LingBaW. Linguistics Beyond and Within 10 (2024), 89–111

Laying in a new course? A bibliometric analysis of L2 vocabulary research 1988-92

Paul Meara
Swansea University & University of Oxford, UK

Abstract

This paper uses an author co-citation analysis to examine the research on L2 vocabulary acquisition published in 1992. Two analyses are presented. The first analysis provides a context for the 1992 data. It looks at work that was being cited in a five year window covering 1988-92. The second analysis is a more detailed account of the 1992 research on its own terms.

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

1. Introduction

This paper is the twelfth in a series of studies in which I have attempted to map out the way L2 vocabulary research has developed over the last 50 years (e.g. Meara 2023). These reports are based on the research outputs identified in the Vocabulary Acquisition Research Group Archive (VARGA) database (Meara n.d.) The present report takes this historical overview a step further by focussing on the research published in 1992. At first glance, 1992 looks like a good year for L2 vocabulary acquisition research. After a relatively stable period with low levels of output, 1992 shows a significant increase in the number of publications appearing and the number of authors engaging in the field. As we will see, however, these obvious signs of growth are not straightforward to interpret. The report begins with an overview of the research published in the five-year window 1988-92, and continues with a more detailed, exploratory account of the 1992 publications.

The analyses that follow use the Author Co-citation method developed by Small (1973). Small's methodology is described in detail in Appendix A for the benefit of readers who are not yet familiar with the approach used in these reports. In brief, the analyses focus on the authors cited in the bibliographies of a list of papers published in the relevant time frame, and identify sets of authors who are frequently cited together. Usually a small number of very

LingBaW. Linguistics Beyond and Within, e-ISSN: 2450-5188
DOI: 10.31743/lingbaw.18012

strong co-citation clusters emerge from this approach, and we take these clusters to be indications of important research themes in the dataset. The maps developed for this report will follow the same format as last year's report, where I used spanning trees, showing only the strongest links between the sources being cited. Readers of this series of reports will realise that the data sets we are describing are becoming increasingly complex as time goes on, and much more difficult to map in a helpful, intuitive way. The spanning tree approach provides a possible solution to this problem.

2. Part 1: The 1988-1992 data set

We begin by describing the superficial characteristics of the research published in a five year window covering 1988-92. These characteristics are summarised in Table 1, alongside the 1987-91 data for comparison purposes. The table shows that there is very little movement between these two windows. The 1988-92 corpus is very slightly smaller than the 1987-91 corpus and has a slightly larger number of contributors. As usual, most of the contributors identified make just one contribution to the corpus – the 1988-92 figure (81%) shows a small increase over the 1987-91 figure (77%). A small group of authors contribute more than a single item: in 1987-91, 15 authors contributed at least six items to the corpus, but this figure falls slightly in 1988-92, with only 12 contributors authors meeting this criterion.

Table 1: *The basic characteristics of the 1987-91 and the 1988-1992 research outputs*

	1987-91	**1988-92**
Total outputs	636	628
Unique authors	600	633
Prolific authors (6+ contributions)	15	12
Authors making a single contribution	488	512

Table 2 lists the most prolific authors in the 1988-92 data. This prolific author list is slightly smaller than the 1987-91 list. Four authors have dropped out of the list (Galisson, Beheydt, Gass and Hartmann), and one new author (Arnaud) has appeared in the list.

Table 2: *The prolific authors in the 1988-92 research outputs (Prolific here is defined as a contribution to at least six outputs)*

	1987-91	**1988-92**
10+	Meara (22) Laufer (14) Carter (11) Zimmerman (11)	Meara (21) Laufer (18) Vermeer (10)
9	Vermeer	
8	Broeder McCarthy Palmberg	Broeder
7	Galisson Johns	Carter Johns McCarthy Nation Zimmerman
6	Appel Beheydt Gass Hartmann Nation	Appel Arnaud Palmberg

The analysis that follows uses the author co-citation method developed by Small (1973) (see Appendix A). By convention, not all outputs are included in author co-citation analyses. Book chapters and papers published in journals are included, but other types of output (monographs, theses, computer programs, and so on) are not. The rationale for these

exclusions is that book chapters and journal papers tend to have a consistent approach to citation of other people's work, while the other types of output often take a different approach to citing their bibliographical sources. These different practices distort the statistical trends in the data. Theses, for example, usually reference enormous numbers of sources, whereas journal papers are typically more sparing in their approach. The next step in our analysis therefore involves pruning the corpus to generate a smaller data set that consists of journal articles and book chapters. The results of this pruning process are reported in Table 3. The table shows a small increase in the number of eligible outputs in the 1988-92 data set, and a similarly small increase in the number of authors contributing to the data set.

Table 3: *The main characteristics of the 1987-1991 and the 1988-1992 data sets*

	1987-91	**1988-92**
Number of outputs in the data set	455	464
Number of authors contributing to the data set	406	421
Number of sources cited in the data set	4738	5210

The table also shows the number of unique authors that these papers cite. This figure is surprisingly large. As usual, most of the people being cited in the 1988-92 data set are cited in only a single paper (3377 cases or 64% of the total), but a small number of cases are cited much more often and more consistently. Table 4 shows the distribution of these citation patterns.

Table 4: *The number of cases cited N times in the 1988-92 data set*

FREQ	**75+**	**74**	**73**	**72**	**71**	**70**	**69**	**68**	**67**	**66**	**65**	**64**	**63**	**62**	**61**
Cases	1	1													
FREQ	**60**	**59**	**58**	**57**	**56**	**55**	**54**	**53**	**52**	**51**	**50**	**49**	**48**	**47**	**46**
Cases							1					1	1		1
FREQ	**45**	**44**	**43**	**42**	**41**	**40**	**39**	**38**	**37**	**36**	**35**	**34**	**33**	**32**	**31**
Cases		1	1					2	1			4	2	3	2
FREQ	**30**	**29**	**28**	**27**	**26**	**25**	**24**	**23**	**22**	**21**	**20**	**19**	**18**	**17**	**16**
Cases	1		3	4	6	5	3	6	4	5	9	5	10	11	9
FREQ	**15**	**14**	**13**	**12**	**11**	**10**	**9**	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**
Cases	11	18	23	21	21	30	31	40	67	80	100	208	333	757	3377

Table 5 lists the most frequently cited authors in this data set. This table shows that the most cited authors list is very stable across 1987-91 and 1988-92. The most cited author in 1988-92 is Nation, cited in 76 outputs (16% of the total). Two authors who figured in the 1987-91 list (Faerch and Levenston) do not appear in the 1988-92 list. Only one new author appears in the 1998-92 list (Lockhart).

Table 5: *The authors cited in at least 40 of the items in the 1988-92 data set.*

1987-91	1988-92
Meara (77)	Nation (76)
Nation (69)	Meara (74)
Carter Richards (49)	Carter (54)
Levenston (47)	Krashen (49)
Faerch (45)	Laufer(48)
Krashen (44)	Richards (46)
Laufer (41)	Sinclair (44)
Sinclair (40)	Lockhart (43)

The analysis that follows is based on the co-citations among the most frequently cited authors in the 1988-92 data set. Clearly, it is not feasible to analyse in detail the connections between all 5210 authors in the data, and in order to keep things simple, it is normal practice in author co-citation studies to work with the 100 or so most frequently cited authors. The data in Table 4 suggest that we can get close to this conventional figure if we adopt an inclusion threshold of 16 citations in the data set. This threshold gives us a list 103 authors. In our analysis of the 1987-91 data set we adopted an inclusion threshold of 15 citations, and 98 authors met this threshold. The 1988-92 data set is therefore very comparable to the 1987-91 data set in size, but slightly more demanding in its threshold (114 authors in the 1988-92 data set are cited 15 times or more). These characteristics are summarised in Table 6.

Table 6: *The main characteristics of the 1987-91 and the 1988-92 data sets.*

	1987-91	1988-92
Inclusion Threshold	15	16
Authors included	98	103
New Authors		18
Lost Authors	13	

Table 6 also shows that there is a relatively small amount of churn between the two data sets. Thirteen authors who appear in the 1987-91 data set no longer meet our inclusion threshold (Atkins, Beheydt, R Brown, Carton, Fries, Jain, Lado, Lorge, Phillipson, Selinker, Stein, Tarone and Tomaszczyk), while 18 new authors appear in the 1988-92 data set (Chen, Coltheart, Durgunoglu, Johns, Kolers, Macnamara, Magiste, McClelland, Meyer, Morton, Potter, Rey, Roediger, Ruddy Schwanenflugel, Sim, So and von Eckardt). Some of these authors are returners from earlier data sets (Kolers, Macnamara, Coltheart) but others are genuinely new, and will be discussed in more detail in Section 3 of this report.

The co-citations among the 103 most highly cited authors in the 1988-92 data set were mapped using the Gephi software package. (Bastian, Heymann and Jacomy 2009). Co-citations which appear only once in the data set are ignored: Gephi identifies 2114 co-citation links that appear in the data set at least two times. In our earlier reports we simplified the co-citation data by setting an arbitrary strength threshold for inclusion. This made the resulting maps easier to read and interpret, but it makes comparisons between data sets more difficult to handle. In the analysis that follows, we asked Gephi to generate a spanning tree map, based

on the strongest co-citations within the data set. This approach is more consistent than the arbitrary threshold approach that we used in our earlier reports. The methodology for building spanning trees is explained in more detail in Appendix C.

Figure 1 shows the basic mapping of the 1988-92 data set. This map shows the 103 authors who are cited at least 16 times in the data set. Each author appears as a node in the map; the size of a node indicates how many other nodes it is connected to.

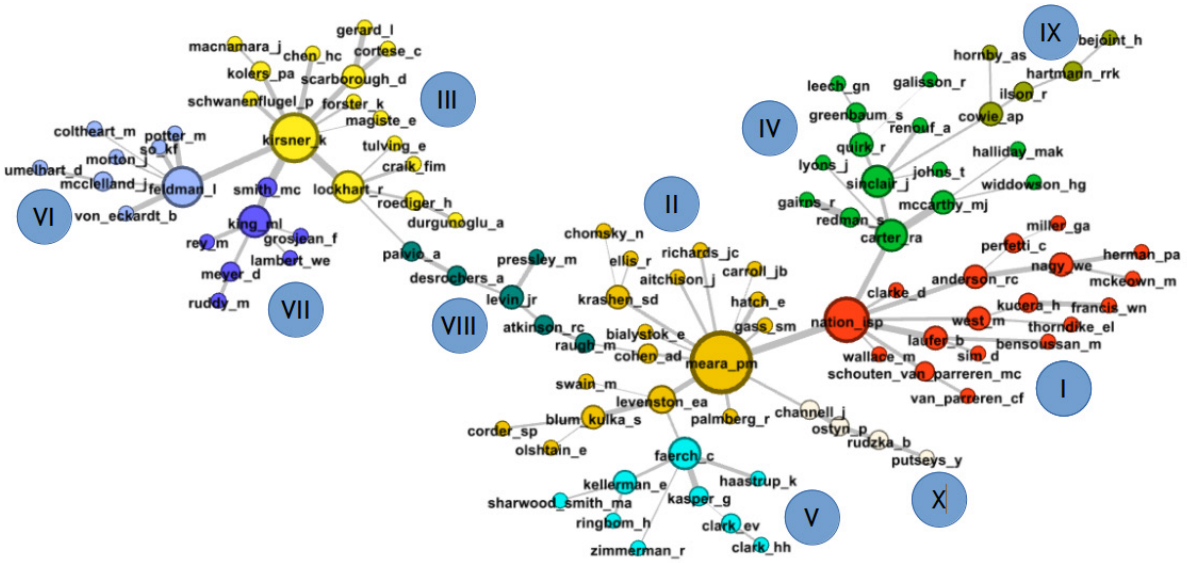

Figure 1: *A spanning tree map of the 1988-92 data set. 103 nodes with at least 16 citations in the data set. Colours indicate the ten thematic clusters identified by Gephi. Nodes are sized according to how many connections they have with other nodes.*

Based on the strongest connections between these authors, Gephi finds 10 clusters in this map:

Cluster I, focussed on Nation, with 18 members, is the largest cluster in the map. There are a number of sub-themes here. West, Kucera & Francis, and Thorndike are a set of word frequency counts; Nagy, Anderson, Herman and McKeown are a group of L1 reading specialists whose work is highly cited by the other members of this cluster reflecting an interest in L2 reading processes. Both these subclusters have appeared in our earlier maps.

Cluster II, focussed on Meara, with 17 members, seems to be comprised of authors whose work is mainly concerned with Second Language Acquisition in general, rather than L2 vocabulary acquisition in particular. This work has strongly influenced the L2 vocabulary research. Particularly noticeable here is the subcluster of authors based in Israel (Levenston, Blum-Kulka, AD Cohen and Olshtain.

Cluster III, focussed on Kirsner (15 members), is a group of psycholinguists whose work mainly deals with formal models of word recognition in bilinguals.

Cluster IV, focussed on Sinclair and Carter (12 members), contains a set of linguists whose main interest lies in descriptions of English and corpus linguistics. This cluster is particularly associated with the Universities of Birmingham and Nottingham.

Cluster V, focussed on Faerch, is a group of (predominantly Scandinavian) applied linguists whose work is mainly concerned with transfer between a bilingual's L1 and other

languages that they are learning. This cluster also contains two members who are best known for their work on L1 acquisition (EV Clark and HH Clark).

Cluster VI, focussed on Feldman, is a second group of psycholinguists interested in bilinguals. This cluster is less focussed on word recognition than are the members if cluster III. Coltheart, for example is a dyslexia specialist, and Morton was publishing papers that developed a model of L1 word recognition.

Cluster VII, is really focussed on Kirsner, and should be seen as an extension of Cluster III. This cluster is generally interested in the linguistic behaviour of bilinguals.

Cluster VIII is a group of six psychologists who are interested in imagery and mnemonics and their applications to L2 vocabulary acquisition.

Cluster IX is a small group of applied linguists with interests in dictionaries and the way L2 speakers use them.

Cluster X identifies the four authors of a series of textbooks that develop a componential analysis approach to L2 vocabulary teaching.

Table 7 summarises the main features of this map, and provides comparison figures for the equivalent map covering the 5 year window 1987-91.

Table 7: *The main clusters in the 1988-92 data set*

cluster	1987-91	1988-92
I	Vocabulary teaching and reading (21)	Vocabulary learning theory (18)
II	Lexical error and transfer (16)	Vocabulary teaching and reading (17)
III	Vocabulary learning theory (16)	Bilingual word recognition (15)
IV	Performance of bilinguals (14)	Corpora and Discourse (12)
V	Corpora and Discourse (13)	Lexical error and transfer (9)
VI	Dictionaries and their use (7)	Psycholinguistics (9)
VII	Imagery and Mnemonics (6)	Performance of bilinguals (7)
VIII	Semantics and Collocation (4)	Imagery and Mnemonics (6)
IX		Dictionaries and their use (5)
X		Applications of Semantics (4)

Broadly speaking, the two maps are very similar: the clusters in the 1987-91 map are easily recognizable in the 1988-92 map, but there are some subtle shifts in the structure of the field which suggest that the field has not yet solidified. The 1988-92 map contains more clusters than the 1987-91 map, and the new clusters are on the whole smaller than the earlier ones. Bilingual word recognition (Cluster III in 1988-92) seems to be a new research theme. Lexical error and transfer seems to be declining in importance. There is also some movement in the membership of the clusters.

The 1988-92 map seems to fall naturally into three sectors. Clusters III, VI and VII make up a set of psycholinguistic sources. Clusters IV and IX make up a set of formal linguistic sources. Clusters I, II, V and X form the main L2 vocabulary acquisition sources. Cluster VIII, the mnemonics and imagery cluster provides an interesting set of links between the psycholinguistics clusters and the rest of the network.

Despite the familiarity of this overall structure, a number of less obvious changes can be found in the data set. These mainly concern the key co-citation links in the 1988-92 data set. Table 8 lists the strongest links in this data set, and the strongest links in the 1987-91 data set for comparison.

Table 8: *The strongest co-citation links in the 1987-91 and the 1988-92 data sets*

1987-1991	1988-92
Faerch ~ Kasper 34	Kirsner ~ Smith 32
Carter ~ McCarthy 32	Carter ~ McCarthy 32
Gairns ~ Redman 32	Gairns ~ Redman 32
Nation ~ Meara 30	Kirsner ~ Lockhart 30
Levenston ~ Blum-Kulka 28	King ~ Smith 29
Levenston ~ Meara 27	Meara ~ Nation 29
Faerch ~ Haastrup 24	Laufer ~ Nation 27
Channell ~ Ostyn 23	Faerch ~ Kasper 26

The main point to note here is the disappearance of four very strong co-citation links from the 1987-91 list (Levenston ~ Blum-Kulka, Levenston ~ Meara, Faerch ~ Haastrup and Channell ~ Ostyn) and the unexpected entry of Kirsner, Lockhart, King and Smith into the 1988-92 list. These last names are co-authors of a 1984 paper that set a methodological agenda for a series of studies using lexical decision tasks. It is not clear why this paper has emerged from obscurity at this time. Kirsner emerges as a major hub in the 1988-92 data set, but to some extent, this may be an artefact of the way we are treating multi-authored papers. Each time Kirsner is cited in the data set, he is also co-cited with his co-authors, and this gives him a prominence that perhaps needs to be interpreted with caution. We will discuss this problem further in Section 3 of this report. A more straightforward feature worth noting is the continued dominance of the L2 vocabulary research by Nation and Meara, and the first appearance of Laufer in the strongest co-citations list. Carter and McCarthy still appear as significant foci in the linguistics cluster, but overall this cluster appears to be less influential than it was in the 1987-91 map.

3. Part2: The 1992 data in more detail

We now turn to a more detailed analysis of the research published in 1992.

Figure 2 shows the distribution of output types in this year, compared with the outputs identified for 1991. The figure shows that there is a very large increase in the number of outputs in 1992. This increase mostly comes in the form of chapters in books rather than papers published in journals. The number of books and monographs dealing with L2 vocabulary actually fell in 1992, compared with 1991. Table 9 lists the four outputs that fall into this category.

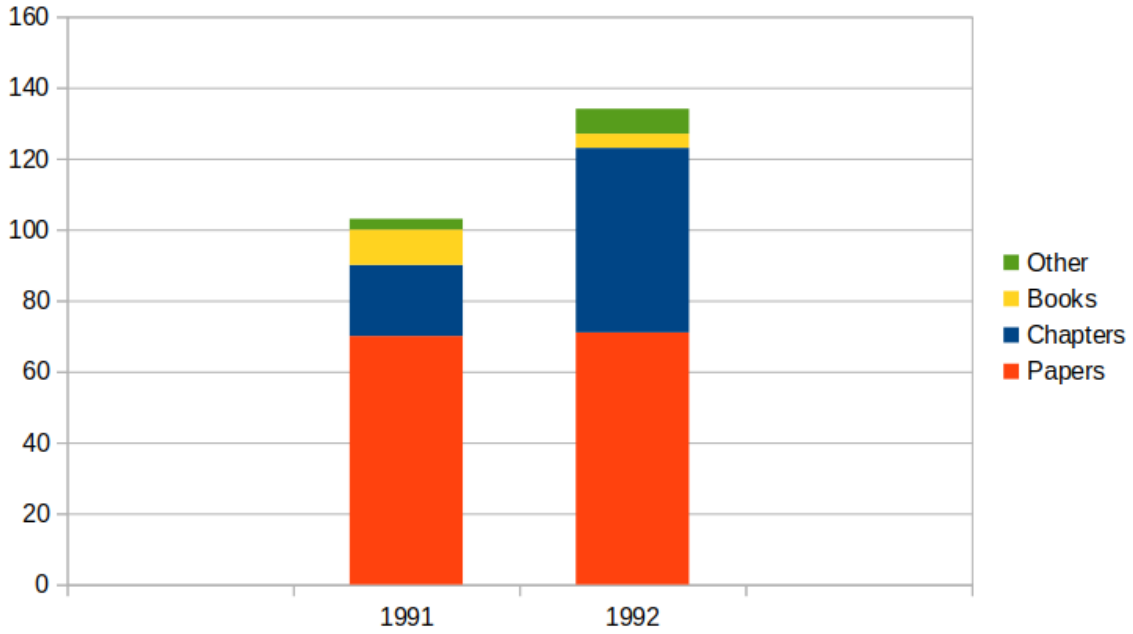

Figure 2: *The 1992 research output by type*

Not included in Table 9 is a massive volume edited by **R. J. Harris** *Cognitive processing in bilinguals* (Harris 1992). This collection contains 33 papers, about half of which deal directly with the way bilingual speakers process words. These papers are included separately in the data set for 1992, and the impact of Harris' book is considered further in the discussion section.

Table 9: *Monographs and edited volumes published in 1992*

Arnaud, P. & H. Béjoint (eds.) *Vocabulary and Applied Linguistics*. London: Macmillan. 1992.
Meara, P. M. Vocabulary in a second language. Vol. 3. *Reading in a Foreign Language*, 9(1992) whole volume.
Picoche, J. *Précis de lexicologie française. L'étude et l'enseignement du vocabulaire.* Paris: Nathan. 1992.
Sánchez Lobato, J. & B Aguirre Beltrán *Léxico fundamental del español: Situaciones, temas y nociones. Glosario multilingüe.* Madrid: SGEL. 1992.

Notes

Arnaud & Bejoint is an edited collection of 17 papers delivered at a conference in Lyon in 1991. Most of these papers are included as separate entries in the 1992 dataset. These entries mainly deal with dictionaries for L2 speakers and some psychological aspects of L2 vocabulary acquisition.

Meara is an annotated bibliography covering material on vocabulary acquisition published between 1960 and 1990. The volume contains some 350 entries, plus a brief glossary, and a short introduction.

Picoche is a textbook that introduces a distinctively French approach to lexicology. Most of this book is concerned with developing Sausurre's idea of the linguistic sign, but part of chapter 2 deals with the pedagogy of vocabulary. This chapter contains some notes on active and passive vocabulary, an account of the *Trésor de la langue française*, a good account of the *Français fondamental* research, and a discussion of the relationship between word frequency and *mots disponibles* – words that readily come to mind in the context of every day tasks. This section ends with a brief account of lexical statistics.

Sánchez Lobato & Aguirre Beltrán is a list of about 2250 Spanish words that seem to have been elicited on the basis of a *Français Fondamental* type investigation. The words are arranged into 15 topics and a set of exercises is provided for each topic. Translations of the 2235 words are provided in English, French and German.

Table 10 is a list of doctoral theses published in 1992. The VARGA database does not systematically monitor theses, so this list may underestimate the amount of work of this type that became available in 1992. The theses listed here are important enough to have been cited in subsequent years, and they hint at research centres and authors which may be influential in the future.

Table 10: *Doctoral theses awarded in 1992*

Crutcher, R. J. *The effect of practice on retrieval of foreign vocabulary learned using the key-word method.* PhD University of Colorado, Boulder. 1992.

Griffin, GF *Aspects of the psychology of second language vocabulary list learning*. PhD thesis, Warwick University. 1992.

Jiménez Catalán, R. M. *Errores en la producción escrita del inglés y posibles factores condicionantes.* Madrid: Universidad Complutense de Madrid. 1992.

Lee, L. *The effect of instructional method and learning style on Spanish vocabulary learning in college students.* Doctoral dissertation. University of Texas, Austin. 1992.

Sanaoui, R. *Vocabulary learning and teaching in French as a second language classrooms*. PhD Thesis. University of Toronto. 1992.

Siramard, Y. *Combining extensive reading and intensive vocabulary study in a Japanese university.* Doctoral dissertation, Temple University Tokyo. 1992.

The 1992 outputs also include a number of unpublished reports and working papers that are not included in the data set analysed later. These include:

Goodfellow, R. CALL and lexical RECALL. Open University CITE Report. No 164. 1992.
This report is a general discussion of the role computers might play in the teaching of vocabulary. These points are illustrated with a discussion of STORYBOARD. Goodfellow briefly reports two informal studies of EFL learners doing processing tasks with decontextualised vocabulary.

Hall, C. J. Making the right connections: vocabulary learning and the mental lexicon. Puebla, Mexico. ERIC Document Reproduction Service ED 363 128. 1992. The status of this report is unclear. It seems to anticipate some later publications dealing with Hall's Parasitic Model of vocabulary acquisition,

Krohn, D. *Grundwortschätze und Auswahlkriterien. Metalexikographische und fremdsprachendidaktische Studien zur Struktur und Funktion deutscher Grundwortschätze.* Göteborg: Acta Universitas Gothoburgensis. 1992.
I was unable to obtain a copy of this work.

Meara, P. M. *EFL Vocabulary Tests*. Swansea University. (ERIC Document Reproduction Service No. ED 362 046)**.** 1992. A set of experimental vocabulary size tests.

Nagy, W., G. Garcia & B. Hancin-Bhatt Cross-language transfer of lexical knowledge: bilingual students' use of cognates. *Technical report No.558. 1992.* I was unable to obtain a copy of this internal report.

Schmidt, K-H. & P. Metzler *Wortschatztest (WST).* Weinheim: Beltz Test. 1992. I think this is a computer program.

A total of 185 authors can be identified in this data set, nearly double the number of authors identified in the 1991 data set. Table 11 reports the number of authors contributing to multiple outputs.

Table 11: *The number of authors contributing to N outputs in 1992*

N outputs	8	7	6	5	4	3	2	1
Cases in 1992				1	2	1	18	163
Cases in 1991					1	2	5	91
Lotka's estimate	3	3	5	7	10	18	41	

As usual, most of the increase in in the number of outputs appearing in 1992 comes from authors who contribute to just a single output, but we can also note a large jump in the number of authors contributing to two outputs. Once again, the most prolific author is Laufer (5 outputs). Arnaud and Meara contribute to 4 outputs each, and de Groot contributes to 3 outputs. 18 authors contribute to two outputs: Béjoint, Doctor, Fernández, Grainger, Harrington, Hartmann, Heredia, Hulstijn, Klein, Leffa, Löschmann, McLaughlin, Oller, Pearson, Thomas, Umbel, Vermeer and Wang.

Béjoint is best known for his work on dictionary use. **Grainger** is a psychologist based in Paris who works on formal models of bilingual lexicons. Together these two authors identify important strands in French vocabulary research. **Harrington's** work is focussed on working memory in bilinguals. In 1992 he was based at the University of California in Santa Cruz. **Hartmann**'s work is concerned with L2 dictionary use. **Leffa**'s two papers deal with electronic glosses. A new feature in this data set is the presence of authorial teams in the prolific author list. **Heredia and McLaughlin** make up a team based in California and working on bilingual memory processes. **Doctor and Klein** were a South African team, also working on formal models of bilingual lexicons. **Umbel, Pearson, Oller and Fernández**, based at the University of Miami, are mainly interested in the development of vocabulary in Spanish speaking children in Florida. **Wang and Thomas**'s work deals with mnemonic imagery strategies This group is also based in Florida.

Table 11 also reports an estimate of how many prolific authors we would expect to find in a data set where 163 authors contribute to only a single output. (See Appendix B for details of this analysis.) We have noted in previous reports that the L2 vocabulary research is heavily dominated by one-off studies, and surprisingly deficient in authors who make more than one contribution to the annual data sets. This trend seems to continue into 1992.

It is worth noting here that almost all of the prolific authors in 1992 are new: only Laufer and Meara also appeared as prolific authors in 1991. Table 12 shows the extent of this churn.

Table 12: *Prolific authors in 1991 and 1992 (here "prolific" means more than one contribution to the data set)*

1991	1992
Laufer Meara Bogaards Scholfield Gruneberg Kelly Mondria Stevens	**Laufer Meara** Arnaud Béjoint de Groot Grainger Harrington Hartmann Hulstijn Leffa Löschmann Vermeer Doctor & Klein, McLaughlin & Heredia, Thomas & Wang, Pearson, Umbel, Oller & Fernández

3.1. The data sources

The VARGA database (Meara n.d.) identified 119 outputs published in 1992 that were eligible for inclusion in the analysis that follows. A small number (10) of these outputs were not traceable, and these items are listed in Table 13.

Table 13: *Items published in 1992 that I was unable to obtain copies of.*

Abe, H. & T. Matsui. An analysis of verbs of utterance. In Y. Shimizu (ed.) *Lexical development of Japanese ESL students*. Dokkyo University: Association for English Language Teaching. 1992.

Akagawa, Y. Can pre-reading activities over-ride Japanese students' poor knowledge of vocabulary? *JACET Bulletin* 23(1992), 1-20.

Grönholm, M. Lexikal strategier hos svenskspråkiga elever vid inlarning av finska. [Lexical strategies of Swedish speaking students learning Finnish.] In M. Axelsson & A. Viberg (eds.) *Forsta forskarsymposiet om Nordens sprak som andrasprak.* Stockholm: 1992.

Heredia, R., M. S. Weldon & B. McLaughlin Conceptually driven vs. data driven processes in bilingual memory: one or two systems. *PALM*, 7(1992), 255-278.

Lahuerta Martínez, C. Adquisición del vocabulario: Aproximación al estudio de la función de las claves lingüísticas en el proceso de interpretación del vocabulario durante la lectura de textos ingleses. In F. Etxeberria & J. Arzamendi (eds.) *Bilinguismo y Adquisición de Lenguas.* Bilbao: Servicio Editorial Universidad del Pais Vasco. 1992, 353-363.

Löschmann, M . Wortschatzarbeit: kommunikativ-integrativ, interkulturell, kognitiv. In U. Jung (ed.) *Praktische Handreichungen für Fremdsprachenlehrer.* Frankfurt/M.: Verlag Lang. 1992, 311-319.

Löschmann, M. Arbeit am Wortschatz. In U. Jung (ed.) *Praktische Handreichung für Fremdsprachenlehrer.* Frankfurt/Main: Verlag Lang. 1992, 311-319.

Patris, J.& N. Vasnick Comment aborder les fautes lexicales en classe? *Enjeux* 26(1992), 46-55.

Porquier, R. Construction de la référence spatiale dans l'intéraction exolingue. [Constructing spacial reference in cross-language interaction.] In R. Bouchard, J. Billiez, J-M. Colletta, V. de Nucheze & A. Millet (eds.) *Acquisition et enseignement/apprentissage des langues.* Grenoble: LIDILEM. 1992.

Most of these items are chapters in books not held in UK libraries. The remaining 109 outputs, (65 journal articles and 44 book chapters) make up the dataset that is analysed in the report that follows. For space reasons, I have not listed all the items in this paper. However, interested readers can identify these included items by accessing the VARGA database at https://www.lognostics.co.uk/varga and entering the search terms **1992 {JA}** and **1992 {CH}.**

3.2. The analysis

Next we look in detail at who is being cited in this data set, and more specifically at the co-citation patterns that emerge from the analysis.

This analysis identifies 2033 unique authors cited in the papers that make up the 1992 data set. The corresponding figure for 1991 was 1486, so we have a substantial increase in the number of authors being cited. As usual, the vast majority of these authors are cited only once, but we have a number of authors who are more frequently cited in the data set. The full distribution is shown in Table 14.

Table 14: *The number of authors cited N times in the 1992 data set*

N	**20**	**19**	**18**	**17**	**16**	**15**	**14**	**13**	**12**	**11**
authors	1		1	2	3	1	2	1	6	6
N	**10**	**9**	**8**	**7**	**6**	**5**	**4**	**3**	**2**	**1**
authors	3	11	8	11	20	31	47	112	256	1511

The most frequently cited authors in this data set are Kirsner (cited in 20 papers), Kolers (cited in 18 papers), Lockhart and MC Smith (cited in 17 papers), Feldman, King and Nation (cited in 16 papers each), von Eckhardt (cited in 15 papers) and Potter and HC Chen (cited in 14 papers) each. Only one of these authors (Nation) appeared in the 1991 list of most cited authors: citations of Nation increased from 13 in 1991 to 16 in 1992. Nearly all of the highly cited sources in 1991 received fewer citations in 1992, despite the increase in the number of papers in the data set. The two exceptions are Laufer (cited 11 times in 1991 and in 1992), and Krashen (cited 8 times in 1991 and 12 times in 1992). It is difficult to see this as anything other than significant shift away from L2 vocabulary research (see Table 15).

Table 15: *The most cited authors in 1991 and their citations in the 1992 data set*

	Meara	**Carter**	**Laufer**	**Richards**	**Atkinson**	**Gairns**	**Redman**	**Aitchison**	**Krashen**	**McCarthy**	**Pressley**
1991	15	12	11	11	10	9	9	8	8	8	8
1992	12	8	11	6	6	2	2	5	12	4	6

For the next step in our analysis, we eliminate the authors who are only infrequently cited, and work with a reduced set of highly cited sources. Conventionally, we work with around 100 highly cited sources, but readers may remember that our analysis of the 1991 data involved a rather smaller figure than this. In 1991, we identified 78 authors who were cited at least five times in the data set, and, for the purposes of comparison, it would be useful for us to identify in the 1992 data set a group of influential authors of around this size. Table 14 indicates that there are 76 authors who are cited at least six times in the 1992 data set, and the analysis that follows will be based on this subset of the full data set. Table 16 shows the main characteristics of this group of highly cited authors, compared with the equivalent group that we identified in the 1991 data set. The higher inclusion threshold for the 1992 data set reflects the fact that the 1992 data sets includes a larger number of outputs than the 1991 data set.

Table 16: *The general features of the 1991 and 1992 data sets*

	1991	**1992**
Authors included	78	76
Inclusion Threshold	5	6
New Authors		48
Lost Authors	50	

Figure 3 shows a mapping based on the co-citation links between the 78 most frequently cited authors in the 1992 data set. The figure shows co-citation links between the authors who appear at least four time in the data set. Weaker links have been eliminated in the interests of simplicity. In spite of this simplification, the map clearly illustrates the problems that arise

when we have a very high level of co-citation between the sources. It is almost impossible to tease out what is going on in the two larger clusters in this figure.

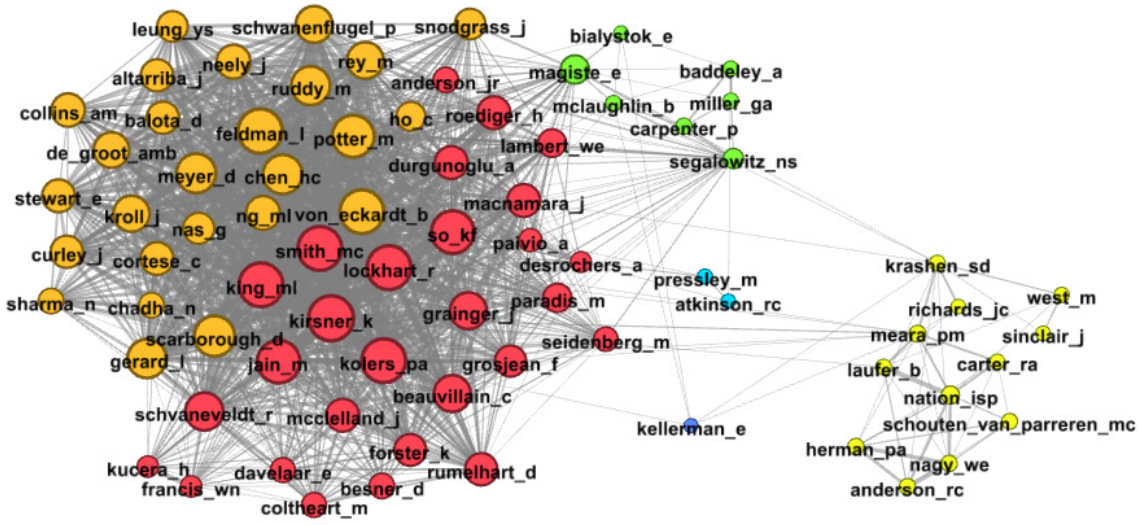

Figure 3: *A map of the co-citations between the 78 most cited authors in the 1992 data set*

This map looks very different from any of the maps that have appeared in this series of reports so far, but in spite of this, it is relatively straightforward to interpret. We have two clusters of authors, shown in red and orange, mainly populated by psycholinguists, and very tightly interlinked by their co-citations. These two clusters account for 56 (72%) of the nodes in the data set. We have a third cluster of psycholinguists, shown in green, who are slightly detached from the two larger clusters, and we have a cluster of 12 applied linguists (shown in yellow). This last cluster has very few co-citation links with the rest of the network: Only Krashen, Meara and Laufer are directly linked to other clusters, and these links are very weak. We also have a cluster consisting of two authors who specialise in the applications of mnemonics (Pressley, RC Atkinson) and a cluster with only one member (Kellerman), who provides some unexpected links between the other clusters.

The main feature to note here is the way that the psycholinguistic research has very suddenly come to dominate the field as a whole, and to a large extent eclipsed the research of applied linguists. In our earlier maps, the psycholinguistics research appeared to be fairly marginal, and most of the cutting edge research could be aligned with research that was more oriented towards linguistic descriptions. Here, the situation appears to be reversed: what we have previously referred to as the "mainstream L2 vocabulary research" has suddenly become marginalised.

The second feature to note here is the huge number of sources appearing in 1992 who did not figure in the 1991 data set. This churn is summarised in Table 17. Fifty authors who were significant in 1991 have disappeared from the 1992 significant author list, while 48 new authors have emerged. Only 28 authors – about a third of the total – have a presence in both the 1991 list and the 1992 list. The changes are listed in Table 17, and they will be discussed in more detail later.

Table 17: *Changes in the composition of the 1991 and 1992 data sets*

Lost authors from 1991 (50 sources)
AITCHISON_J BECK_I BENSOUSSAN_M BERRY_J CARROLL_JB CHANNELL_J COHEN_AD CRAIK_FIM CURTIS_M DELANEY_H FAERCH_C GAIRNS_R GALISSON_R GOODMAN_KS GREENBAUM_S HAASTRUP_K HALL_JW HATCH_E JOHNS_T KELLY_P LEECH_GN LEVENSTON_EA LEVIN_JR LYONS_J MARSLEN_WILSON_W MCCARTHY_MJ MCCORMICK_C MCDANIEL_M MCKEOWN_M MILLER_GE OBLER_LK OLLER_J OLSHTAIN_E PALMBERG_R PEARSON_P PERFETTI_C QUIRK_R RAUGH_M REDMAN_S RINGBOM_H SHARWOOD_SMITH_MA SHIFFRIN_R STERNBERG_R SVARTVIK_J SWAN_M TULVING_E VAN_PARREREN_CF WATERS_G WILSON_KP ZIMMERMAN_R
Authors appearing in 1991 and 1992 (28 sources)
ANDERSON_RC ATKINSON_RC CARTER_RA DESROCHERS_A FORSTER_K FRANCIS_WN HERMAN_PA KELLERMAN_E KING_ML KIRSNER_K KRASHEN_SD KUCERA_H LAMBERT_WE LAUFER_B LOCKHART_R MAGISTE_E MEARA_PM MILLER_GA NAGY_WE NATION_ISP PAIVIO_A PRESSLEY_M RICHARDS_JC RUMELHART_D SCHOUTEN_VAN_PARREREN_MC SINCLAIR_J SMITH_MC WEST_M
New Authors appearing in 1992 (48 sources)
ALTARRIBA_J ANDERSON_JR BADDELEY_A BALOTA_D BEAUVILLAIN_C BESNER_D BIALYSTOK_E CARPENTER_P CHADHA_N CHEN_HC COLLINS_AM COLTHEART_M CORTESE_C CURLEY_J DAVELAAR_E DE_GROOT_AMB DURGUNOGLU_A FELDMAN_L GERARD_L GRAINGER_J GROSJEAN_F HO_C JAIN_M KOLERS_PA KROLL_J LEUNG_YS MACNAMARA_J MCCLELLAND_J MCLAUGHLIN_B MEYER_D NAS_G NEELY_J NG_ML PARADIS_M POTTER_M REY_M ROEDIGER_H RUDDY_M SCARBOROUGH_D SCHVANEVELDT_R SCHWANENFLUGEL_P SEGALOWITZ_NS SEIDENBERG_M SHARMA_N SNODGRASS_J SO_KF STEWART_E VON_ECKARDT_B

The density of the co-citation links within the two largest clusters in Figure 3 makes it difficult to see the fine detail of the co-citation patterns in this part of the map. However, Figure 4 provides a simpler mapping in which most of the weaker links in the map have been pruned so that we are left with a spanning tree where each node is directly linked only to its strongest co-citation partner. This visualisation is directly comparable with the 1991 spanning tree that appeared in our last report.

The first point to note here is the emergence of three new hubs in the 1992 data set: Kirsner, Kolers and Chen. Kirsner had a relatively small role in the 1991 maps, but is clearly dominant in the 1992 map, a rise to dominance that can only be described as dramatic. Kolers has not been a significant influence for some time. He was an important figure in some of our earlier maps, but most of his work was published in the 1960s and 70s, and he does not appear to have published any relevant papers after 1980. Again, his importance in the 1992 map comes as a surprise. Chen appears to be a complete newcomer to the list of significant influences. He published a series of experimental studies in 1989 (Chen & Ng 1989 and Chen & Leung 1989). Chen has not figured in our previous maps, so his appearance as a major influence in 1992 is surprising too. The big losers in 1992 are Nation, Meara and Carter. These three authors were identified as significant influences in the 1991 map, but they fail to achieve this status in 1992. The biggest loser in 1992 is Meara, a very significant influence in 1991, but here reduced to a relatively minor role. Meara's co-citation link with Kirsner is one of the weakest links in the entire network (there are only two co-citations linking these two sources). This link plays an important role in preventing the network from falling into two separate networks, however.

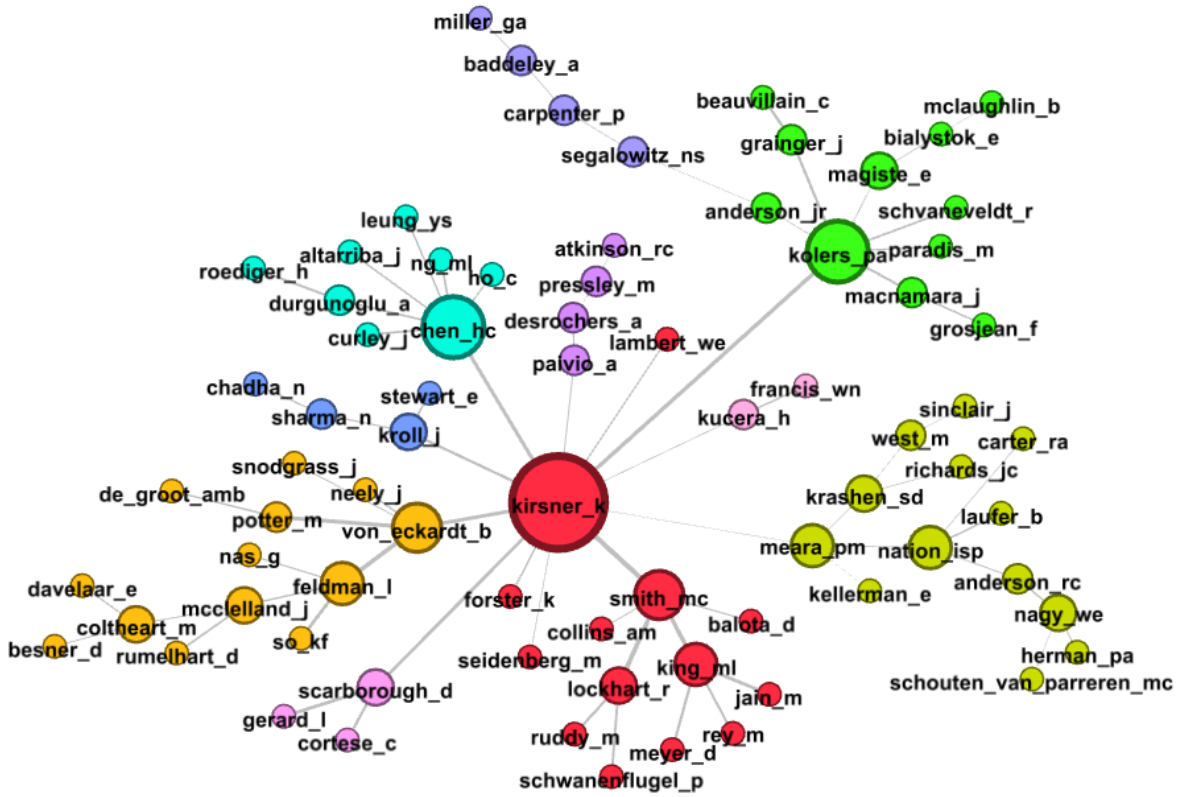

Figure 4: *A spanning tree analysis of the 1992 data set*

Table 18 lists the strongest co-citation links in the 1992 data set. As usual, the equivalent data from 1991 is included for the purposes of comparison. The strongest links in 1992 are much stronger than the equivalent links in 1991 and none of the strongest links in 1991 figure in the 1992 list. Tellingly, some of the authors who appear in the 1991 strong links list do not have any presence in the 1992 map (Gairns & Redman, McCarthy, Aitchison, Raugh, & Atkinson, Levin). This looks like a serious change in direction for the field as a whole.

Table 18: *The strongest co-citation links in the 1991 and 1992 data sets*

Link Weight	1991	1992
17		Kirsner~Smith
16		Lockhart~Smith
15		King~Smith Feldman~Smith
14		Potter~von Eckardt Kirsner~Kolers
13		Kirsner~von Eckardt
		Gerard~Scarborough Kirsner~Scarborough
		Jain~King Feldman~So
9	Gairns~Redman	
7	Carter~Meara Carter~McCarthy	
	Carter~Aitchison	
6	Nation~Schouten-van Parreren	
	Laufer~Nation Raugh~Atkinson	
	Nation~Meara Levin~Pressley	
	Kucera~Francis	

Only three of the strongest links that we identified in 1991 appear in the 1992 map: Kucera and Francis are co-cited slightly more often in 1992 than they were in 1991; Laufer and Nation, likewise are cited slightly more often in the new data set; Meara and Nation are co-cited slightly less often in the new data set (only four times).

The data in displayed in Figure 4 were submitted to a formal analysis using the Gephi program. The analysis identified 10 clusters in this data set:

Cluster I (14 members and dominated by Kirsner) is a group of psycholinguists whose main interest is formal models of lexical storage in bilinguals. The key source here is a 1984 paper co-authored by Kirsner, Smith, Lockhart, King and Jain.

Cluster II (13 members focussed on von Eckardt) is a second group of psycholinguists interested in formal models of lexical storage in bilinguals. The key source here is another 1984 paper co-authored by Potter, So, von Eckardt and Feldman.

Cluster III (13 members focussed on Meara and Nation) is a much reduced set of authors whose main concern is L2 vocabulary learning, with a particular emphasis on L2 reading.

Cluster IV (11 members focussed on Kolers) is another group of psychologists working on bilingual performance. This cluster is less focussed on formal models than Cluster I and Cluster II are.

Cluster V (8 members focussed on Chen) largely consists of people who have co-authored papers with Chen. Chen is particularly interested in Chinese, while Durgunoglu and Roediger work with French speakers.

Cluster VI (4 members focussed on Kroll) seems to be distinguished from the other psycholinguistic groups by a methodological interest in translation.

Cluster VII (4 members) is a set of authors who deal with imagery and mnemonics.

Cluster VIII (4 members) is a set of authors interested in memory processes and skilled reading. The key author here is Norman Segalowitz, an active researcher in the 1970s and 1980s, a very significant figure in Canada, but only rarely cited in the applied linguistics research in Europe at this time.

Cluster IX (3 members, focussed on Scarborough). The members of this cluster are particularly interested in cognate effects in laboratory studies of word recognition.

Cluster X with 2 members is a standard word frequency count.

These clusters are listed in Table 19. This table also identifies the main clusters that emerge in the 1991 spanning tree map. The new list of clusters is heavily dominated by authors who work on formal models of L2 storage. *Imagery and Mnemonics*, an important cluster in 1991, has been reduced to a small cluster with four members. The L1 reading skills cluster, the *Corpora and Discourse* cluster and the *Descriptive Approaches to English* cluster have all collapsed and become absorbed into an all-purpose L2 vocabulary cluster, where they are represented by Carter and Sinclair. We will explore this collapse further in the discussion that follows.

Table 19: *The clusters identified in the 1991 and 1992 data sets*

Cluster	1991	1992
I	Vocabulary Acquisition and Transfer (14)	Formal models of lexical storage (14)
II	Mnemonics and Imagery (13)	More formal models of lexical storage (13)
III	L1 reading skills (11)	Vocabulary uptake and inferencing (13)
IV	Performance of Bilinguals (10)	Performance of bilinguals (11)
V	Corpora and Discourse (10)	Chinese/French (4)
VI	Vocabulary uptake and inferencing (9)	Translation effects in bilinguals (4)
VII	Descriptive approaches to English (5)	Imagery and Mnemonics (4)
VIII	(Johns, Oller)	Memory processes and skilled reading (4)
IX		Cognate effects (3)
X		Word Frequency Count (2)

4. Discussion

The simplest account of the changes from 1991 to 1992 suggests that the traditional concerns of L2 vocabulary research have been replaced by an upsurge in psycholinguistics research and formal modelling approaches. Unlike the traditional approaches, this new work is laboratory-based, rather than classroom based. It uses highly technical research methods which would be difficult to emulate in classroom situations. It is also striking that most of this research is concerned with bilingual speakers, rather than the (low level) second language learners,who usually serve as participants in L2 vocabulary research. It is difficult to see how the insights gained in the psycholinguistic research could be applied in real world situations. Some of the research, for example, suggests that bilingual speakers are very slightly slower at recognising words in their weaker language, but the difference is tiny – often not much more than a few tens of milliseconds – and not really likely to be registered outside of laboratory studies. This does indeed feel like a serious change of course – a new set of priorities, a new set of methods, and a serious shift in perspective.

At the same time, the mapping shown in Figure 4 strongly suggests that research in foreign language vocabulary learning has (at least temporarily)been eclipsed by the research in psycholinguistics. A number of research areas that were important in 1991 have disappeared from the 1992 map, and a very large number of authors who were significant influences in 1991 have stopped being cited. For example, it looks as though research on dictionary use and corpus linguistics is only barely registering in 1992. The only identifiable themes that appear in both 1991 and 1992 are mnemonics, lexical inferencing and the importance of reading for L2 vocabulary acquisition.

However, this interpretation is not as straightforward as it looks. One possibility is that the 1992 maps have been seriously skewed by the appearance of Harris' edited collection of papers in 1992. This volume is tightly edited, and it adopts citation practices which are notably different from anything that we have seen so far in this series of studies. The chapters in Harris' collection frequently cite each other, for example, a feature which is not usually found in papers that are published in a single year in more disparate sources. The papers also have much more extensive bibliographies than is normal for journal papers at this time. In

1992, the average number of outputs cited in a paper was 31 but nearly all the papers in the Harris collection exceed this figure. (In 1990, the equivalent figure was only 24.)

Many of the same comments apply to the collection of papers that appeared as a special issue of the *European Journal of Cognitive Psychology* edited by de Groot and Berry (de Groot & Berry 1992). This volume contains seven papers which cover much of the same ground as the Harris volume. Taken together, these two volumes make up nearly 20% of the outputs that constitute the 1992 data set, and it is not surprising that they dominate the co-citation map. The obvious solution to this problem is to eliminate the papers in these volumes from the 1992 data set and see what we are left with. Surprisingly, this radical pruning does not make much difference to the overall structure of the maps, and this suggests that the shift away from applied linguistics driven research towards more psycholinguistics concerns is even more serious than it might appear at fist sight. (See Figure 5.)

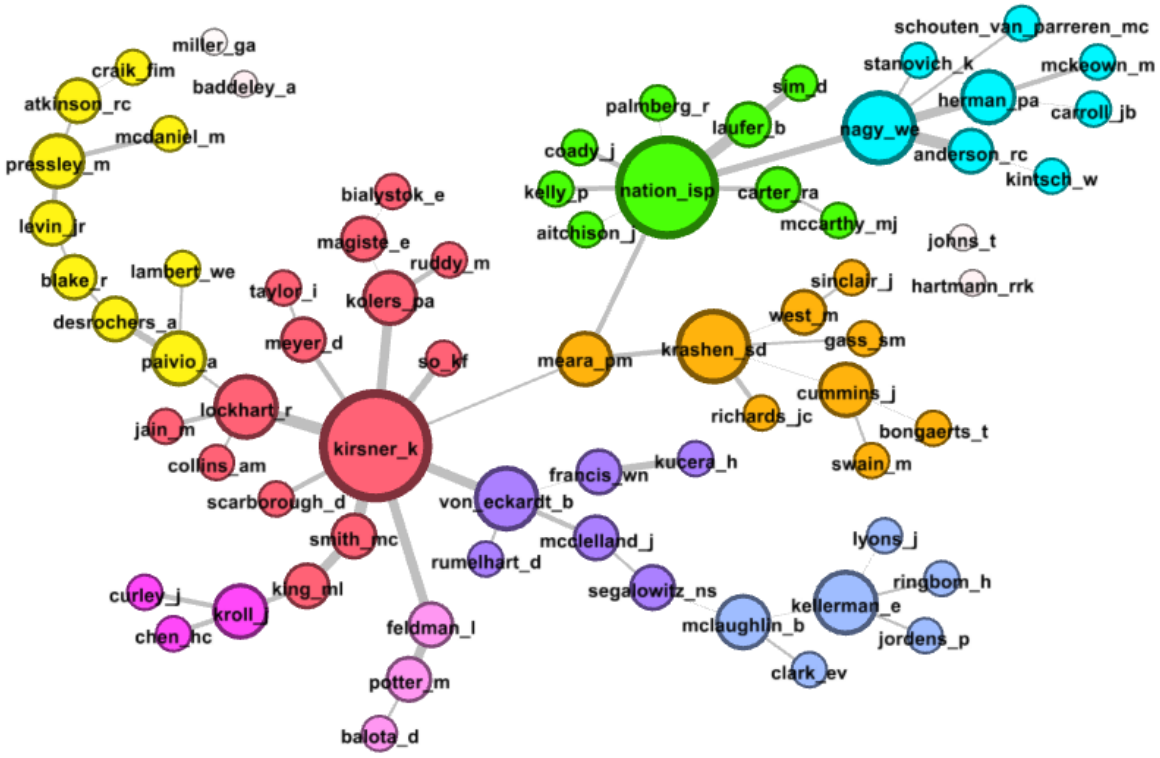

Figure 5: *A spanning tree map of the co-citations in 1992 data set (Citations within the papers in Harris 1992 and de Groot & Barry 1992 have been eliminated)*

Kirsner still emerges as the most significant influence in the reduced 1992 spanning tree. Meara still plays a key role in linking the psycholinguistic work to the L2 vocabulary acquisition work. The biggest difference between these two maps is the emergence of Nation as the second very significant influence – the first time that he has appeared in this role in these reports. It is also worth pointing out that a number of linguists who were absent from Figure 4 do make an appearance in Figure 5, and many of these influences are strongly co-cited with Nation (Aitchison, Kelly, Coady, Palmberg, Sim and McCarthy). All these figures appeared in the 1991 spanning tree. The L1 reading cluster has developed into a distinct cluster dominated by Nagy. Kellerman emerges as the nucleus of a small cluster dealing with

semantics and meaning, and there has been a notable expansion in the mnemonics cluster which now includes Lambert, Blake, Levin, McDaniel and Craik. These shifts suggest that the field as a whole is fairly stable, but it is not yet large enough or resilient enough to be unaffected by the appearance of a large collectionsof highly focussed papers with a shared methodological focus.

The strongest co-citation links in Figure 5 are listed in Table 20. These weights are very comparable to the weights that we reported in the 1991 data set, and this suggests that the very strong weights reported in Table 18 are indeed an aberration. Nevertheless, there are some warning signals that need to be heeded in Figure 5. The majority of the strong co-citation links involve Kirsner and colleagues, and only the strong link between Nation and Laufer runs against this trend. (See Table 20).

Table 20: *The strongest co-citation links in Figure 5*

Link Weight	
7	Laufer~Nation Kirsner~Lockhart Kirsner~Smith Nagy~Anderson
6	Kirsner~von Ekardt Kirsner~Feldman Kirsner~Kolers Feldman~Potter King~Smith Nagy~Herman

Generally speaking, the linguistics clusters in Figure 5 are held together only by the most tenuous of links. It is difficult to tell whether this represents a genuine structural change in the field. However, it IS noticeable that a number of authors who were important in the 1991 map no longer play a role in the 1992 maps. Gairns & Redman, who made up the strongest co-citation link in 1991 do not appear in the 1992 map. They represent a strand of research that was strongly involved with practical applications of more theoretical research, and their loss is a significant shift in emphasis. Also absent from the 1992 maps are Galisson, Zimmerman and Faerch & Haastrup, representatives of the French, German and Scandinavian research traditions that we have identified in our earlier maps. Levenston, Bensoussan and Olshtain have been succeeded by Laufer as the most cited author in Israeli vocabulary research. Perhaps the most striking loss is the disappearance of Leech, Greenbaum, Svartvik & Quirk, who we identified in our earlier maps as an important cluster dealing with linguistic descriptions of English. This strand of research is here represented only by Sinclair, Carter & McCarthy. Dictionary research is represented only by Hartmann, who appears as an unattached source. Johns, too (an advocate of hands-on concordancing), appears as an unattached source in this analysis.

To sum up, it seems that many of the concerns that figure in our 1991 maps have been replaced by new research foci. This does not mean that the main themes that we identified in 1991 have stopped altogether. Rather, they do not reach the critical mass that allows them to appear in a map of the most significant themes in the 1992 map.

Conclusion

Clearly, we have identified some very significant shifts in the 1992 dataset, and there are hints that some very large changes are appearing over the horizon. After a period of reduced output,

1992 shows a modest increase in the number of outputs published in a single year, and it provides the largest set of outputs in the 1988-92 window. This increase in activity hints that we may need to be on the lookout for new emerging research themes in future years. Perhaps the most important feature in this report is the discrepancy between the five year 1988-92 analysis and the more detailed analysis of the 1992 data. The data that underlies the five year map appears to be very stable, with only a few minor changes compared to the 1987-91 window. In contrast, the smaller 1992 data set appears to be wildly different from the 1991 data set, with very large structural changes taking place. This suggests that the field might be entering a period of rapid development and change.

We can anticipate that the field will continue to grow at a rapid rate, both in terms of the number of research outputs recorded, and in terms of the number of authors contributing to these outputs. It is more difficult to foresee the direction that these changes will move the field as a whole. Nevertheless, two developments in particular do hint at important changes to come.

Firstly, it is worth noting that 1992 saw a small meeting of psycholinguists and L2 vocabulary researchers which resulted in a collection of papers published in 1993. (Schreuder & Weltens 1993). This meeting brought together a number of significant figures from both the psycholinguistics and the applied linguistics approaches, and the resulting publication looks to be a text that might successfully bridge the growing divide between these two diverging research traditions.

Secondly, in my earlier reports, I noted the importance of Nation's book *Teaching and Learning Vocabulary* (Nation 1990). Most of the work appearing in 1992 will have been written before the publication of this book – the research cycle was much slower in the days before the internet than it is now, and it is not surprising the find that *TLV* does not seem to be influencing the co-citation maps to date. By 1993 and 1994 we can expect the impact of this book to appear in the more recent maps. In the meantime, however, the steady rise in the importance of Nation in these maps seems to indicate that the applied linguistics strand of L2 vocabulary might be able to recover from the hits it received in 1992.

All in all, 1992 looks as though it was a pivotal year for vocabulary research, and we can confidently look forward to interesting new developments in 1993.

Acknowledgements

Special thanks to John Read for some very insightful comments on an early draft of this paper.

References

Bastian, M., S. Heymann., and M. Jacomy. 2009. Gephi: An open source software for exploring and manipulating networks. *International AAAI Conference on Weblogs and Social Media*.

Chen, H-C., and Y. S. Leung. 1989. Patterns of lexical processing in a non-native language. *Journal of Experimental Psychology: Learning, Memory and Cognition* 15: 316–325.

Chen, H-C., and M. L. Ng. 1989. Semantic facilitation and translation priming effects in Chinese-English bilinguals. *Memory and Cognition* 17: 454–462.

de Groot, A. M. B., and C. Barry (eds). 1992. *European Journal of Cognitive Psychology* 4(4).

Harris, R. J. (ed.) 1992. *Cognitive processing in bilinguals.* Amsterdam: North Holland.

Kirsner, K., M. C. Smith, R. S. Lockhart, H. L King, and M. Jain. 1984. The bilingual lexicon: language-specific units in an integrated network. *Journal of Verbal Learning and Verbal Behaviour* 23: 519–529.

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

Meara, P. M. 2023. Shifting sands: A bibliometric analysis of L2 vocabulary research in 1991. *Linguistics Beyond and Within* 9: 112–132.

Meara, P. M. (no date) VARGA: The Vocabulary Acquisition Research Group Archive. https://www.lognostics.co.uk/varga/ (Accessed 1 October 2024).

Nation I. S. P. 1990. *Teaching and Learning Vocabulary*. Rowley, Mass.: Newbury House.

Prim, R. C. 1957. Shortest connection networks and some generalisations. *Bell System Technical Journal* 36(6): 1389–1401.

Potter, M. C., K-F. So, B. von Eckardt, and L. B. Feldman. 1984. Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning And Verbal Behavior* 23: 23–38.

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

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.

Appendix A

Author co-citation analysis: The methodology

The author 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 Price (1965), has been extensively used to analyse research in the natural sciences (e.g. White & 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 mapping 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 B

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 one author who makes two contributions, and it would likewise be very unusual to find that a single author who 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.

***Table A**: An illustration of Lotka's Law with x = 2 and $N_1 = 250$*

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 1988-1992 data set when x = 2.9. This is lower than the equivalent figure for 1987-91, which was x = 3.01, though both figures need to be treated with some caution because the data sets are 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, but the fall in the 1988-92 figure seems to suggest that the field is becoming slightly more "normal" than it was in 1987-91.

Appendix C

Spanning trees

The maps presented in this paper are a simplification of the maps that appeared in the earlier papers in this series. The earlier maps tried to capture the relationships between the authors by including any co-citation link which was stronger than a chosen threshold value – for example, we might include any link with a weighting of 8 occurrences or more in the data set. The threshold values were chosen to avoid cluttering up the visuals with very weak connections, but they varied from one report to another, and were essentially arbitrary.

In this paper, I have adopted an alternative solution to this problem, by displaying the data in the form of a spanning tree. In this alternative approach, we start with a list of authors, a list of all the co-citation links between them, ordered by their weight, and an empty map containing no nodes. We then build a map by working through the ordered list of links, and following the steps outlined in an algorithm developed by Prim (Prim: 1957). Starting with the strongest link, we add nodes and edges to the empty map as long as the new edge does not lead to a cycle. That is, if we have a new edge A ~ B, and our tree does not already contain a link (direct or indirect) between node A and node B, then we add the edge A ~ B to the map, adding new nodes as necessary. The map grows in a piecemeal way at first, adding pairs of strongly connected nodes to the map, but eventually, the algorithm finds a set of links that connects each node to another by its strongest connection.

Cite this article as:

Meara, P. (2024). Laying in a new course? A bibliometric analysis of L2 vocabulary research 1988-92. *LingBaW. Linguistics Beyond and Within, 10*, 89–111.