Ground Zero: A bibliometric analysis of L2 vocabulary research 1986-1990

Paul Meara
Swansea University, United Kingdom
University of Oxford, United Kingdom

Abstract

This paper uses a co-citation analysis to examine the research on vocabulary acquisition that was published in 1990. Two analyses are presented. The first is a detailed account of the 1990 research on its own terms. The second analysis places this work in a larger context by looking at research published in a five-year window covering 1986-1990.

Keywords: L2 vocabulary acquisition, vocabulary research, bibliometric, author co-citation

1. Introduction

This paper is the tenth in a series of studies in which I have been mapping out the way L2 vocabulary research has developed over the last 50 years. Beginning with 1982, *LingBaW* has published a set of papers in which I have presented bibliometric mappings of the research that appeared in each year up until 1989 (Meara 2014–2021). This paper takes this historical overview another step forwards. It covers the research published in 1990 and recorded in the Vocabulary Acquisition Research Group Archive (VARGA. Meara: n.d.), a very large collection of papers that is not limited to the obvious English language sources.

The paper falls into two parts. Part 1 reviews the 1990 research in its own terms. Part 2 puts this research into a wider context by summarising the main trends that appear in a five-year window covering 1986-90. Both parts use the author co-citation method, developed by Small (1973). Small's method is described in detail in Appendix 1 for readers who are not yet familiar with the approach used in these papers.

2. Part 1. The new research published in 1990

At first sight, 1990 does not look like a good year for L2 vocabulary research. 1989 *was* a fairly good year in terms of the number and variety of papers that appeared. In 1990, however, we

have a return to a lower level of activity. The VARGA database (Meara n.d.) identified 142 relevant sources published in 1989. The equivalent figure for 1990 is only 113 published sources – the lowest total since 1985. The obvious interpretation of this decline is that the surge in research that began in the early 1980s has now peaked. However, this interpretation turns out to be somewhat over-simplified. The main reason for this is that 1990 actually sees a significant shift in the type of work being published. A large proportion of the 1989 output appears as book chapters – with the collection by Tickoo (1989) alone accounting for a significant fraction of the total output. There are fewer special issues of journals and edited volumes in 1990, and as a result both the number of book chapters and the number of ordinary papers has declined in this year's output (cf. Figure 1).
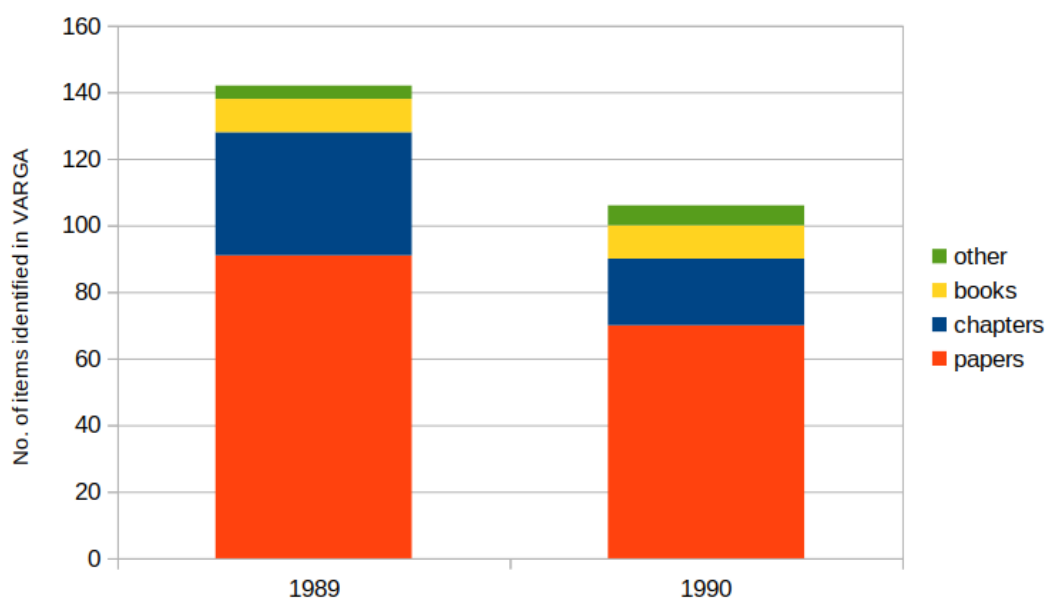

Figure 1: *The 1990 research by item type*

The outstanding publishing event of 1990 is the appearance of Paul Nation's book *Teaching and Learning Vocabulary* (Nation 1990). As we shall see in further papers in this series, it is difficult to overstate how influential this book would turn out to be. Basically it defined a research agenda that came to dominate the field for the next two decades and beyond. Nation's book was not universally welcomed at the time – partly because it shifted the research away from the dominant linguistic areas of research to more psychological aspects of vocabulary – reading, inferencing, learning, testing – which had played only a limited part in the growth of vocabulary research in the previous few years. It is not often that a single book revolutionises a field, but this book is definitely one of them, and we will see its influence coming to dominate the field in later papers in this series.

However, Nation's work was not the only book length treatment to appear in 1990. McCarthy's *Vocabulary* (McCarthy 1990) is a more traditional volume, that focusses more on the linguistic features of vocabulary. The book falls into two halves. Section one (Explaining vocabulary) deals with word-formation, lexical relations, prototypical vocabulary, vocabulary in use, and measurable characteristics of words in texts. These are all features that rely heavily on the corpus research that we noted in the 1989 research. Section two (Demonstrating

vocabulary) is more concerned with vocabulary teaching and the role of the teacher in this process. It covers vocabulary selection, organising vocabulary in a non-random way, presenting vocabulary in classrooms, vocabulary learning strategies and lexical reference (dictionaries and similar support material).

Taylor's book, also called *Teaching and Learning Vocabulary* (Taylor 1990), covers much the same ground as the second half of McCarthy's text. A general introduction to words is followed by chapters that cover communicative vocabulary teaching, repetition and interaction, and exercises for consolidating newly learned vocabulary. The book ends with a short chapter on vocabulary in discourse.

A radical departure was Willis' *The Lexical Syllabus* (Willis 1990). Like McCarthy (1990), this book also leans very heavily on research carried out by the COBUILD team at Birmingham University. It argues that most EFL text books pay too much attention to grammar, and not enough attention to words, and suggests that much of the grammatical structure of English can be found by studying the behaviour of words in a corpus. This was not just a question of emphasis: Willis argues that it radically changes the roles of teachers and syllabus designers, and that it encourages learners to become active explorers of a language, rather than passive recipients of their teachers' wisdom.

Ljung's book *A study of TEFL vocabulary* (Ljung 1990) also relies heavily on a corpus based approach to vocabulary, though unlike the three previously mentioned texts, it is mainly concerned with a specific set of textbooks. The book falls into two parts: Part One is a detailed critical analysis of the vocabulary that appears in a set of 56 English language textbooks aimed at secondary school pupils in Swedish schools. This analysis is actually quite short – only 40 pages. Part Two – just short of 400 pages contains detailed word lists that inform the discussion in Part One. This is an unusually thorough analysis of the vocabulary taught in textbooks, and it deserves to be much better known than it appears to be. (At the time of writing, *Google Scholar* lists fewer than 100 citations of this work, compared to 348 for Taylor, 1242 for Willis and 2191 for McCarthy.)

A number of other titles are worth mentioning briefly. Dretzke and Nester (1990), Tréville (1990) and Milan and Sunkel (1990) rehash some familiar discussions about cognates and false friends.

Diab (1990) is a longer version of a study that was first reported in 1989. It looks at how nurses access ESP vocabulary by using dictionaries. Murphey (1990) analyses the lyrics of pop songs, and discusses how they can be used to teach words to learners of English. Schrameier (1990) and Willems and Oud-de Glas (1990) were unobtainable due to COVID restrictions in place at the time this work was being carried out.

Four cited theses dealing with L2 vocabulary acquisition were published in 1990. These are listed in Table 1.

Table 1: Theses published in 1990 that are cited in later years

Altarriba, J
Constraints on interlingual facilitation effects in priming in Spanish-English bilinguals. PhD Dissertation. Vanderbilt University. 1990.
Cunningham, L
L2 vocabulary: A study of the word association responses of beginning learners of Irish. MPhil Dissertation. University of Dublin. 1990.
de Jong, E
Woorden leren met prentenboeken. Een vergelijking van twee instructiemethodes. PhD thesis. Katholieke Universiteit Brabant.– Tilburg. 1990.
Dekkers, R
Woorden schieten tekort. Een onderzoek naar de optimalisering van woordenschatverwerving bij allochtone kinderen. PhD thesis. Katholieke Universiteit Brabant. Tilburg, 1990.

2.1. The data sources

As usual in these reports, the analysis presented in this section is based on papers published in journals or as book chapters. The monographs and theses listed in the previous section are not included in the analysis as these longer works tend to use citations in ways which differ from what we find in shorter research papers.

The VARGA data base (Meara: n.d.) lists a total of 95 sources eligible for the analysis, but a surprisingly large number of these items were not available – partly because of COVID travel restrictions and library closures at the time I was collecting this data. These items – 17 in total – were excluded from the analysis, but they are listed here in Table 2.

It is unlikely that the exclusion of these items significantly affects the analysis that follows, but a number of points are worth noting. Firstly, several of these items are German sources, and their appearance in Table 2 reflects the ongoing library issues arising from the political events in Germany in 1989. Secondly, the paper by McCarthy was one of several short think-pieces that appeared in a thematic volume of *English Studies* published by the British Council. These papers are best described as and somewhat eclectic, in their coverage, but they probably reflect a growing interest in vocabulary in the British Council, an influential player in English Language Teaching at the time. Thirdly, Hayes (1990) is the first paper dealing with acquisition of Chinese that we have identified in this series of analyses. Fourthly, the omission of Lado is probably the most serious omission. Lado wrote a series of important vocabulary research papers in the1960s and 1970s where he looked at "massive vocabulary expansion". At the time vocabulary acquisition was not a significant topic of research and experimental studies of vocabulary acquisition were not fashionable with the result that this work largely went unnoticed. (Google Scholar indicates that Lado's *Final Report* is cited only two times.) Finally, the omission of Zimmermans' paper in Milwaukee Studies significantly reduces the importance of this German strand of research.

The remaining 78 studies are not listed here for space reasons, but interested readers can find the complete list on the VARGA web-site: *https://www.lognostics.co.uk/varga/*. Entering **1990 ##** in the search box will return a complete list of the papers that are included in this year's analysis. Although the number of eligible papers is considerably lower than the figure of

118 papers that we identified in 1989, it is actually close to the number of papers and book chapters published in 1988 (83). This suggests that the relatively large size of the 1989 data set might be an unusual blip in the numbers, rather than an indicator of a reliable upward trend in the number of outputs.

Table 2: *Eligible items published in 1990, but not included in the formal analysis*

Abe, H. 1990. The acquisition of communicative vocabulary: A case of procedural knowledge. *FL Reporter,* 22:33-45.

Anon. 1990. Wortsschatzarbeit = Vokabellernen? *Neusprachliche Mitteilungen,* 43:101-103.

Appel, R. and A. Vermeer. 1990. Woordenkennis sleutel voor schoolsucces: uitbreiding woordenschat in kleuterklassen. *Stimulans,* 8,7:6-8.

Fanning, H. 1990. Falsche Freunde in zweisprachiger Erziehung (Englisch/Deutsch). In: H. J. Zobel (ed.) *Studien zur Sprachdifferenzierung: Roland Arnold zum 60. Geburtstag.* Wissenschaftliche Beiträge der Ernst-Moritz-Arndt-Universität Greifswald. 47-51.

Farrell, P. 1990. Vocabulary in ESP: a lexical analysis of the English of electronics and a study of semi-technical vocabulary. *CLCS Occasional Paper* No. 25. Trinity College Dublin.

Hayes, E. 1990. The relationship between 'word length' and memorability among L2 readers of Chinese Mandarin. *Journal of Chinese Language Teachers Association,* 25,3:31-41.

Hinz, K. 1990. Wortschatzarbeit in der leitungschwachen Englischklasse. *Der fremdsprachliche Unterritcht,* 23: 22-26.

Kühn, P. 1990. Das Grundwortschatzwörterbuch. In: FJ Hausmann, O Reichmann, HE Wiegand and L Zgusta (eds.) *Wörterbücher/Dictionaries/Dictionnaires. Ein internationales Handbuch zur Lexikographie, 2.* Teilband. Berlin. 1353-1364.

Lado, R. 1990. Towards a lexico-semantic theory of language and language learning. *The Georgetown Journal of Language and Linguistics,* 1,1: 96-100.

Lipczuk, R. 1990. Über ein Wörterbuch der faux amis. In: A Kątny (ed.) *Deutsche Sprache im Kontrast und im Kontakt.* Rzeszów. 225-237.

McCarthy, M. J. 1990. Minding your words: two important areas of vocabulary learning. *English Studies,* 5:6-7.

Quetz, J. 1990. Wortschatzarbeit. *Zielsprache Englisch,* 20:29-31.

Reiner, E. 1990. "False friends" und "doublets" JA Pfeffers *Germanismenwörterbuch der englischen Sprache. Moderne Sprachen,* 34,1-2:41-48.

Scherfer, P. 1990. Vom Nutzen des Vorwissens im Vokabelunterricht. *Der fremdsprachliche Unterricht,* 24:30-36.

Vater, B. 1990. Schuler verlangen nach erfullter Gegenwart! Wortschatzarbeit in einem handlungsorientierten Englischunterricht in der Sekundarstufe 1. *Der fremdsprachliche Unterricht,* 23: 11-16.

Zimmerman, R. 1990. Productive and unproductive lexical strategies in L2 learners. *Milwaukee Studies on Language.* 1990.

We begin by reporting the usual superficial analysis of authorship. A total of 94 authors make a contribution to the 1990 output. Again, this number is considerably down on the 1989 figures, where 134 unique authors were identified, but close to the figure that we reported for 1988 (83 unique authors). In 1990, the proportion of authors contributing to just one paper is 88%. Table 3 shows the distribution of authors in 1989 and 1990 in terms of the number of contributions that they make to the two data sets.

Table 3: *The number of authors contributing to N outputs in the 1989 and 1990 data sets*

Contributions	6	5	4	3	2	1
1990 data			2	0	6	87
1989 data			1	4	18	109

The best interpretation of these figures seems to be that that the proportion of authors contributing to only one paper has risen slightly in the 1990 data set (85% compared with 82% in 1989).

The number of authors contributing to two or three papers has fallen dramatically, but there is a small rise in the number of authors contributing to 4 papers in this data set. Laufer and Meara both contribute to four papers; Appel, Broeder, Colpaert, Decoo, Schouten-van Parreren and Swartz contribute to two papers each. It is worth noting that the first five of these authors all worked in the Netherlands, underlining the importance of vocabulary research in that country at the time. Swartz is a scholar working on CALL at the US Army Research Institute, Alexandria. Of the authors who contributed two or more papers in 1989, most have fallen out of the major contributors list for 1990. Of the 1990 authors, only Meara and Laufer were also significant contributors to the 1989 output and all the authors of two papers are new entrants into this list. This is a surprisingly high level of churn, and it suggests that the field as a whole is still a long way from settling down into a steady pattern of outputs.

2.2. The analysis

The main analysis reported in this section is not directly concerned with who published in 1990, but rather with the sources that are cited in the 1990 data set. Cited authors provide us with clues as to the ideas that were important at the time, and allow us to plot the growth of research trends. The first step in our analysis, then, is to identify the important sources that are cited in the 1990 data set. The methodology for doing this has been described in detail in the previous papers in this series (see Appendix 1). For 1990, the analysis identifies 1363 sources – very close to the number we identified in the 1988 data set (1391), but considerably fewer than we identified in the larger 1989 data set (1911). As usual, most of these sources are cited only once, but a small number of sources are cited much more than this. The data is summarised in Table 4.

Table 4: The number of times sources are cited in the 1990

frequency	25	24	23	22	21	20	19	18	17	16	15	14
cases							1					
frequency	13	12	11	10	9	8	7	6	5	4	3	2
cases			1	3	2	4	5	5	18	25	55	168

The most cited sources in this data set are Meara (19 citations), Nation (11 citations), Carter, Cohen and Krashen (ten citations each), Richards and Laufer (nine citations each), Palmberg, McCarthy, Lockhart, and Levenston (eight citations each). Most of the sources that appear in this list are new: only Nation, Meara, Carter and Levenston were also highly cited in the 1989 data. Direct comparisons between the 1990 data set and the 1989 data set are difficult because the 1990 dataset is a lot smaller than that of the previous year. Inclusion in the ten most cited list for 1989 required a source to be cited at least 11 times, whereas the equivalent threshold for 1990 was only eight citations. Meara's 19 citations (22% of the total data set) is striking, as it exceeds the previous "record" set by Nation in 1989.

The next step in our analysis of the 1990 data is to identify the most frequently cited sources in the data set. Once these sources have been identified, we can construct a map based on the co-citations between them. The convention here is to identify a set of about 100 sources, but for 1990 it is difficult to do this. For the 1989 data set, we adopted an inclusion threshold of 5 citations, and this gave us a set of 80 sources to map. With the 1990 data, adopting this threshold would leave us with only 43 sources to map. Lowering the threshold to four citations adds a further 25 sources, bringing the total to 68 cases, and this is as close as we can get to the 1989 figures. In percentage terms, a source that is cited four times in this data set is cited in nearly 5% of the papers published in 1990.

The co-citation data for the 68 most cited sources in the 1990 data set were extracted from the complete data set (all 1363 sources cited in 1990), and the results analysed using the Gephi software package (Bastian, Heymann and Jacomy 2009). Gephi's analysis of these co-citations is shown in Figure 2. The interpretation of this map is relatively straightforward. Gephi identifies seven clusters in this data set.
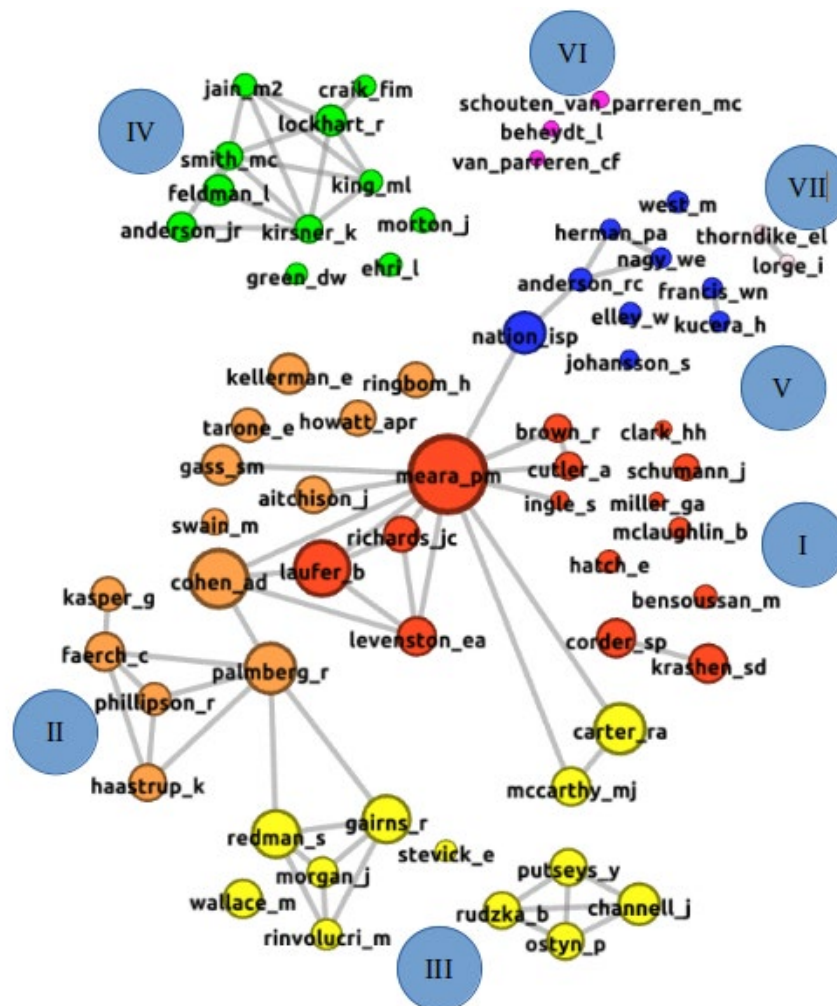

Figure 2: *A co-citation map of the 1990 data set. The map contains 68 nodes, each representing a source that is cited at least four times in the data set. The number of citations is reflected in the size of the node. Edges between nodes indicate that the two nodes are co-cited at least four times. The nodes are grouped into clusters taking into account all co-citation links that appear at least twice in the complete data set. Weaker edges have been removed from the map in the interest of simplicity and clarity.*

The core of the map is made up of **Cluster I** (15 sources, centred on Laufer, Meara, Richards and Levenston). These sources are perhaps best described as the current mainstream of L2 vocabulary research, but the sub-cluster made up of Corder and Krashen is influential too. The detached sources in this cluster signal psychological influences on the L2 vocabulary research. Brown is an important figure in L1 acquisition; Hatch and Clark work in the area of semantics; Schumann represents a strand of neurolinguistic research; Miller is best known in this context for his work on the limitations of short term memory; Cutler is best known for her work on slips of the tongue; Bensoussan is closely associated with Laufer; McLaughlin (1987) is a psychologically oriented text book dealing with L2 acquisition.

Cluster II is a group of mainly European researchers who are loosely interested in the psycholinguistics of L2 vocabulary acquisition. The core of this cluster is a group of Scandinavian researchers that we have identified in our earlier analyses (Faerch and Kasper, Haastrup, Phillipson, Palmberg and Ringbom). Andrew Cohen is the most cited source in this cluster, mainly cited for his interest in mnemonic approaches to vocabulary acquisition (e.g. Cohen 1987).

Cluster III is a loose group of sources who are mainly concerned with vocabulary pedagogy. We can identify three sub-clusters here. McCarthy and Carter represent input from corpora and discourse analysis (e.g. Carter and McCarthy 1988); Rudzka, Ostyn, Putseys and Channel authored a series of influential textbooks that focussed on semantic relations between words (e.g. Rudzka et al. 1985); Stevick (1976) is an important methodological source; Gairns and Redman (1986), Wallace(1982) and Morgan and Rinvolucri (1986) are methodology textbooks mainly aimed at teachers. With the exception of Stevick, all the sources in this cluster are UK based.

Cluster IV is the now familiar group of cognitive psychologists who work with experimental studies of bilinguals. This cluster is detached from the other clusters in the map. The sources here are not really interested in the pedagogical aspects of L2 vocabulary acquisition, though some of this work – notably Craik and Lockhart's (1972) paper, and Craik and Tulving's 1975 paper – are strongly cited in the L2 literature.

Cluster V, centred on Nation, is principally concerned with reading in an L2. Nagy, Herman and Anderson are important L1 reading theorists. Kucera and Francis, Johansson and West are word lists.

Cluster VI is a group of Dutch speaking vocabulary researchers.

Cluster VII (Thorndike and Lorge) is another word frequency count, and is probably best treated as an extension of the word counts in Cluster V.

Most of these clusters will be familiar from the maps published earlier in this series of papers.

Figure 3 shows the changes that have taken place between the 1989 data set and the 1990 data set.
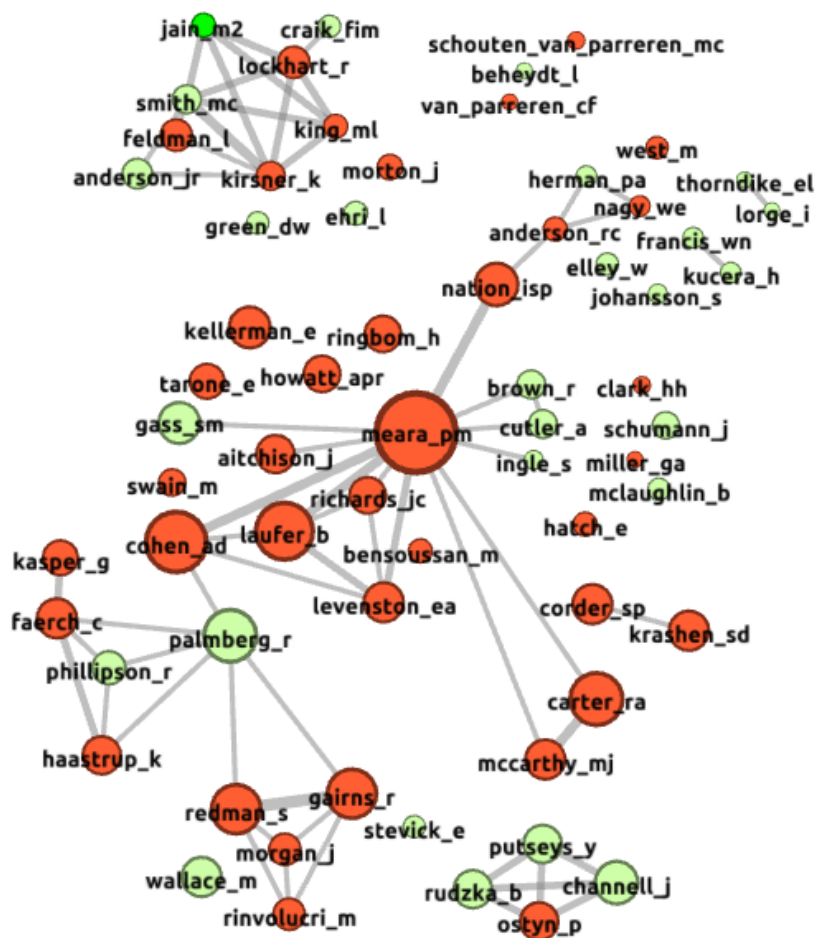
Figure 3: *New entries in the 1990 data set (shown in green)*

The figure preserves the cluster layout of the 1990 data set, but uses colour to show the appearance of new sources. The 1990 "survivors" – sources that appear in both the 1989 and 1990 data sets are shown with dark shading. There are 35 survivors in 1990, about 54% of the total, indicating a fair degree of stability in this data set. The 30 new cases – sources who did not appear in the 1989 data set – are shown with light shading. However, many of these apparently new entries had already appeared in our earlier maps, and are best described as "returners" rather than new entries. Only seven genuinely new entries appear in the 1990 data set: Cutler, Ingle, Schumann and McLaughlin in Cluster I, Phillipson in Cluster II, Stevick in Cluster III and Green in Cluster IV. Clusters V, VI and VII consist entirely of familiar sources. Cutler represents an important strand of L1 research on speech errors. Schumann is mainly cited for his work on neurolinguistics. McLaughlin's textbook, as we have already noted, is more psychologically oriented than other works of this type. These new sources may indicate that Cluster I is becoming more oriented towards the psycholinguistics of L2 vocabulary development, while the more traditional linguistic influences are moving to other clusters, and perhaps becoming less influential. Phillipson, in Cluster II, is working closely at this time with Haastrup, and represents a significant strengthening of the Scandinavian sub-cluster that we have noted in previous years. Stevick, in Cluster III, is a methodological textbook. Green, new in Cluster IV, signals an additional research theme among the psycholinguistic sources.

Green is mainly cited for his work on the way bilinguals control access to their two languages (e.g. Green 1986).

The most striking feature of the 1990 data set, one which is not immediately apparent in the maps, is that two of the clusters that made up the 1989 map have almost completely disappeared in 1990. The most important cluster in 1989 was a densely connected dictionaries/semantics cluster. Surprisingly, all the members of this cluster fail to appear in the 1990 map. The implications of this change are not entirely clear. It is possible that the dictionary theme has played itself out by 1990, but this interpretation will need to be confirmed in later maps. The 1989 map also contained a small cluster whose focus was best described as L2 vocabulary learning in German and French. This cluster too has disappeared from the 1990 map. A small L1 acquisition cluster from 1989 has been reduced to a single node in 1990 (Clark in Cluster I).

We also find some smaller changes in the 1990 map. There is a suggestion that vocabulary course books (cluster III) are becoming more coherent as a feature, but less strongly connected to the empirical and theoretical work thinking that characterises Clusters I and II. There is also a suggestion that the psycholinguistics cluster has grown in 1990, and is more diverse in its membership than it was in 1989. It is not clear whether this is a genuine shift, or just a reflection of different authorship patterns in psycholinguistics – papers in this area tend to have multiple authors, who all cite each other, and this can sometimes inflate the importance of a source. Cluster VI, the Dutch/Flemish cluster, appears in the 1990 map as a detached cluster, no longer widely co-cited by the key sources in Cluster I and Cluster V.

3. Part 2. A wider perspective 1986-90

In this section, we will place the 1990 data into a larger context by looking at a five-year window covering all the research published between 1986 and 1990. Working with a five year window smooths out some of the short-term fluctuations in the annual data, and the results are less obviously affected by the size of the individual data sets.

For the purposes of comparison, Table 5 summarises the main features of the 1985-1989 data.

Table 5: *The main characteristics of the 1985-1989 data set*

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

Table 6, summarises the main characteristics of the 1986-1990 data set. Here, the 1985 data has fallen out of the five year window, and the 1990 data has been added to the window. Because 1985 was a (relatively) good year for L2 vocabulary research, and 1990 saw a slight fall in the number of publications, the 1986-1990 data set is marginally smaller than the 1985-1989 data set that we reported in last year's analysis, and the number of authors contributing to the 1986-90 data set is a lot smaller. In other respects the two data sets are comparable.

Table 6: *The main characteristics of the 1986-1990 data set*

Number of papers in the data set	465
Number of authors contributing to the data set	411
Number of sources cited in the data set	4699
Inclusion threshold for this data set	14 citations
Number of cited sources meeting the inclusion threshold	99
Identifiable co-citation clusters	6
I: vocabulary acquisition (41)	
II: dictionaries, corpus analysis and semantics (25)	
III: L2 word recognition, performance of bilinguals, imagery (20)	
IV: Reading and word frequency counts (7)	
V: Dutch vocabulary research (4)	
VI: French vocabulary research (1)	

Table 7 shows the number of contributions each of the 411 authors made to the 1986-1990 data set.

Table 7: *The number of authors contributing to N papers in the 1986-1990 data set, and the expected number of authors based on Lotka's Law*

Papers	19	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
Authors	1				1		1	1	2		2	6	9	16	53	335
Lotka	0	1	2	2	2	3	3	4	5	7	9	13	21	37	84	335

Once again, we find that the vast majority (82%) of the authors contribute to only one paper. This figure is slightly larger than the equivalent figure for the 1985-1989 window. However, the number of authors contributing to multiple publications still falls short of what we would expect to find. 14 authors contributed to five or more publications in the 1986-1990 window. Thirteen of these authors were already identified as prolific authors in the 1985-89 data set. Here, their status appears to be consolidated: Meara (19 papers), Zimmerman (12 papers), Laufer (10 papers), Palmberg (9 papers), Broeder and Carter (8 papers each), Beheydt (6 papers), AD Cohen, Extra, Nation, Robinson, van Hout and Vermeer (5 papers each). Schouten-van Parreren (6 papers) is the only new addition to the most prolific author list. Two prolific authors from the 1985-89 have fallen off the list in the 1986-1990 window: McCarthy contributes to 4 papers in the 1986-90 window, and Alfes contributes to two papers in this data set.

The bottom line of Table 7 shows the number of authors we would expect to contribute to N publications in a data set of this size. These figures are based on a suggestion by Lotka (1926). Lotka's method is outlined in Appendix 2 for readers who are not already familiar

with this approach. The table shows that, given 335 authors who contribute to just one output, we might expect 84 authors who contribute to two outputs, 37 authors who contribute to three outputs, and so on. The table shows that the actual figures fall quite a long way short of Lotka's projection for this data set. Meara is the only author who bucks the general trend at this time.

Table 8 summarises the number of times each source is cited in the 1986-90 data set. As usual, most sources are cited only once in the complete set of 465 papers – 2967 authors, or 63% of the total of 4699 sources fall into this category. However, there is some evidence that the number of sources who are very frequently cited is increasing: the most frequently cited sources are Meara (cited in 72 papers), Nation (60), Krashen (51), Levenston (50), Faerch and Richards (46), Carter and Sinclair (37), Kasper (36), Kellerman, Lockhart and Schouten-van Parreren (35 citations each). Most of these names will be familiar from our earlier analyses. This list is largely identical to the list of highly cited sources in the 1985-89 data set, but four new sources appear in the 1986-90 data set – Carter, Sinclair, Kasper and Lockhart – replacing four sources who have fallen out of the 1985-89 data set – AD Cohen, West, Blum-Kulka and Corder.

Table 8: *The number of times sources are cited in the 1986-90 data set*

FREQUENCY	60+	59	58	57	56	55	54	53	52	51	50	49	48	47	46
CASES	2									1	1				2
FREQUENCY	45	44	43	42	41	40	39	38	37	36	35	34	33	32	31
CASES									2	1	3	1	2	1	
FREQUENCY	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16
CASES	2	1		3	1	2	4	6	3	8	7	7	6	10	6
FREQUENCY	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
CASES	5	10	14	26	24	25	35	41	44	89	113	181	327	673	2967

Next, we turn to an analysis of the co-citation patterns among the most frequently cited sources. Standard practice in this type of analysis is to work with the 100 most frequently cited sources, and Table 8 shows that we can use a threshold of 14 citations to identify a suitable set of 99 sources for the analysis that follows. This threshold is identical to the threshold we adopted in last year's analysis.

Data for these 99 sources was analysed using the Gephi package. The results of Gephi's analysis are shown in Figure 4, which maps the 99 most frequently cited sources in the 1986-90 data set, using 14 citations as a threshold for inclusion. The sources are linked by 575 edges each of which occurs at least six times. However, in the interests of simplicity, Figure 4 shows only those edges which occur eight or more times. Gephi finds six clusters in this data set.
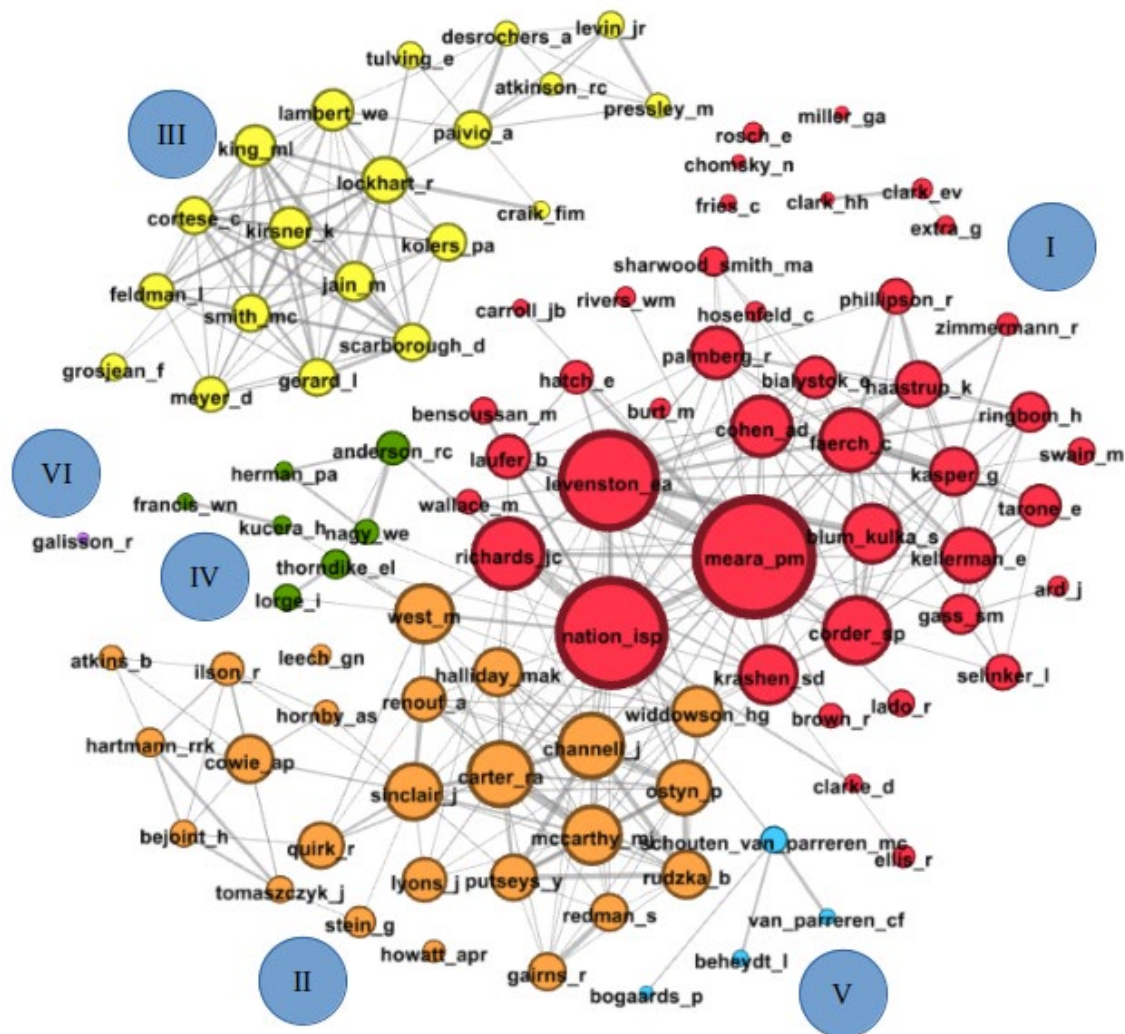
Figure 4: *A map of the co-citations found in the 1986-1990 data set. Nodes are sources that are cited at least fourteen times in the data set. Co-citation links that occur less than eight times in the data set have been eliminated in the interests of simplicity.*

The four most cited sources in the dataset (Meara, Nation, Levenston and JC Richards) form the core of **Cluster I.** This cluster is best described as the mainstream L2 vocabulary cluster at this time. It consists of 36 sources that are strongly co-cited, a small detached subcluster that is clearly a set of L1 acquisition sources (H Clark, E Clark and Extra) and a small number of detached outliers (GA Miller, cited most often for his work on memory chunking; Rosch, cited most often for her work on semantics; Fries, cited most often for his work on semantic frames; and Noam Chomsky as the goto reference for generative grammar.) A number of other sub-clusters can also be identified here. The most important of these is the Scandinavian group (Faerch, Kasper, Haastrup, Phillipson, Ringbom and Palmberg). We have noted the significance of this group in our earlier maps, but by 1990 this group is becoming a key feature of the L2 vocabulary research enterprise. The Faerch/Kasper co-citation, for example, is by far the strongest co-citation in this data set (34 occurrences). We can also identify an influential Israeli sub-cluster (Levenston, Laufer, AD Cohen, Blum-Kulka and Bensoussan), a small Edinburgh sub-cluster (Corder, Selinker), and a subcluster consisting of

Krashen and Burt. These last two sub-clusters were important in previous maps, but their influence here seems to be in decline.

Cluster II is the now familiar Dictionaries and Semantics cluster that has also appeared in our earlier maps (24 sources + 1 detached outlier, Leech). Three sub-clusters can be identified here, one dealing specifically with dictionaries (Hartmann, Stein, Tomaszczyk, Cowie), a second that deals more broadly with corpora-based approaches to vocabulary (Quirk, Sinclair, Carter, McCarthy and Halliday) and a third subcluster consisting of important textbooks (Gairns and Redman, Channel, Ostyn, Rudzka and Putseys). A cluster of this type was clearly identified in our analysis of the 1985-1989 data: here it has shrunk slightly. The distinguishing characteristic of this cluster is that its members are co-cited with Nation and JC Richards, but only rarely with the other sources who make up Cluster I.

Cluster III is the familiar group of psychologists and psycholinguists whose work influences L2 vocabulary research (20 sources). At first glance, this cluster appears to be about the same size as the psycholinguistics group we identified in last year's map. However, its composition has changed a bit. Sources that were previously important in this cluster, notably Lambert and Tulving, now seem to be less so. A couple of sources have dropped out of the cluster (Albert and Obler, Macnamara, and Caramazza), and their place has been taken by a group of sources who work on imagery that formed an independent cluster in the 1985-1989 map. The cluster continues to be dominated by a team led by Kirsner.

Cluster IV, also familiar from previous years, is mainly composed of frequency counts. It contains a subcluster of L1 reading specialists who are strongly co-cited with Nation, but not with other members of Cluster I.

Cluster V is a group of L2 vocabulary researchers who publish mainly in Dutch. The key figure in this group is Schouten-van Parreren who is often co-cited with Nation, but only infrequently co-cited with other sources in Cluster I.

Galisson stands as the sole member of **Cluster VI.** He is the one remaining representative of French vocabulary research in this map, but his work is not well-integrated with that of the other sources.

4. Discussion

Sharp-eyed readers will have noticed that the map in Figure 4 looks rather different from the five-year maps in our previous analyses, which were becoming increasingly complex and difficult to interpret. Gephi offers a number of different map lay-outs, and Figure 4 is computed using the Fruchterman-Reingold algorithm, which places well-connected nodes at the centre of the map, and less well-connected nodes towards the periphery. Hopefully, use of this format will allow future analyses to be compared more easily with earlier maps, by introducing a greater degree of standardisation to the maps. For example, the Fruchterman-Reingold layout allows the size of the maps to be standardised in a way that was not possible with the layouts that we have been using up till now. Furthermore, it is possible that this format will make it easier to carry out some further analysis that takes greater account of the strength of the connections between the more important nodes. These features will be

exploited in future analyses. (See https://www.lognostics.co.uk/maps/ for a further discussion of this issue.)

Meanwhile, a number of themes are beginning to emerge from this analysis of the 1986-1990 dataset. Firstly, and most importantly, the main L2 vocabulary cluster has grown substantially, mainly by absorbing what appears as a separate cluster in the 1985-1989 map. The cluster has almost doubled in size by 1990, and now accounts for about half of the nodes in the map. The members of this cluster are very strongly cited alongside each other, and almost all of the members of the cluster can themselves be identified as active L2 vocabulary researchers. I think this is a sign that the L2 vocabulary research is, by this time, becoming self-sufficient, and less reliant on external influences. The few external influences that remain are clearly becoming less central to the structure of this cluster: L1 vocabulary acquisition has almost become detached (Clark and Clark); error analysis (Corder and Selinker) is clearly marginal; Krashen's influence on the field also seems to be on the decline. At the same time, this cluster has started to become isolated from the rest of the map. This is not immediately apparent from Figure 3, but a closer look at the data suggests that Nation, Levenston, Meara and Richards are the only sources who have strong connections outside the cluster. The implications of this drift are not entirely clear.

The second theme to note concerns Cluster II, the dictionaries, corpora and semantics cluster. The membership of this cluster is almost unchanged in the five year map. It is worth remembering, however, that this theme no longer has a strong presence in the one-year 1990 map, so the strength of this theme might be expected to diminish in future years. Meanwhile, Whitcut has dropped out of the cluster, Kucera and Francis have moved to another cluster, and the cluster has gained three new sources in their place (Gairns and Redman, and Howatt). The cluster as a whole is notably self-contained: members of this cluster are co-cited with the main sources in Cluster I, but there are few other strong co-citations. This level of stability in a cluster is unusual, and it strengthens the view that research in this area may have peaked.

The third theme emerging from this map concerns the role of national groupings. The main feature here is that the Français Fondamental cluster no longer figures in the 1986-90 map. Robert Galisson (Cluster VI) remains as the sole representative of this important French research strand (though he is actually quite critical of the approach that the Français Fondamental project took). Likewise, the German research that we noted in our earlier maps is again represented only by Zimmerman, who plays a minor role in Cluster I. Dutch vocabulary research has become detached from the main L2 research cluster, and now stands alone as Cluster V in the 1986-90 map. The increasing dominance of English language research – both in the sense of research written in English and research about learning English vocabulary – is very clear here. Other languages barely get a look in.

Finally, the main fault line in the 1986-90 map lies between Cluster III and the other clusters. The map suggests that the co-citation links between sources in Cluster III and the other sources in the map are non-existent. In reality, there are some weak links, but they are not strong enough to show up here. The map also suggests that the composition of this cluster is stable: the cluster contains 19 members in our 1985-1989 map, and 20 in the 1986-1990 map. However, a number of features of the new Cluster III are worth commenting on. The key figure in this cluster is Paivio, whose dual-coding approach to bilingual memory provides

a unifying theme for work on mnemonics and bilingual storage. The mnemonics and imagery sub-cluster that is built around Paivio is largely based on research that was carried out in the late 70s and early 80s. This work has consistently failed to find a home in mainstream L2 vocabulary research, but it shares little with the rest of the psycholinguistic cluster other than an emphasis on experimental methods. At the moment, the position of this sub-cluster looks precarious, and it would not be surprising to find that it disappears from our maps in the future. The second sub-cluster in Cluster III seems to be more robust. Kirsner's group (King, Jain, Lockhart, Smith) and Scarborough's group (Scarborough, Cortese, Gerard) both use lexical decision tasks as a way of investigating bilinguals' mental lexicons. Most of the citations to these sub-clusters appear in the latter half of the 1985-90 window, so it looks as though they will persist into the future. In contrast to Paivio, Grosjean, whose work at this time is largely concerned with code-switching, looks rather isolated at the edge of Cluster III. He is frequently cited alongside the other members of this cluster, but only rarely cited by them.

Notably missing from Cluster III are Albert and Obler. Their 1978 edited collection, *The Bilingual Brain,* (Albert and Obler 1978) was an important influence throughout the early part of the 1980s, but it has only one citation in 1990 (Channel 1990) suggesting that interest in neurolinguistic research on how bilinguals handle L2 words is rapidly falling off.

Overall, the 1986-90 map shows a very high degree of stability. Ninety sources that appeared in the 1985-89 map also appear in the new map; only nine new sources make an appearance. Five of these – Ard, Burt, R Ellis, Fries and Sharwood-Smith – are additions to Cluster I. Two of the new sources appear in Cluster IV (Herman and Lorge); Howatt is a new addition to Cluster II; Bogaards is an addition to the Dutch research group (Cluster V).

5. Conclusion

The analyses in this paper suggest that the consolidation we noted in earlier reports is continuing into 1990. There are some small changes in the prolific author lists, and the relative importance of the major sources, but these changes do not amount to a serious change of direction. Nation, Meara, Levenston and Richards are clearly an important set of pivotal nodes in the maps, as they provide most of the links between the main clusters. Most of the major players in L2 vocabulary research are already in place in these maps, and we can clearly see a sort of consensus beginning to emerge. Meara (2020b) has argued that these developments can be seen as a "First Paradigm" for L2 Vocabulary Research.

Looking forward, we might ask: what research themes will take the place vacated by the dictionary theme? will the French research continue its decline? will the Dutch language group achieve a breakthrough? will the growing Scandinavian research group and the strong Israeli sub-cluster start to have a major influence on the main vocabulary cluster? We might also note that some sources that rise to prominence in later maps, still play only a small role here. How quickly will they develop their full potential? And will their increasing importance create new research themes that we can observe through these maps?

Finally, we might expect that Nation's 1990 textbook will start to emerge as the key source for vocabulary researchers in the very near future, and we might expect to see a big increase in

the number of research papers which pick up the agenda that this book inspired. The next paper in this series will examine the research published in 1991 in the context of a five-year window covering 1987-1991. For the moment, the likely shifts in L2 vocabulary research remain just around the corner.

References

Albert, M. and L. K. Obler. 1978. *The bilingual brain: neuropsychological and neurolinquistic aspects of bilingualism*. New York.

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. and M. J. McCarthy (eds.) 1988. *Vocabulary and Language Teaching*. London.

Channell, J. 1990. Vocabulary acquisition and the mental lexicon. In J. Tomaszczyk and B. Lewandowska-Tomaszczyk (eds.) *Meaning and lexicography,* 21-31. Amsterdam.

Cohen, A. D. 1987. The use of verbal and imagery mnemonics in second language vocabulary learning. *Studies in Second Language Acquisition* 9: 43-62.

Craik, F. I. M. and R. S. Lockhart. 1972. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior* 11: 671-684.

Craik, F. I. M. and E. Tulving. 1975. Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General* 104: 268-294.

da Solla Price, D. 1965. Networks of scientific papers. *Science* 49(3683): 510-51

Diab, T. 1990. *Pedagogical Lexicography: A case study of Arab nurses as dictionary users.* Berlin.

Dretzke, B. and M. I. Nester. 1990. *Student's guide to false friends, new friends and old friends.* Berlin.

Gairns, R. and S. Redman. 1986. *Working with words: a guide to teaching and learning vocabulary*. Cambridge.

Green, D. W. 1986. Control, activation and resource: a framework for the control of speech in bilinguals. *Brain and Language* 27: 210-223.

Ljung, M. 1990. *A study of TEFL vocabulary.* Stockholm.

Lotka, A. J. 1926. The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences* 16(12): 317-324.

McCarthy, M. 1990. *Vocabulary*. Oxford.

McLaughlin, B. 1987. *Theories of second-language learning.* London.

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 Gonzálvez García and A. Downing (eds.), *The Functional Perspective on Language and Discourse: Applications and implications,* 111-129. Amsterdam.

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

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

Meara, P. M. 2017. A new beginning? A bibliometric analysis of L2 vocabulary research in 1985. *Linguistics Beyond and Within* 3: 136-154.

Meara, P. M. 2018. Laying the foundations. A bibliometric analysis of vocabulary research in 1982-86. *Linguistics Beyond and Within* 4: 108-128.

Meara, P. M. 2019. Steady as she goes? A bibliometric analysis of L2 vocabulary research in 1987. *Linguistics Beyond and Within* 5: 103-119.

Meara, P. M. 2020a. The British are coming...! A bibliometric analysis of L2 vocabulary research in 1989. *Linguistics Beyond And Within* 6: 111-129.

Meara, P. M. 2020b. The emergence of a First Paradigm in vocabulary research: The bibliometrics of *System*. *Vocabulary Learning and Instruction* 9(1): 1-32.

Meara, P. M. 2021. A time for consolidation: A bibliometric analysis of L2 vocabulary research 1985-89. *Linguistics Beyond and Within* 7: 65-85.

Meara, P. M. (no date) VARGA: The Vocabulary Acquisition Research Group Archive. Available at: https://www.lognostics.co.uk/varga/ (accessed 25th October 2022).

Milan, C. and R. Sünkel. 1990. *Falsche Freunde auf der Lauer: Dizionario di false analogie e ambigue affinità fra tedesco e italiano.* Bologna.

Morgan, J. and M. Rinvolucri. 1986. *Vocabulary.* Oxford.

Murphey, T. 1990. *Song and music in language learning.* Bern.

Nation, I. S. P. 1990. Teaching and Learning Vocabulary. Rowley, Mass.

Rudzka, B., Channell, J., Putseys, Y. and P. Ostyn. 1985. *More Words You Need.* London.

Schrameier, A. 1990. *Wortbedeutung im Gedächtnis.* Göttingen.

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.

Stevick, E. 1976. *Memory, meaning, and method.* Rowley, Mass.

Taylor, L. 1990. *Teaching and learning vocabulary.* New York.

Tickoo, M. L. (ed.) 1989. *Learners' dictionaries. State of the Art.* Singapore.

Tréville, M-C. 1990. *Good friends: the use of English cognates for learning French.* Quebec.

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

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.

Willems, M. M. and M. M. B. Oud-de Glas. 1990. *Vocabulaire selectie voor het vreemdetalenonderwijs.* Nijmegen: ITS.

Willis, D. 1990. *The Lexical Syllabus.* London.

Acknowledgements

Appendix 1: Co-citation analysis: The methodology

The co-citation method used in this paper was developed by Small in a number of papers published in the 1970s (e.g. Small: 1973). This approach, which was actually built on earlier bibliometric work by da Solla Price (1965), has been extensively used to analyse research in the natural sciences (e.g. White and Griffith 1981) but does not seem to have been adopted as a standard tool by researchers in the Humanities.

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

of the clusters and the relationship between them. The clusters are generally taken to represent "invisible colleges" in the data.

Appendix 2: Lotka's model

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

$$E_N = T / N^x$$

where   T is the total number of authors who contribute a single paper to the data set,
          N indicates 2,3,4,5... outputs,
and E_N is the expected number of authors contributing to N outputs.

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

contributions	10	9	8	7	6	5	4	3	2	1
Expected E_N	2	3	4	5	7	10	16	28	63	250

Clearly, this model predicts that the number of papers an active researcher might be expected to produce falls off rather quickly. Empirical tests of what has become known as "Lotka's Law" do seem to work well. However, the model works best when we are dealing with well-established fields, and very large data sets. The single year data sets that I have discussed in this series of papers are not a close match to Lotka's expectations, but the larger 5-year data sets are generally a better fit to the power law model. In both cases, however, we get a much better fit when the value of N^x is raised above 2. For example, we get the best fit for the 1986-1990 data set when x= 2.7, 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.