

The British are coming...! A bibliometric analysis of L2 vocabulary research in 1988

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 1988. Two analyses are presented. The first is a detailed account of the 1988 research on its own terms. The second analysis places this work in a larger context by looking at the research published in a five-year window covering 1984–88. The analyses identify important themes in the research and significant sources who are influencing the way the research is developing at this time. A particularly important new research theme centred around corpus linguistics appears in the 1988 data, and there are some surprising changes to the list of influential sources.

Keywords: L2 vocabulary acquisition, vocabulary research, bibliometric analysis

1. Introduction

This paper is the eighth 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, 1987 and 2006 (Meara 2012, 2014, 2015, 2016, 2017, 2018 and 2019). This paper is a sequel to my earlier LingBaW papers, in that it covers the research output of 1988. The paper contextualises this research by means of an additional analysis of the research published in the five-year window of 1984–88.

The paper falls into two parts. Part I reviews the new research that appeared in 1988 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 1984–1988. Both parts make use of the co-citation methodology that was used in the earlier papers. For readers who are not yet familiar with this approach, I have provided a methodological summary in an Appendix.

2. Part 1. The new research published in 1988

At first glance, 1988 seems to be a poor year for vocabulary research. The previous four years had shown a small but steady increase in the number of research outputs published, but for 1988 the total number of outputs falls just below 100 for the first time since 1985. However, on

closer inspection, this fall turns out to rather deceptive. The Vocabulary Acquisition Research Group Archive (VARGA) database for 1988 (Meara n.d.) identifies only 99 outputs, but the make-up of these sources is rather different from what we found in earlier years.

2.1. *The data sources*

In 1988, we have a relatively large number of books (9) and a handful of computer programs (2), as well as a grand total of 21 book chapters that deal with L2 vocabulary acquisition in one way or another, and 50 standard research papers. The VARGA database for 1988 also records five PhD theses, and three cited Masters theses. However, since the research literature tends not to cite theses reliably, VARGA typically under-reports work of this type, and it is possible that some sources of this type have been missed. VARGA includes theses which are cited in later research work, but does not attempt to log every thesis presented, and the work listed here should not be taken as an exhaustive list. Not all of the work published in 1988 is included in the analysis that follows. Books, theses and similar works are conventionally excluded as raw data in bibliometric analyses of the type reported in this paper, on the grounds that they cite the research in a way which is different from what we expect in a normal research paper. A few works of this type have been excluded from the analyses reported in this paper. I have also excluded from the analysis two computer programs, developed by Esser & Widdig and by Scott & Johns. Both these works are reported more fully in separate research papers which *are* included in the analysis. Table 1 lists the complete set of 1988 publications which are excluded from the analysis that follows.

Table 1: *Work published in 1988 but excluded from the analysis in this paper*

BOOKS

- Boch, R. 1988. *Les faux amis aux aguets: Dizionario di false analogie e ambigue affinità fra francese e italiano*. Bologna.
- Broeder, P., G. Extra, R. van Hout, S. Stromqvist and K. Voionmaa. 1988. *Processes in the developing lexicon*. Tilburg.
- Browne, V. 1988. *Odd pairs & false friends: Dizionario di false analogie e ambigue affinità fra inglese e italiano*. Bologna.
- Carter, R. and M. McCarthy (Eds.) 1988. *Vocabulary and Language Teaching*. London: Longman.
- Leiste, D., C. Döll and A. M. Tereso Domingos. 1988. *Kleines Wörterbuch der 'falschen Freunde': Deutsch-Portugiesisch, Portugiesisch-Deutsch*. Leipzig.
- Lozanov, G. and E. Gateva. 1988. *The Foreign Language Teacher's Suggestopedic Manual*. Gordon and Breach Science Publishers.

SOFTWARE

- Esser, R. and W. Widdig. 1988. *WE/WT, Autorensystem für die Wortschatzarbeit. Benutzungsanleitung*. Cologne: RRZK.
- Scott, M. and T. Johns. 1988. *Oxford English Software: Microconcord 1.0*. Oxford: Oxford University Press. 1988.

THESES and DISSERTATIONS

- Ahmed, M. O. 1988. *Vocabulary Learning Strategies*. PhD Thesis, UCNW Bangor. 1988.
- Mei Lin. 1988. *An assessment of the treatment of vocabulary in a series of pilot coursebooks used by English-Language learners at tertiary level in the Peoples' Republic of China*. MA Thesis. York University.
- Novda, L. M. 1988. *The word retrieval process and reading acquisition and development in bilingual children*. PhD thesis. Harvard University.
- Sonaiya, Q. C. 1988. *The lexicon in second language acquisition: a lexical approach to error analysis*. PhD thesis. Cornell University.

- Söderman, T. 1988. Word associations of foreign language learners and native speakers – a shift in response type and its relevance for a theory of lexical development. Masters Thesis. Åbo Akademi.
- Verkaik, P. and P. van der Wijst. 1988. Taal verlies en woordherkenning in het Frans als vreemdetaal. [Language loss and word recognition in French as a foreign language.] Masters thesis: Katholieke Universiteit, Nijmegen.
- Weltens, B. 1988. The attrition of French as a foreign language. PhD Thesis, Katholieke Universiteit Nijmegen.
- White, W. H. 1988. Vocabulary acquisition from reading. PhD dissertation, University of Southern California, Los Angeles.

The books and monographs listed in Table 1 are particularly interesting. Three of these are studies that deal with cognates and false friends: Boch (1988) and Browne (1988) both treat false friends in French and Italian; Leiste, Döll and Tereso Domingos (1988) treats false friends in German and Portuguese. Studies of this type have appeared occasionally in previous years, but this set is larger than most, and extends the range of languages. Lozanov and Gateva (1988) deal with Suggestopedia – the idea that languages can be learned subconsciously. This book is not specifically concerned with vocabulary, but vocabulary plays a large role in studies of subconscious language learning, and for this reason, it has been included in the VARGA listing.

Two of the book-length studies are substantial project reports. Broeder et al (1988) summarise the on-going work of the European Science Foundation Project as far as the lexicon is concerned (Purdue 1982, 1993). The report consists of 5 main chapters dealing with: the learner's use of words; the informants and data base used in the studies; richness and variety of the developing lexicon; pronominal reference to person; and word formation processes in talking about entities. Coenen and Vermeer (1988) is another substantial study that attempts to assess how much L2 vocabulary is known by migrant learners of Dutch. Part One of this volume describes how the basic word lists were collected. Part Two consists of a list of 4332 different words used by more than one of the speakers tested.

The rationale for excluding book length treatments is that their normal citation practices differ from what we find in standard research papers. However, the final book, Carter and McCarthy (1988), is odd in this respect. This volume is an edited collection of ten papers, four of which were written by the editors. After some thought, I decided to treat this volume as ten separate entries, despite the overwhelming contribution of the editors.

A new development in the 1988 outputs is the appearance of two computer programs. The most important of these is Scott and Johns' *Microconcord*, a concordancing program. This program was unusual at the time in that it ran on small microcomputers, and did not require a mainframe computer. It also had a very short learning curve, which meant that it could readily be used in ordinary classroom situations. It was rapidly taken up as a tool that language teachers could use to help their students learn the finer points of vocabulary. I excluded this item from the analysis that follows, treating it as a book, rather than a paper. *Microconcord* is not completely neglected, however: it is described and evaluated in other papers which are included in the analysis. The second computer program (Essen and Widdig 1988) appears to be an authoring program for developing computer-based vocabulary exercises. This work was not included in the analysis. It is described in another paper by the same authors, and this paper is included in the data set.

Table 2 lists the small number of papers which I was not able to source, mainly due to the closure of libraries and restrictions on travel during the coronavirus pandemic.

Table 2: *Papers published in 1988 which were unobtainable, and were consequently excluded from the analysis*

- Armand, E. (1988). Typologie des exercices de vocabulaire contenus dans les manuels de français édités en France entre 1970 et 1984 et destinés à l'enseignement-apprentissage du français aux étrangers. [A typology of vocabulary exercises found in French course-books.] *Bulletin de l'Unité de Recherche Linguistique* 4: 97–183.
- Galisson, R. (1988). Le vocabulaire revient: Le vocabulaire en pénitence. Brève histoire d'une disgrâce chronique. *Reflète* 27: 14–19.
- Noyau, C. (1988). Le développement du lexique en langage étrangère: lemmatisation de données orales d'acquisition. [The development of the lexicon in a foreign language: analysis of spoken language data.] *Actes du 2ème Colloque de Linguistique Hispanique*, 199–216. Brest: Université de Bretagne Occidentale.
- Taylor, A. (1988). Learners and English dictionaries: some assumptions and challenges. *Institute of Language in Education Journal* 4: 88–92.
- Tréville, M.-C. (1988). Faut-il enseigner le vocabulaire de la langue seconde? [Do we need to teach vocabulary in a second language?]. In R. LeBlanc, J. Compain, L. Duquette and H. Séguin (eds.) *L'enseignement des langues secondes aux adultes: recherches et pratiques*. Ottawa: Presse de L'Université d'Ottawa.

The four French papers in this table are probably an important omission. We have seen in our earlier papers that French research on vocabulary was beginning to go through something of a resurgence in the late 1980s, and the omission of these papers means that this strand of research is not properly represented in the main analysis. The papers by Armand and Noyau are technically available, but they could only be found in libraries that were not functioning because of the corona virus pandemic. The paper by Galisson was completely untraceable. Again, I suspect that this is an important omission, as Galisson was a very vocal critic of vocabulary teaching in France in the 1980s, and his work developing an alternative pedagogy of vocabulary and lexis was both influential and very controversial. Tréville is part of a vocabulary research group based in Ottawa which was developing an approach to the teaching of French vocabulary that was more empirical and less philosophical than the approaches being developed in France at that time.

The paper by Taylor proved to be untraceable: the paper seems to be a large scale study of dictionary use by L1 Chinese learners of English at University level, and would have made a useful addition to the other research on dictionary use which appeared in 1988.

The remaining 81 items, all conventional book chapters or papers in journals, are not listed here in full for reasons of space. Readers who want to access this list can do so by using the VARGA database: <https://www.lognostics.co.uk/varga/> and entering the search term ## 1988.

The usual superficial analysis of this data set identifies 92 unique contributors. As usual, we find that most of these authors contribute to only a single paper in the data set, and only a handful of authors contribute to more than one paper. The data is summarised in Table 3.

Table 3: *Authors contributing to N papers in the main data set*

No of papers (N)	5	4	3	2	1
Actual data	0	4	1	4	83
Lotka's model N=83	3	5	9	20	83

Carter, McCarthy, Meara and Palmberg all contribute 4 entries, making them the most prolific authors in this year's output. McCarthy is a new entrant to the prolific authors list. Johns, also a new entrant, contributes three papers. Four authors, Nation, Robinson, Summers and Tono all contribute two papers to the data set. Of these, only Nation has appeared in previous counts.

Table 3 also shows that this profile is considerably shallower than we would expect to find in a mature research field. With 83 authors contributing just one paper, we would expect to find about 20 authors who contribute to two papers, nine who contribute to three papers, and so on. In 1988, the field is still short of contributors with a substantial output, just as it was in the previous years. (Lotka 1926. For readers who are unfamiliar with Lotka's approach, a brief account can be found in Appendix 1.).

2.2. *The analysis*

The main analysis in this paper is a co-citation analysis of all the sources cited in the 1988 data set. The methodology is described in Appendix 2. This analysis identified 1391 authors who are cited in the data set, a substantial fall on the 1987 figure of 1587 that I reported in the previous paper. The number of times each of these sources is cited in the data set is summarised in Table 4.

Table 4: *The number of times sources are cited in the 1988 data set.*

frequency	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
cases	1	1	1	1	3	3	2	4	5	10	12	39	55	186	1068

The most cited sources in this data set are Sinclair (15) Nation (14), Meara (13), Channell (12), Krashen, Ostyn and Renouf (11), Carter, McCarthy and Richards (10), and Halliday and West each cited in 9 papers. The main point to note here is the emergence of Sinclair as the most cited source in the data set. Allowing for the smaller size of the 1988 data set, these figures are very comparable with the data I reported for 1987, but it is worth pointing out that once again there is a considerable turnover in the list of frequently cited sources. Only Nation, Meara, and Richards appear in both the 1987 and 1988 lists of highly cited sources. Sinclair, Channell, Krashen, Renouf, Carter, Halliday, McCarthy, Ostyn and West are all new entrants to the 1988 list of significant sources. With the exception of West and Krashen, all these new sources are part of the Nottingham~Birmingham group of researchers located in the UK. In contrast, Faerch, Kasper, Kellerman, Levenston, Hartmann, Blum-Kulka, Carroll, Haastrup and Lockhart, who were all significant sources in 1987, have dropped out of the highly cited sources list, though all of them continue to be cited to some extent in the 1988 data set.

The 1988 data set is actually quite difficult to work with. Conventional practice is that we base our bibliometric maps on the co-citations among the 100 most cited authors in a data set. With this data set, applying a cut-off of four citations leaves us with 82 authors, while a cut-off of three citations gives us 138 authors, where neither of these figures is close to 100. Our analysis of the 1987 data used four citations as an inclusion threshold, and so for the sake of comparison with earlier reports, I have adopted the same inclusion threshold for the 1988 dataset. This decision means that authors need to be cited in just over 4% of the papers published in 1988 to be included in our analysis. The Significant Influences are cited much more than this, of course: Sinclair, for example is cited in 18% of all the 1988 papers.

The citation data for the 82 most cited sources were analysed using the Gephi software package (Bastian, Heymann and Jacomy: 2009), and the results of this analysis are reported in Figure 1. Gephi's analysis identifies six clusters, based on how often the members of the cluster are co-cited in the data set.

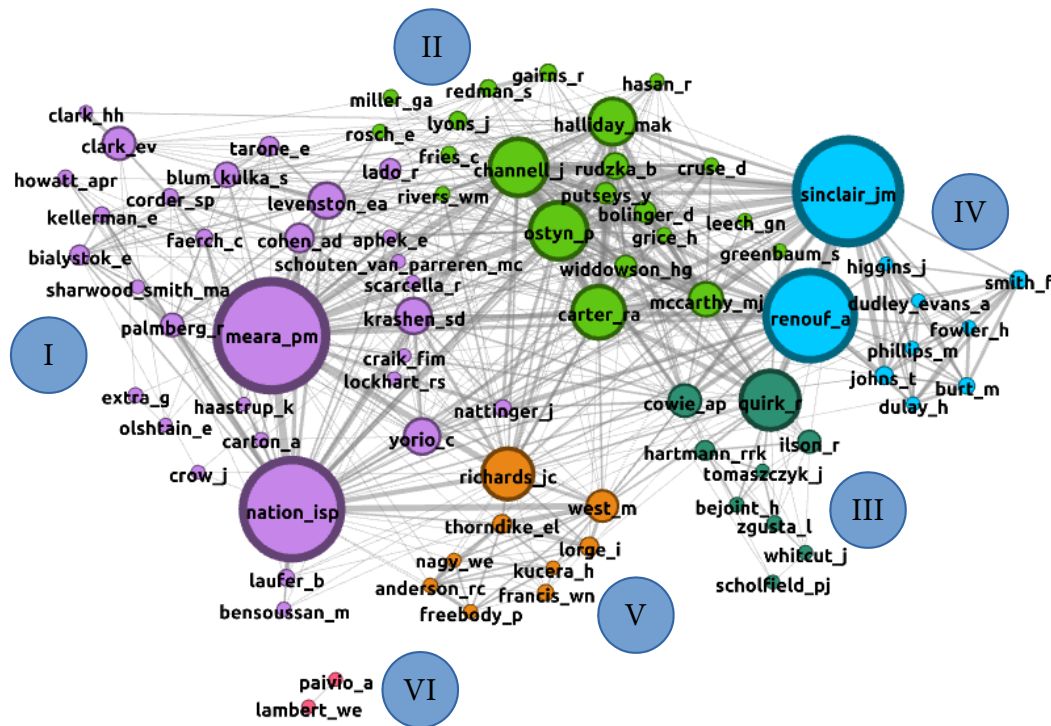


Figure 1: The main clusters in the 1988 data set. The weakest links have been excluded. Nodes are sized according to their betweenness centrality.

Cluster I, at the western edge of the map, dominated by Meara and Nation, is the main vocabulary acquisition cluster. This cluster contains about a third of the sources for 1988, including three of the prolific authors identified earlier. The cluster seems to be mainly concerned with the lexical performance of non-proficient speakers, whether these are bilinguals, L2 learners, or children learning their L1.

Cluster II in the centre of the map, dominated by Channell, Ostyn and Carter, seems to be mainly concerned with meaning, and how semantics can be exploited for language learning. The key source in this cluster is the set of textbooks produced by Rudzka, Channell, Putseys and Ostyn (1981, 1985). These texts, which emphasised an approach to vocabulary based on a componential semantic analysis were identified as important in our earlier papers, and they continue to be influential in 1988.

Cluster III, at the south-eastern sector of the map, is a dictionary research cluster. A cluster of this sort first emerged in the 1987 data set. Here, this cluster appears to be consolidating and reaffirming its position in the map.

Cluster IV, at the Eastern edge the map, dominated by Sinclair and Renouf, is largely composed of researchers working at the University of Birmingham. It reflects the growing importance of corpus linguistics in vocabulary research.

Cluster V, at the southern central part of the map, dominated by J. C. Richards, identifies a number of important English word frequency counts, and research that is informed by this work. A cluster of this sort has appeared in all of our maps so far. Anderson, Nagy and Freebody are mainly concerned with how L1 readers acquire the meanings of words that they encounter while reading.

Finally, **Cluster VI**, the small two person cluster at the foot of the map, seems to be the remains of the psycholinguistics clusters which dominated the research earlier in the decade.

Our earlier analyses suggested that this theme has been becoming steadily less important to L2 vocabulary acquisition research throughout the 1980s, and the isolated position of this cluster in the 1988 map seems to confirm this assessment.

In broad terms, there are three fault lines in this map. The main fault line separates Cluster I and Cluster VI from all the other clusters. The sources that make up this grouping are mainly concerned with experimental data, whereas the other clusters might be described as more concerned with descriptive data, and more specifically concerned with descriptive analyses of English. A second fault line separates cluster III and Cluster IV from the other clusters. The sources in this grouping are mainly descriptive linguists, with specific interests in Corpora and Dictionaries. It is slightly surprising that the dictionary researchers in Cluster V are not more closely associated with the corpus linguists. The third fault line seems to lie between Clusters II and III and the rest of the map. The emphasis in this group is on pedagogy, and how linguistic tools can improve L2 learners' experience.

In structural terms, the 1988 map is rather different from the maps that appeared in our earlier reports. The 1988 map has strong interconnections between most of its clusters, and these connections are less dependent on a few key figures who provide the links between clusters. There are two exceptions to this generalisation. The dictionary cluster, cluster III, is not strongly linked to the rest of the network: most of the sources in this cluster are co-cited with each other, but only rarely with other researchers. The same point could also be made for cluster IV. Again here we have a large group of sources who are cited together, but are only rarely cited with sources outside their own cluster. The exceptions here are Sinclair and Renouf, who are frequently co-cited with sources in other clusters, particularly Cluster II.

Probably the most significant feature of the 1988 map is the almost complete absence of any psycholinguistic sources. In 1987 almost a third of the sources fell into two large clusters that we identified as psychologists whose work had influenced research into L2 vocabulary acquisition. Cluster VI is all that is left of this strand of research in 1988. This does not mean that psychological research has stopped, of course, but the analysis does suggest that it has stopped being influential in the work done by applied linguists. This has been something of a trend throughout the 1980s, but it is surprising to see such a rapid collapse. A small number of psychologists do appear in the map, but their influence appears to be limited. Craik and Lockhart are co-cited with Krashen, but not with other members of cluster I. George Miller and Eleanor Rosch, who were previously loosely attached to the main L2 vocabulary cluster, now appear in cluster II. This change in emphasis – a general loosening of ties between two research traditions – feels like a significant shift in the way the field is structured.

We can see the extent of this shift in Figure 2 and Figure 3. Figure 2 shows the new entries in the 1988 co-citation map. There are 42 new entries in this map, slightly more than half of all the sources that appear in figure 1. But more importantly, we are not dealing with a few isolated new entrants here. Rather, the map shows two well-established research areas – one dominated by Sinclair and Renouf, the other by Carter, McCarthy and Halliday – that have quickly become a core part of the L2 vocabulary research. Methodologically, these two areas rely on formal linguistic analysis, and they seem to represent a return to a more traditional relationship between L2 vocabulary acquisition and linguistic theory (cf. Corder 1973). Both themes have

been present in our earlier maps. Nonetheless, the way these two clusters seem to have completely replaced the psycholinguistics clusters that we found in our earlier maps is striking.

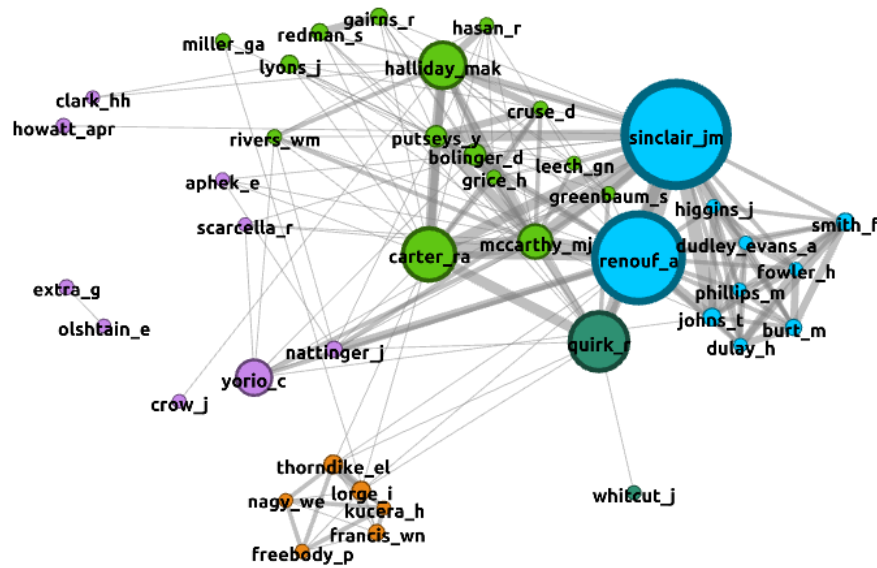


Figure 2: New entries in the 1988 Co-Citation map

Figure 3 shows a map of the sources who appear in both the 1987 map and the 1988 map – the “survivors” map. As usual, this map needs to be treated with some caution, as the 1988 map contains more nodes than the 1987 map. Nonetheless, the number of survivors in 1988 is strikingly small: some 60% of the sources in the 1987 map fail to make it into 1988. The majority of the survivor sources belong to cluster I, with Meara and Nation both consolidating their central position within this group. Krashen remains a surprisingly influential source in the cluster. The dictionary research group is remarkably stable: almost all of its members appear in both the 1987 and the 1988 maps, but as we have already noted, this group appears to be somewhat isolated from the mainstream of L2 vocabulary research, and this may be problematic in future years.

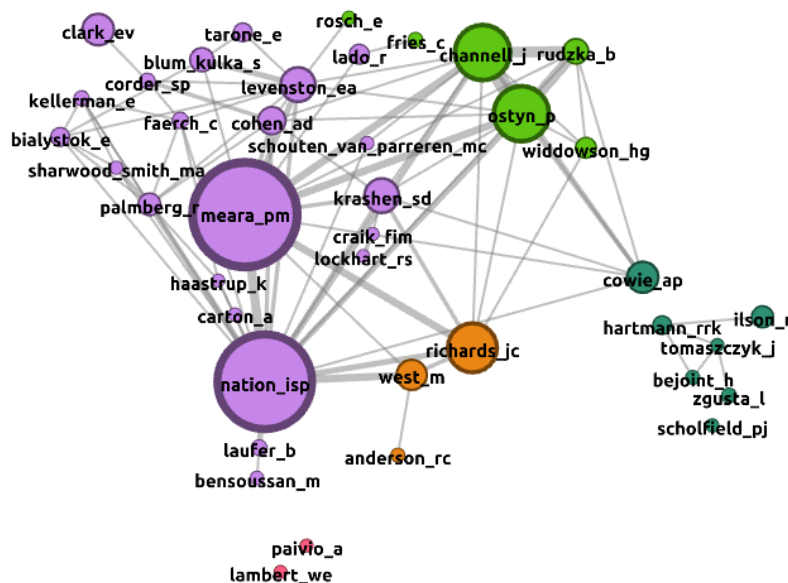


Figure 3: Sources that appear in both the 1987 and the 1988 co-citation maps

3. Part 2. A wider perspective: 1984–1988

We now need to look at the 1988 data in a larger context. In our earlier analysis, we were able to do this by looking at a rolling five-year window. This wider context smooths out some of the fluctuations in the annual reports – people who publish a lot in one year but not in adjacent years, for example – and it allows us to identify longer-term trends in the data. In this paper, our five-year window covers the period 1984–1988.

Table 5 recapitulates the main characteristics of the 1983–1987 window which were discussed in Meara (2019). The raw statistics for the 1984–88 data set are broadly in line with the earlier figures, but generally show an increase over the 1983–87 data. The main features of this new data set are reported in Table 6. The data for 1984, 1985, 1986 and 1987 are included in both data sets, but in the current analysis, the 1983 data has been replaced by the larger, and more coherent data set published in 1988. The number of papers included in the new data set is 403, an increase of some 13%, despite the relatively low number of outputs in 1988. As usual, I have not listed all the included papers here, but interested readers can access the list via the VARGA database: (<https://www.lognostics.co.uk/varga/>). Set the search start date to 1984 and set the finish date to 1988. Then enter ## as the search term, and the program will return a complete list of all the papers included in this data set.

Table 5: *The main characteristics of the 1983–1987 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	93
Number of cited sources meeting the inclusion threshold	5+2
I: overviews, lexical inferencing, transfer, lexical errors	
II: word recognition in an L2, performance of bilingual speaker	
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)	

Table 6: *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
Number of cited sources meeting the inclusion threshold	7+3
I: vocabulary acquisition (30+1)	
II: word recognition in an L2, performance of bilingual speakers (26)	
III: meaning, corpus analysis (22)	
IV: word frequency counts (5)	
V: Français fondamental (4+1)	
VI: dictionary research (4+1)	
VII: Dutch research (3)	

375 unique authors contribute to these papers, and increase of 15% on the 1983–87 data set. Table 7 shows the number of authors who make N contributions to the data set. As usual, there is a heavy predominance of authors who make only a single contribution to the data set: 77% of the authors fall into this category – almost identical to the 1983–87 figure.

Table 7: *The number of authors contributing to N papers in the 1984–88 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	1	5	5	19	53	287
Lotka:	1	1	1	1	1	1	3	4	4	6	8	11	18	32	72	287

Ten authors contributed to five or more papers. Meara contributed to 16 papers; Palmberg contributed to nine papers; Laufer, to eight. Zimmermann and Broeder contributed to seven papers, Carter to six. Arnaud, Beheydt, Extra, McCarthy and van Hout each contributed to six papers in the data set. Three of these authors – Carter, McCarthy and Arnaud – are new to the prolific authors list.

The bottom line of Table 7 shows the number of contributions we would expect to find compared with the predictions made by Lotka’s Law (Lotka 1926). Lotka’s model (See Appendix 1) suggests that we ought to have many more authors contributing two, three, four or more papers to the data set than we in fact get. The 1983–88 data does look like a power law distribution, but it deviates substantially from Lotka’s model when we count the number of authors who produce multiple papers. (Technically, the best fit for this data is has a larger exponent than Lotka’s model predicts. The best-fitting curve for the data in Table 7 has an exponent of 2.49. This is a slight improvement on the equivalent figure for the 1983–87 window (2.55). It perhaps indicates that the field is slowly becoming more normalised, though at this stage in its evolution, the field still has a serious over-reliance on authors who contribute to only a single paper in the data set.)

Of course, the fact that some authors contribute several works to a data set does not necessarily mean that their work is influential, so we turn next to the citation data found in the 1984–1988 data set. A total of 4080 sources are cited in this data set, a significant increase on the figure of 3816 that we reported for the 1983–87 window. The data is summarised in Table 8, which shows the number of times the sources are cited in the data set. 65% of the sources are cited in only one paper.

Table 8: *The number of times sources are cited in the 1988 data set.*

frequency										51	50	49	48	47	46
cases										1					
frequency	45	44	43	42	41	40	39	38	37	36	35	34	33	32	31
cases			1			1		1		1			1		3
frequency	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16
cases	1		1	1		5	2		4	1	3	4	12	5	6
frequency	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
cases	10	13	9	14	19	18	31	34	34	49	106	151	252	646	2640

At the other end of the scale, a handful of sources are substantially cited. Meara is cited in 51 papers, Krashen in 43, Richards in 40. Levenston is cited in 38 papers, Nation in 36, and Kellerman in 33. Corder, Faerch and Lambert are all cited in 31 papers, Cohen in 30, Eve Clarke in 28 and Channell in 27. This list of significant influences has changed little from the 1983–87 data: Michael West has dropped out of the list; Joanna Channell is a new entry. Meara, Krashen and Richards are all substantially more cited in the 1984–88 data set than they were in 1983–87. Levenston, Nation, Kellerman, Faerch and Cohen show smaller increases in their citation counts. Lambert's count has increased by 1; Corder's count does not change. The number of sources cited at least 10 times in the data set is 138 – almost a 50% increase over the 1983–87 figure (93).

Custom and practice is that co-citation analysis works with the 100 most-cited sources. For our data set, 101 sources are cited at least 12 times, so the analysis that follows is based on this subset of the data. It is worth noting, however, that only 65 of the 1983–87 sources would have met this threshold. It is also worth noting that restricting the analysis to the 101 most cited sources means that we are ignoring a lot of activity which fails to meet the arbitrary threshold. Figure 4 shows the basic map for 1984–88.

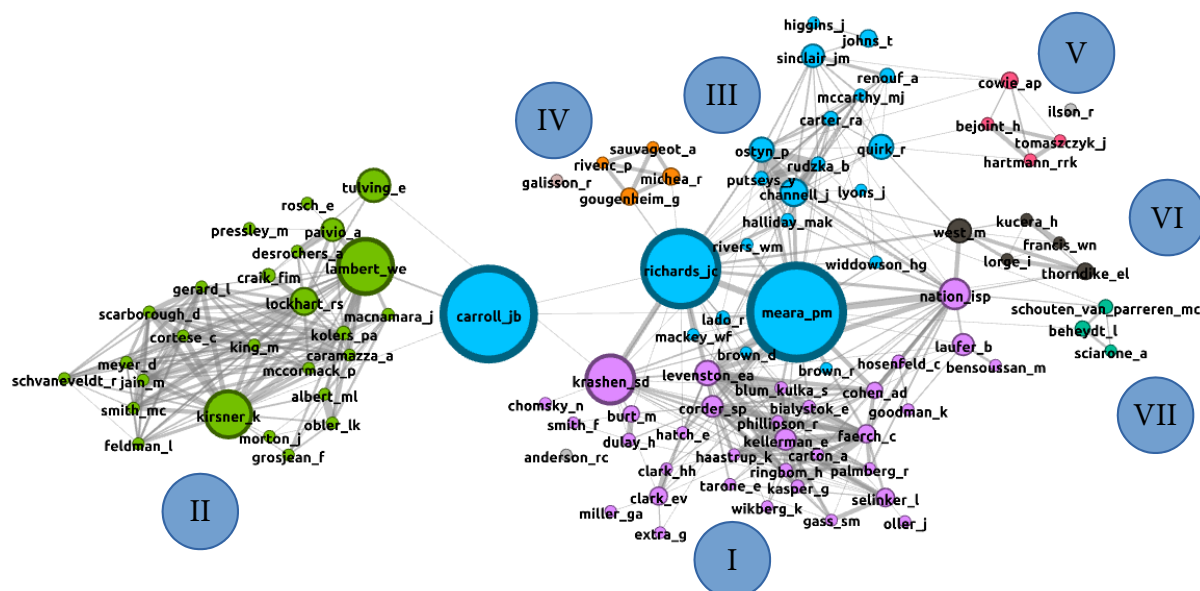


Figure 4: Patterns of citation among the 101 most cited sources in the 1984–88 data set. Threshold for inclusion is 12 citations in the data set with a minimum co-citation strength of six. Nodes are sized according to their betweenness centrality value.

Gephi identifies seven clusters in the data set, along with three singletons (Galissou, Ilson and Miller) frequently cited, but not strongly co-cited with other sources who appear in the map. These clusters are broadly in line with the clusters identified in our analysis of the 1983–87 data, but the addition of the 1988 citation data has introduced some changes of emphasis.

Cluster I, the large dense cluster at the bottom of the map, can clearly be identified both in 1983–87 and 1984–88. This cluster contains most of the empirical work on L2 vocabulary acquisition, with a particularly large sub-group of Scandinavian researchers. Several sub-themes can be identified within this cluster: a sub-cluster centred on Krashen, a lexical errors sub-cluster centred on Corder, a reading sub-cluster focussed on Nation and Laufer, a set of L1

acquisition sources (Eve Clark), and a sub-cluster dealing with transfer issues (Kellerman, Ringbom).

Cluster II is the familiar psycholinguistics group of sources at the western edge of the map. This cluster continues to be dominated by Lambert, and the cluster as a whole is very densely interconnected, though it has hardly any immediate connections with the other clusters in this map.

Again, there are a number of identifiable sub-clusters here: Lambert is a key figure in the performance of bilinguals; Kirsner leads a group of researchers looking at word recognition in bilinguals; Craik, Lockhart, Paivio, Desrochers, Pressley and Tulving are all writing about aspects of memory. Rosch, who previously appeared in an L2 semantics and meaning cluster, is a key source in the psychology of meaning. This cluster is about the same size as the equivalent cluster in the 1983–87 map, but it seems to be less densely connected in 1984–88. Our analysis of the 1988 data suggests that few of these sources are consistently cited in the 1988 research, and this suggests that cluster II is likely to shrivel in future analyses.

The main change in the 1984–88 map is the emergence of a strong new cluster apparently centred on Richards, Carroll and Meara. In fact, this cluster, **Cluster III** at the top centre of the map, is mainly composed of new entrants, with Richards, Carroll and Meara appearing to take a leading role mainly because of their high betweenness centrality scores (see below). The key figure in this new cluster is actually Sinclair. The main theme in this cluster is Corpus Linguistics and its application to vocabulary teaching. The cluster also seems to have absorbed an earlier cluster that dealt with meaning and vocabulary acquisition.

That leaves us with four small clusters which formerly appeared as sub-clusters in a larger L2 vocabulary cluster. These are:

Cluster IV (Gougenheim, Michea, Rivenc and Sauvageot) is a group of French researchers, particularly associated with the work on *Français Fondamental*. I have also included Galisson in this cluster, though his work is highly critical of the *Français Fondamental* approach on the grounds that it ignores the cultural and ethnographic aspects of lexical knowledge.

Cluster V (Cowie, Bejoint, Tomaszczyk and Hartmann) is an L2 dictionary group, that previously appeared as a dictionary and frequency count cluster. In this map, the frequency counts emerge as a separate cluster, **Cluster VI** (West, Kucera & Francis and Thorndike & Lorge).

Cluster VII is a small group consisting of Behydt, Schouten-van Parreren and Sciarone. This cluster is basically a Dutch language research group.

There are very few changes between this map and the equivalent map based on the 1983–87 data set. The main clusters remain largely intact in both maps, though there are some small variations where new clusters have budded off from the main cluster and become more independent. In general, these new clusters are characterised by their having very few co-citation links with the other clusters in the map. Sources in Cluster IV for example, are occasionally co-cited with Richards, but they are not co-cited with the sources in cluster VI, despite their overlapping interests. Similarly, the dictionary researchers are sometimes co-cited with sources in cluster II, but they do not figure strongly in the co-citations with Cluster I. The divorce between cluster II (the psycholinguistics cluster) and the rest of the map has become very clear by 1988.

The main characteristic of the new map is that it is extraordinarily stable, despite the changes which have appeared on a year by year basis. We can see this in Figure 5, which shows the “survivor” sources who appear in both the 1983–87 map and the 1984–88 map.

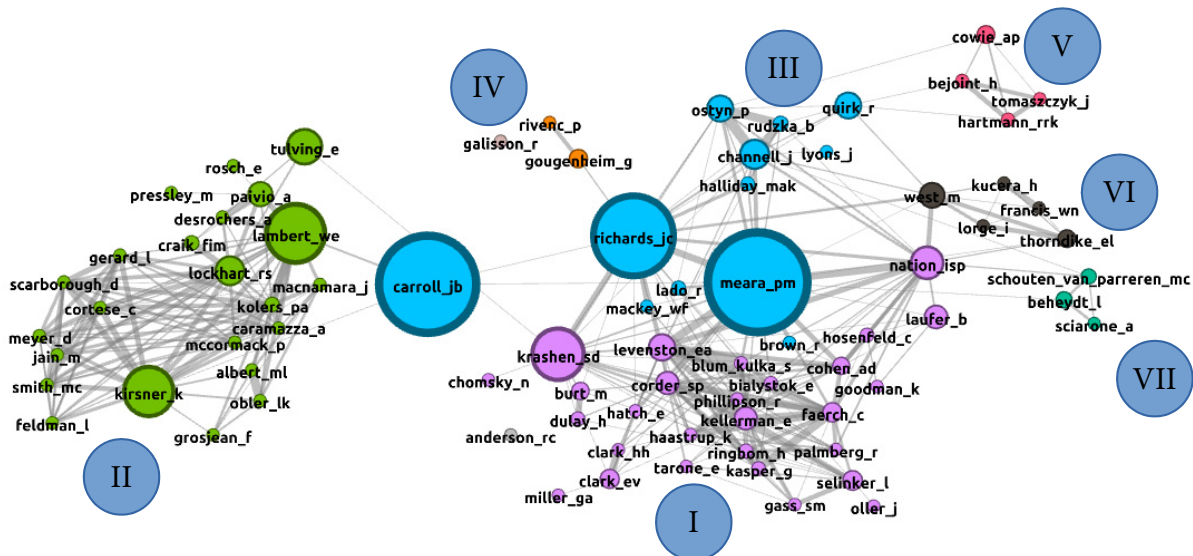


Figure 5: The “survivors”. Sources who appear in both the 1983–87 data set and the 1984–88 data set.

Figure 5 clearly shows that the number of survivor sources is surprisingly large. In fact, 80% of the sources in figure 5 also appeared in the equivalent map for 1983–87.

The 21 new entrants are shown in Figure 6. Most of these new entrants are single additions to existing groups. Cluster VI and Cluster VII have no new additions. The outstanding feature in this map is the very large new cluster focussed on Sinclair. In thematic terms, this group is a corpus linguistics cluster, concerned with the way corpora can be exploited in language teaching. However, it probably makes more sense to see this cluster as a UK-based geographical cluster. Two UK Universities contribute most of the new sources: Carter and McCarthy were both based at Nottingham University, while Sinclair, Renouf, Higgins and Johns all worked at Birmingham University. For those unfamiliar with UK geography, the distance between these two centres is a mere 80 kilometers, and several of the sources who make up this new cluster worked in both centres. Joanna Channell, a “survivor” who appeared as a significant source in this year’s map, was also associated with both centres during this period. Unlike some of the new developments we have seen in previous years, this one looks unlikely to peter out quickly, and we can probably expect a significant growth in this area in future years.

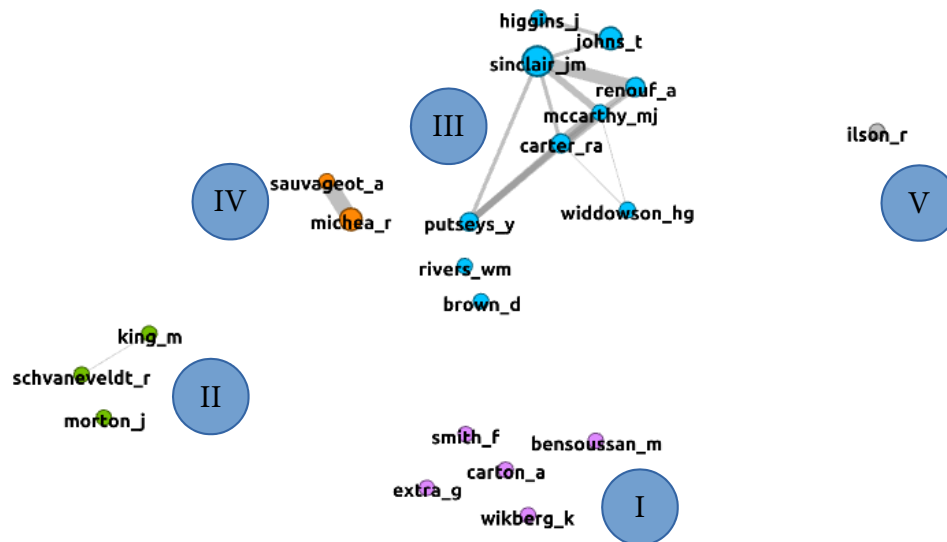


Figure 6: The new entrants in the 1984–88 map.

4. Discussion

A number of interesting ideas emerge from these analyses. The first idea is that the overall picture which emerges is much more stable than the pictures we have reported for previous years. Nonetheless, some changes to the field are evident, and the strong emergence of the new corpus linguistics cluster shows this very clearly. What is surprising about this new group is that it does not appear to be strongly co-cited with the other L2 vocabulary sources. We might have expected the corpus linguistics cluster to be strongly connected with the earlier word frequency count cluster, and with the dictionary use cluster, and though there are some links here, they do not appear to be very strong. If anything, the dictionary use cluster seems to have become more detached from the main L2 vocabulary cluster than before.

The second idea is that the research as a whole continues to be overwhelmingly focused on English research. A couple of small clusters comprising French and Dutch-speaking sources are in evidence, but by 1988 these clusters too have become more detached from the main L2 vocabulary research cluster. I think the main reason for this is simply that none of the sources in these clusters publish very much in English. In contrast, the main L2 vocabulary research does include a very large group of Scandinavian sources (Haastrup, Palmberg, Ringbom, Faerch, Kasper, Phillipson) who do publish in English. Collectively these sources make up a significant proportion of the Cluster I, and their emphasis on lexical inferencing is an important research theme at this time. The danger, of course, is that other researchers who also work in this area, but do not publish much in English (Schouten-van Parreren, for instance) are likely to get overlooked and squeezed out. The small *Français Fondamental* cluster (Cluster IV) looks as though it might be bucking this trend, but this cluster is really an internal dispute among French researchers about the value of the much earlier research on *disponibilité*. English language researchers (with the exception of Richards) seem to have missed the significance of this work.

The small cluster of German researchers which we noted in last year's report fails to make a mark on the 1984–88 map, despite the relatively large number of papers in German which

contribute to the 1984–88 data set. This is partly due to these papers following a different tradition of citation practice from the practices which are by this time becoming normal in the English language research. More importantly, perhaps, while the German researchers cite the main English language sources, English researchers, on the whole, cite each other frequently, but only rarely cite German sources. A good example of this problem is Zimmermann, who was identified in the 1983–87 data set as a prolific author with six publications. A further publication in 1988 brings his total to seven – more than enough for him to retain his place in the list of prolific authors. Altogether, though, these seven publications garner only ten citations in the data set (and seven of these are self-citations). This means that Zimmermann fails to appear in the 1984–88 map in spite of being a prolific author, and the related German research disappears with him. Clearly, publishing a lot of papers does not automatically mean that they will be widely cited.

The third feature which emerges from the 1984–88 map concerns the role played by a few key figures in the structuring of the map. We noted in last year's report that the key figure in the 1983–87 map was Krashen, with Richards, Meara and Lambert playing lesser roles. This judgement was based on a measure called **betweenness centrality**. The theory behind the betweenness centrality measure is that some nodes in a map are key to its structure because they provide links between the different clusters, and thus represent important points of contact between different research groups. The measure is based on the probability of a node being found along a path that links randomly chosen pairs of nodes. Nodes which appear frequently in these random paths score highly on the betweenness centrality measure, and when the clusters are highly divergent, the few nodes that link them score very highly on this feature. In practice, most sources tend to be co-cited with other sources in their immediate cluster, and only a few sources are co-cited with sources from two or more clusters.

In the 1984–88 map, there has been a significant shift in the make-up of these key nodes. Lambert continues to dominate cluster II, but is not generally cited in the broader L2 vocabulary acquisition literature. Krashen's influence has not entirely disappeared, but it is much reduced. Richards and Meara have become much more influential in the 1984–88 map: both are highly co-cited with the members of the new corpus linguistics cluster, and Richards' early work on *Français Fondamental* provides a strong co-citation link between Cluster II and the French researchers in Cluster IV. The surprise feature in the 1984–88 map is the importance of J. B. Carroll. In the 1983–87 map, Carroll is closely associated with a cluster that is focussed on word counts and dictionaries, but in the new 1984–88 map, he appears as the key link between Cluster II – the psycholinguists – and the rest of the network. My immediate reaction to this was that the analysis was over-estimating Carroll's real influence on the L2 vocabulary research, since, unlike the other key figures, he does not actually publish on L2 vocabulary acquisition during the period we are analyzing here. On reflection, however, I think this reaction was wrong. Twenty-two of the papers in the 1984–88 data set cite Carroll's work, but these citations are not limited to a single influential paper, or a paper that accidentally bridges the gap between the linguistic and the psychological approaches to L2 vocabulary. Most of the citations relate to Carroll's word frequency count work (Carroll, Davies and Richman 1971), or his Age of Acquisition norms (Carroll and White: 1973a, 1973b), but the citations are not limited to these useful tools. Particularly important are a 1964 paper published in the *Harvard Educational*

Review (Carroll 1964a) and the monograph *Language and Thought* published in the same year. (Carroll 1964b). In this case, the co-citation approach does appear to have successfully identified a significant influence who might otherwise have been overlooked.

Finally, it is worth noting that the 1984–88 map highlights the continuing decline in the influence of psycholinguistic research on L2 vocabulary research. Although Cluster II has seen the arrival of three new sources in 1988 (Morton, King and Schwaneveldt), it does not show any real growth in this time window. This cluster is slowly crystallising into three sub-themes (bilingual word recognition, the skills of bilingual speakers, and some residual work dealing with imagery and depth of processing). The rich connections in this cluster are at least partly due to differences in authorship practices in psychology and linguistics. Papers published in psychological journals are more likely to have multiple authors than are papers in linguistics. This results in dense clusters of co-citations, which make this work look more important than it might really be. The important links in a co-citation analysis are those which strongly link the different clusters: the 1984–88 map strongly suggests that cross-cluster links of this sort are in short supply.

5. Conclusion

The main point to emerge from the analysis presented in this paper is that L2 vocabulary research between 1984 and 1988 is remarkably stable, though we still find significant changes on a year by year basis. New research areas have added a degree of focus to the research published in 1988 year, and we have seen that the research is becoming more obviously dependent on linguistics for its main points of reference, whereas the research in earlier years cited a more eclectic set of sources. Some of the Significant Influences who played pivotal roles in our earlier analyses are beginning to be much less influential.

The next paper in this series will examine the research published in 1989 in the context of a five-year window covering 1985–89. Will the field continue to develop in the same way, or will we see significant shifts of direction in this period? It is hard to tell. All we can say at the moment is that by 1988 something that resembles a recognisable vocabulary research agenda is beginning to emerge.

References

- 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.
- Carroll, J. B. (1964a) Words, meaning and concepts. *Harvard Educational Review* 34(2): 178–202.
- Carroll, J. B. (1964b) *Language and thought*. New York: Prentice-Hall.
- Carroll, J. B., P. Davies and B. Richman. (1972). *The American heritage word frequency book*. Boston: Houghton Mifflin.
- Carroll, J. B. and M. N. White (1973a). Age-of-acquisition norms for 220 picturable nouns. *Journal of Verbal Learning and Verbal Behavior* 12(5): 563–576.
- Carroll, J. B. and M. N. White (1973b). Word frequency and age of acquisition as determiners of picture naming latency. *Quarterly Journal of Experimental Psychology* 25(1): 85–95.
- Corder, S. P. 1973. *Introducing applied linguistics*. Harmondsworth: Penguin Education. 1973.

- da Solla Price, D. (1965). Networks of scientific papers. *Science* 49(3683): 510–515.
- Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, 16(12): 317–324.
- 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 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. (no date) *VARGA: The Vocabulary Acquisition Research Group Archive*. Available at: <https://www.lognostics.co.uk/varga/> (accessed 22nd October 2020).
- Perdue, C. (1982). *Second language acquisition by adult immigrants: A field manual*. Strasbourg: European Science Foundation.
- Perdue, C. (ed.) (1993). *Adult language acquisition: cross-linguistic perspective*. Vol. 2. Cambridge: Cambridge University Press.
- Rudzka, B., J. Channell, Y. Putseys and P. Ostyn. (1981). *The words you need*. London: Macmillan.
- Rudzka, B., J. Channell, Y. Putseys and P. Ostyn. (1985). *More words you need*. London: Macmillan.
- 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 co-citation: A literature measure of intellectual structure. *Journal of the American Society for Information Science* 32: 163–171.

Acknowledgements

Thanks to Hilary Nesi and Reinhart Hartmann who provided some useful insights into the Dictionary research in 1988.

Thanks to Marlene Schwarz for help with sourcing the German literature.

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

Appendix 1: 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.

contributions	10	9	8	7	6	5	4	3	2	1
Expected E_N	2	3	4	5	7	10	16	28	140	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 1988 data when $x = 5.1$, 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.

Appendix 2: 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.

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.