co-citation strength of four. Nodes are sized according to their betweenness centrality.

The general features of this map are not very different from the 1982–86 map (see Meara: 2018). We have a broad split between the sources who would identify themselves as psychologists (the cluster on the west of the map), and those who would identify as linguists, and we have very strong clustering within these groups, but rather weaker links between the clusters. The details of the map have changed however.

Gephi identifies five main clusters and two disconnected singletons in the data.

Cluster I, the large cluster at the Eastern edge of the map is the core of the applied linguistics sources. Surprisingly, this cluster is still dominated by Krashen, who appears to act as the main linking node between this cluster and the psychologists in Cluster II. Levenston, Meara, and Richards continue to play a central role in this cluster, but significantly, they are joined by Nation who was previously identified with a different group of sources. This cluster accounts for about 40% of the sources that make up the map.

Cluster II, the larger cluster at the West end of the map, is the now familiar group of psychologists whose work influences L2 vocabulary research. This cluster is about the same size as the equivalent cluster in the 1982–86 map – it comprises about 30% of the sources – but its membership has changed quite substantially.

Cluster III, accounting for 11% of the sources, is the small blue cluster in the centre of the map. This cluster appears to be made up of two sub-groups. The first is the familiar list of frequency counts that we noted in the earlier map. The second is a group of lexicographers who are interested in the role dictionaries might play in L2 learning. The links between these two groups are tenuous, as are the links from this cluster to the rest of the network.

Cluster IV, the second small cluster in the centre of the map, is readily identified as a semantics and meaning cluster. The key source here is Lyons.

Cluster V, the last small cluster, at the top central portion of the map, is a small group of European vocabulary researchers, which again can be seen as a combination of two sub-groups. Gougenheim and Rivenc represent the remnants of the *Français Fondamental* project, while the remaining members of this cluster are a geographical group based in Belgium and the Netherlands. The key source here is Schouten-van Parreren.

Of the **two singletons**, RC Anderson is a reading psychologist, while Galisson is an important French researcher, who was heavily engaged in revitalising French vocabulary research. Both are extensively cited in the data set, but they have many weak co-citation links, rather than a few strong ones, and this leaves them disconnected from the main network.

The addition of the 1987 data and the loss of the 1982 data has brought some changes of emphasis to the 1983–87 map. The simplest way to compare this network with the earlier one is to look at the new sources which appear in the map as a result of the addition of the 1987 data. Somewhat surprisingly, given the very large turnover that we identified in the 1987 data set, only 16 new sources appear in this new five-year map.

The majority of these new sources appear in Cluster 1: **Haastrup, Kasper** and **Phillipson** are indicative of the surge in Scandinavian research that we have noted earlier in this paper; **Schachter, Twadell** and **Tarone** reflect an increased number of North American influences in the data set. Cluster II has four new entries (**Scarborough, Cortese, Gerard** and **Smith**). All four underline the increasing importance of experimental studies of bilingual word recognition. Cluster V has two new entries: **Rivenc** and **Beheydt**, represent a strengthened French language strand in the map. Cluster III has two new entries **Tomaszczyk** and **Hartmann**, both underlining the growing importance of dictionary research as a distinct strand in the L2 vocabulary research of the mid 1980s. **Rudzka** and **Lakoff** strengthen the semantics centre in cluster IV.

There are some fairly obvious changes here. The most important ones are listed below:

- a. a separate semantics and meaning cluster seems to have broken away from the main L2 vocabulary research cluster;
- b. the word lists and frequency count cluster has been extended by the addition of a dictionary use theme.
- c. the separate L2 reading cluster has been absorbed into the main L2 vocabulary cluster, along with most of the new entrants to the data set.

Alongside these obvious changes, it is also worth noting some more subtle changes in the structure of the map. These mainly concern the role of individual sources as links between the clusters. In the 1983–86 map, we identified Krashen and Lambert as the key sources in terms of their betweenness centrality. (This is a measure which identifies sources that provide links

between the different components of the map.) Krashen and Lambert both continue to play this role in the 1983–87 map, but their importance seems to have diminished. This might be an indication that we could expect them to play a much smaller role in future maps. H Clark was the third most important influence in 1983–86 map, but he plays only a minor role in this one. In contrast, Nation, Meara, Levenston and Richards all seem to have grown in terms of the betweenness centrality measure. These sources make up a very strong group that is clearly the core of the Applied Linguistics cluster, and we can expect their roles to become enhanced in future research. The surprise winner in this data set is Lyons, who has moved from a relatively insignificant position in the 1982–86 data set to a much more central role in the 1983–87 data set. I think this reflects a general increase in the amount of L2 vocabulary research that is focussed on semantics and meaning during this period. However, Lyons does not appear as a significant source in the single year 1987 map, and my impression is that this strand of research is relatively short-lived. We might expect some significant changes to this part of the map in the years to come.

4. Discussion

Four important themes seem to emerge from this analysis.

The first theme is that both the one year 1987 data set and the five year 1983–87 data set show that the field is far from settling down into a steady state. New research themes seem to be coming on line with some speed, and new names are replacing older ones in our list of Significant Influences. Sources who were previously key figures in our network appear to be fading in favour of newer ones with a specific interest in L2 vocabulary acquisition, rather than a general interest in language teaching or psycholinguistics. 1987 seems to be the year when L2 vocabulary researchers make up the largest cluster – a significant change from some of our earlier maps when only a handful of cases who fit this description could be identified. The two largest clusters in our map seem to be particularly volatile. Surprisingly, although the size of the psycholinguistics cluster remains fairly constant, its membership changes dramatically in 1987, suggesting that specific psycholinguistic research has a relatively short-lived impact on L2 vocabulary research – a sort of band-wagon effect, perhaps. The only long-term survivors in this group are Lambert, Macnamara and the Montreal group, but their influence is clearly diminishing by 1987, and it is not clear whether their role will be taken up by different strands of psycholinguistic research, or whether, in the longer term, it will just disappear. The Applied Linguistics Cluster, on the other hand, seems to be volatile in a different way. Most of the growth in 1987 accrues to this cluster, most of the new sources are attached to it, and a number of sources who belonged to other, smaller clusters in our earlier maps have now become part of this core cluster. The two obvious cases here are Nation and Laufer, who will go on to become central figures in the L2 vocabulary research enterprise. There is also some volatility in the small clusters III, IV and V in Figure 4. All these clusters have grown, but they still remain relatively detached from the main clusters I and II, and it is not all clear whether they will go on to be permanent features of the vocabulary landscape.

A second feature worth commenting on is the role of national research groupings in the maps. Not surprisingly, most of the significant sources in the 1983–87 map are English writers

– some working in the UK, many working in North America – and this clearly reflects the dominance of UK and North American publishers. Not all research groups have these advantages. German language research, for example, is beginning to form a large part of L2 vocabulary research by 1987, but its impact on the five year map is minimal. Hardly any of this very interesting research is cited in the English language literature. The same problem is to be found with the Dutch, French and Belgian sources that make up Cluster V in Figure 4. This work, like the German work mentioned earlier, is also rarely cited in the English language research literature. It does figure often enough for it to form a separate cluster of sources, but this cluster is not well integrated with the clusters in the map. Schouten-van Parreren is a major figure in Dutch language research, but is mostly cited by other Dutch researchers. Her work deserves to be much better known than it is. To some extent, the insularity of small national groupings is a problem will be solved by two important developments that took place in the late 1980s. One, surprisingly, is the advent of cheap air-travel, the concomitant growth of international conferences and seminars and the possibility of actually meeting people whose work one has long admired from afar. The other, of course, is the development of the Internet, and the way it breaks the monopoly held by academic publishers, and changes the way researchers interact with each other. It remains to be seen how these changes will alter the appearance of our maps in future years. In the meantime, what goes on in L2 vocabulary research is strongly influenced by other research that goes on in the local environment. A good example of this is the new dictionary use sub-group, for example, which is almost entirely made up of researchers with strong UK connections. This may reflect not just a long tradition of dictionary work among British Applied Linguists (e.g. Palmer and West), but also the appearance of two major learners' dictionaries (the *Oxford Advanced Learners' Dictionary* and the *Longman Dictionary of Contemporary English*), and the growing importance of corpus linguistics in the UK at this time. Clearly L2 vocabulary research does not operate in a vacuum.

A third point that deserves comment is the way that the larger 1983–87 map masks changes which are taking place within a single year. Two illustrations will explain why I think this is important.

Our prolific authors list for 1987 contains two authors whose contribution to the 1987 data set are outstanding. Zimmermann contributed four papers in 1987, and had already contributed three papers in 1986, making him a very significant contributor by 1987. Hardly any authors have larger contributions than this during the 1983–87 period. However, this work is not cited often enough for Zimmermann to appear as a major figure in the co-citation map. He does have a small role in the 1987 map, but not one which draws attention to his status as perhaps the major figure in German vocabulary research. There is clearly a disjunction between being a prolific author and being a highly cited source, and it would be worth looking at disjunctions of this sort in more detail. The second author, who highlights a similar issue, is Scholfield, who appears in the 1987 prolific authors list with three publications. Three publications in a single year was a significant achievement in 1987 – only three authors produced this amount of material. Scholfield duly appears as a new entry in the 1987 map, though like Zimmermann, he is not cited often enough to play a role in the larger five-year map. At the very least, these anomalies suggest that there is a time-lag between authors appearing in the prolific authors list and their appearance as frequently cited sources in the co-citation maps. More interestingly,

perhaps, they suggest that an early appearance in the prolific authors list might be turn out to be a useful characteristic of sources who are going to become very strong influences in future years.

Finally, it is worth noting that the changes we have identified in this paper were also being anticipated by writers at the time. In a particularly insightful paper, Laufer (1986) noted that the development of communicative approaches to language teaching would require a rethink of the role of vocabulary in language teaching. She identifies a number of research areas which are beginning to impact on L2 vocabulary practice. These include the development of structural semantics, an increased focus on the performance of learners, especially advanced learners rather than beginners, the importance of fluency over accuracy, and the development of functional and notional approaches to the language syllabus. It is relatively easy to identify all these strands in the maps I have presented here. And it is particularly pleasing to note that the post hoc co-citation method is able to pick up the concerns that were being articulated by key researchers at the time.

5. Conclusion

The next paper in this series will examine the research published in 1988 in the context of a five-year window covering 1983–98. In the meanwhile, readers who find these studies illuminating might be interested in another paper which covers the whole of the vocabulary output published in a single journal (*System*) between 1973 and 2017 (Meara in press). This paper explores the idea of a *first paradigm* in vocabulary acquisition research (Kuhn 1962).

References

- Arik, B. T., and E. Arik. 2017. Second language writing publications in Web of Science: A Bibliometric Analysis. *Publications* 5(4): npn.
- Bastian, M., Heymann, S., and M. Jacomy. 2009. Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media.
- Carter, R. 1987. *Vocabulary: applied linguistic perspectives*. London: Allen and Unwin. 1987.
- de Schryver, G-M. 2009. Bibliometrics in Lexicography. *International Journal of Lexicography* 22(4): 423–465.
- Koster, C. 1987. *Word recognition in foreign and native language*. Dordrecht: Foris. 1987.
- Kuhn, T. 1962/1971. *The structure of scientific revolutions*. Chicago, Il.: University of Chicago Press.
- Laufer, B. 1986. Possible changes in attitude towards vocabulary acquisition research. *International Review of Applied Linguistics* 24(1): 69–75
- Lei, L., and D. L. Liu. 2019. The research trends and contributions of *System's* publications over the past four decades (1973–2017): A bibliometric analysis. *System* 80: 1–13.
- Lei, L., and D. L. Liu. 2019. Research trends in *Applied Linguistics* from 2005 to 2016: A bibliometric analysis and its implications. *Applied Linguistics* 40(3): 540–561.
- Lotka, A. J. 1926. The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences* 16(12): 317–324.
- Meara, P. M. 1987. *Vocabulary in a second language. Volume 2*. London: CILT. 1987
- 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: John Benjamins.

- 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 L2 vocabulary research in 1982–1986. *Linguistics beyond and Within 4*: 108–128.
- Meara, P. M. *in press*. The emergence of a first paradigm in Vocabulary Research: The bibliometrics of *System. Vocabulary Learning and Instruction*.
- Meara, P. M. VARGA: the Vocabulary Acquisition Research Group Archive. Available at: <http://www.lognostics.co.uk/varga/> (accessed November 2019).
- Price, D. 1965. Networks of scientific papers. *Science* 149(3683): 510–515.
- Ringbom, H. 1987. *The role of the first language in foreign language learning*. Clevedon: Multilingual Matters. 1987.
- 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.
- White, H. D., and B. C. Griffith. 1981. Author cocitation: A literature measure of intellectual structure. *Journal of the American Society for Information Science* 32: 163–171.
- Wotjak, G., and U. Herrmann. 1987. *Kleines Wörterbuch der "falschen Freunde" Deutsch-Spanisch*. Leipzig.

Appendix 1: Co-citation analysis: The methodology

The co-citation method 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 Price (1965), has been extensively used to analyse research in the natural sciences (e.g. White & Griffith 1981), but does not yet seem to have been adopted as a standard tool by researchers in the Humanities. Some recent work on the bibliometrics of linguistics *does* exist. An early example is de Schryver (2009), which deals with Lexicography. More recent examples include Lei & Liu (2019a and 2019b) which analyse the entire output of *System* and *Applied Linguistics*, respectively. Arik & Arik (2017) is a more focussed piece that looks at second language writing.

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.

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, & 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

Theses

Duriau, N *Reading techniques and vocabulary retention*. Thesis: Louvain la Neuve. 1987.

Nyikos, M *The effect of color and imagery as mnemonic strategies on learning and retention of lexical items in German*. PhD, Purdue University. 1987.

Unobtainable papers

Alfes, L Etymologie – philologische Nostalgie? *Neusprachliche Mitteilungen* 1(1987), 9–17.

Corrales, O and ME Call At a loss for words: the use of communication strategies to convey lexical meaning. *Foreign Language Annals* 20(1989), 227–240.

French, MP and N Cook The introduction of social studies vocabulary by semantic feature analysis: using a microcomputer data base program – a description. *The Language Teacher* 22(1987), 16–19.

Holtwisch, H Vokabelarbeit im ersten Lernjahr Englisch. *Praxis des neusprachlichen Unterrichts* 34(1987), 390–395.

Kaźmierczak, E Die “falschen Freunde” als Fehlerquelle in der polnisch-deutschen Übersetzungspraxis. *Kwartalnik Neofilologiczny* 34,3(1987), 319–329.

Schouten-van Parreren, MC Woorden leren en lezen. [Learning words and reading.] *Engels* 13,1(1987), 2–7.

Silva, HHG Os cognatos como factor de legibilidade de texto em lingua estrangeira. [Cognates as a factor in the legibility of texts in a foreign language.] *ESpecialist* 16(1987), 35–58.

Vigner, G Competence textuelle et competence lexicale. [Textual competence and lexical competence.] *Le français dans le monde* 210(1987), 47–53.

Voionmaa, K En studie av lexikal överföring hos vuxna språklinärare. [Lexical development in adult language learners.] In: **E Warde et al.** (eds.) *Aspects of Bilingualism*. Uppsala. 1987.

Wu Qiku How to master English and Russian words effectively? *Higher Education Studies: Nanjing Chemical Engineering College* 3(1987), 43–51.

Monographs and bibliographies

Carter, R *Vocabulary: Applied linguistic perspectives*. London: Allen and Unwin. 1987.

Koster, C *Word recognition in foreign and native language*. Dordrecht: Foris. 1987.

Meara, PM *Vocabulary in a second language. Volume 2*. London: CILT. 1987.

Ringbom, H *The role of the first language in foreign language learning*. Clevedon: Multilingual Matters. 1987.

Wotjak, G and U Herrmann *Kleines Wörterbuch der “falschen Freunde” Deutsch-Spanisch*. Leipzig 1987.

Edited collections (the papers that make up these collections are included in the 1987 data set)

Carter, R and M McCarthy (eds.) *Vocabulary and language teaching* London: Longman. 1987.

Cowie, AP (ed.) *The Dictionary and the Language Learner*. Tübingen: Niemeyer. 1987

Diller, H-J, et al. (eds.) *Wortschatzarbeit*. Heidelberg: Winter. 1987.

Gass, SM (ed.) The use and acquisition of the second language lexicon. *Studies in Second Language Acquisition* 9,2(1987) (whole volume).