Online edition : ISSN 2187-2759 Print edition : ISSN 2187-2767

Vocabulary Learning and Instruction

VLI

A Journal of Vocabulary Research

Volume 3 Number 1 December, 2014

Papers by Paul M. Meara
Raymond Stubbe
Masaya Kaneko
James Rogers, et al.

Commentary by John P. Racine

Vocabulary Learning and Instruction

Volume 3, Number 1,

December 2014

doi: http://dx.doi.org/10.7820/vli.v03.1.2187-2759

VLI Editorial Team

Editor: Raymond Stubbe

Editorial Board: Charles J. Anderson, Peter Carter, Luke Fryer, Aaron Gibson, Jeffrey Stewart

Reviewers: Thomas Boutorwick, Dale Brown, Peter Carter, Averil Coxhead, Frank Daulton, Yuko Hoshino, Tomoko Iishi, Kiwamu Kasahara, Masumi Kojima, Stuart McLean, Paul Meara, Kaori Nakao, Quint Oga-Baldwin, John P. Racine, John Read, Vivienne Rogers, Tim Stoeckel, Stuart Webb

Copy Editors: Alex Cameron, Mark Howarth, Tim Pritchard, Andrew Thompson, Alonzo Williams

The Editorial Team expresses a sincere thank you to *Mana Ikawa*, who designed the cover for *VLI*.

Copyright © 2014 *Vocabulary Learning and Instruction*, ISSN: Online 2187-2759; Print 2187-2767. All articles are copyrighted by their respective authors.

Vocabulary Learning and Instruction

Volume 3, Number 1, December 2014
doi: http://dx.doi.org/10.7820/vli.v03.1.2187-2759

Table of Contents

Article	Page
Letter from the Editor *Raymond Stubbe*	iv
Vocabulary Research in the *Modern Language Journal*: A Bibliometric Analysis Online first March 15, 2014 *Paul M. Meara*	1
Do Japanese Students Overestimate or Underestimate Their Knowledge of English Loanwords More than Non-loanwords on Yes–No Vocabulary Tests? Online first March 15, 2014 *Raymond Stubbe*	29
Is the Vocabulary Level of the Reading Section of the TOEFL Internet-Based Test Beyond the Lexical Level of Japanese Senior High School Students? Online first March 15, 2014 *Masaya Kaneko*	44
A Methodology for Identification of the Formulaic Language Most Representative of High-frequency Collocations *James Rogers, Chris Brizzard, Frank Daulton, Cosmin Florescu, Ian MacLean, Kayo Mimura, John O'Donoghue, Masaya Okamoto, Gordon Reid, & Yoshiaki Shimada*	51
Commentary	
Reaction Time Methodologies and Lexical Access in Applied Linguistics Online first April 1, 2013 *John P. Racine*	66

Letter from the Editor

Dear Readers,

Vocabulary Learning and Instruction (*VLI*) has been pleased to bring you three special issues to date, each featuring papers presented at the annual JALT Vocabulary SIG Vocabulary Symposiums.

It gives me great pleasure to present to you *VLI*'s inaugural regular issue. All the papers appearing below were submitted to the journal and peer-reviewed by two or three of the reviewers mentioned on the previous page. Many thanks to all of our reviewers and copy-editors.

In the pages to follow you will find articles by Paul M. Meara, Masaya Kaneko and James Rogers, et al., plus a commentary by John P. Racine. Reflecting the diversity of research occurring within the field of vocabulary, topics range from a bibliometric analysis of vocabulary research in the *Modern Language Journal* to a methodology for identification of the formulaic language most representative of high-frequency collocations.

As a reminder, *VLI* is an open-access international journal that provides a peer-reviewed forum for original research related to vocabulary acquisition, instruction and assessment. Submissions are encouraged from researchers and practitioners in both first language and EFL and ESL contexts.

We hope you will enjoy the papers that follow and encourage everyone involved in vocabulary research to submit a paper soon.

Hoping you a prosperous 2015,

Raymond Stubbe,

Editor, VLI

Vocabulary Learning and Instruction
Volume 3, Issue 1, December 2014
http://vli-journal.org

Article

Vocabulary Research in the *Modern Language Journal*: A Bibliometric Analysis

Paul M. Meara
Swansea University and Cardiff University
doi: http://dx.doi.org/10.7820/vli.v03.1.meara

Abstract

This paper reports a bibliometric analysis of a set of 201 articles which was published in *The Modern Language Journal* (MLJ) between 1916 and 2010. All these articles deal with vocabulary acquisition. The paper reports an all-inclusive author co-citation analysis of this data, in an attempt to sketch out the historical development of vocabulary acquisition research. The paper presents a set of maps which shows whose work is being cited in the Journal. Co-citation links between cited sources allow us to identify research clusters which are characterised by patterns of citations. This paper uses these maps to show how the predominant research focus has changed significantly over the period studied. Much of the earlier work published in MLJ no longer figures in more recent research. The more recent research appears to be much more inward-looking and self-referential than is the case for the earlier research. This paper suggests that a co-citation analysis of research in a single journal does not capture the full richness of vocabulary research, which in turn raises some interesting questions about the selectivity of journals and their research biases.

1 Introduction

The massive increase in research outputs in the last 20 or 30 years has given rise to a growing interest in bibliometric techniques that can provide coherent accounts of the main trends in a large body of work that might otherwise be intractable. One of the best established methods of doing this appears to be the co-citation model developed by Small (1973) in a number of papers published in the 1970s. This work, 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 (Hellqvist, 2010). My own interest lies in the area of vocabulary acquisition by second-language learners, and I have used the co-citation method to successfully generate analyses of small parts of the research literature in this area. Specifically, I have analysed the entire research output for 1982 and 2006. This work, reported in Meara (2012, forthcoming), is interesting in that it provides snapshots of the research that was taking place in these years, and allows us to identify research foci within the wider research community (Price, 1986; Zuccala, 2006). These foci are clusters of researchers – people who tend to be cited together in the same papers and represent a thematic feature within a research landscape.

Please cite this article as: Meara, P.M. (2014). Vocabulary research in The *Modern Language Journal*: A bibliometric analysis. *Vocabulary Learning and Instruction*, *3* (1), 1–28. doi: 10.7820/vli.v03.1.meara

Doing an exhaustive analysis of this sort is not as simple as it sounds. The main problem is that we can never be sure that we have in fact identified all the relevant research published in a single year. For L2 vocabulary research, the best source is the Vocabulary Acquisition Research Group Archive (VARGA) database (Meara, n.d.). This database is a very large bibliographic source that contains some 5000 papers dealing with L2 vocabulary acquisition, and is probably the closest thing we have to a comprehensive listing.[1] However, even a well-maintained resource of this sort is, in the last analysis, only a personal bibliography, and we cannot be sure that it has recorded every piece of research that is to be had. VARGA is good on English-language research, Dutch-language research and Spanish-language research, but its coverage of research in other languages, particularly German and Chinese sources, is less reliable. Furthermore, VARGA's attempts to provide comprehensive coverage means that it contains a lot of derivative literature which many researchers would not consider to be cutting-edge work. This introduces some interesting biases into the citation patterns: for example, it tends to give more prominence to secondary sources, and downplays the importance of seminal work that is well known to only a few key researchers, but not widely known to people working on more routine research topics. More importantly, when we analyse the work published in a single year, the resulting snapshots fail to capture the dynamics of the research activity – how research in vocabulary acquisition changes its focus over time, how some authors make long-term contributions to the field, how other authors flourish only briefly before their influence fades away, how the field persistently fails to follow up important ideas and so on. In order to investigate these ideas, we need to make a longitudinal study of vocabulary research rather than a cross-sectional one.

Doing a complete longitudinal analysis of all the work included in a database like VARGA would be a massive undertaking, but we can make a first stab at a historical account of L2 vocabulary research by looking at all the relevant research published in a single long-running journal. Using as source material papers published in a single journal provides us with a tractable set of data which is comprehensive in a limited kind of way. It also solves the problem of quality, in that articles which appear in a high quality scholarly journal come with some sort of guarantee that the editors, at least, think these papers deal with important topics that are relevant to their readers' interests. Hopefully, a complete analysis of this sort might serve as a framework on which other, more specific analyses can be positioned.

This paper, then, presents a bibliometric analysis of the vocabulary research published in a single journal, and the question we are asking is whether data of this sort can provide a basic historical framework that can be used as scaffolding for further studies on how researchers have approached L2 vocabulary acquisition over an extended period. The journal selected for analysis is *The Modern Language Journal* (MLJ). MLJ is a good candidate for a study of this sort. It first appeared in 1916 and has an unbroken history of publication since that date. MLJ rapidly established itself as the flagship journal of the Modern Language Association (MLA) and it attracted submissions from people who in their day would have been considered major figures in the field. It is currently rated as 16 out of 121 Linguistics Journals according to the ISI Journal Citation Reports – a very significant achievement when we consider that the Linguistics list includes a

number of very heavyweight publications. MLJ is not particularly partisan from a theoretical point of view, but it does seem to be closely in touch with contemporary developments in research and responds quickly to new developments, such as the growth of Computer Assisted Language Learning (CALL), or the emergence of a new geopolitically significant language. More importantly, from our point of view, is that the journal has consistently shown an interest in vocabulary research over its long history. In recent years, it has published a number of papers which have gone on to become very influential in vocabulary research, and a surprisingly large number of the Significant Influences in vocabulary research that were identified in my 2012 paper have appeared as authors in its pages.

2 Methodology

Although it seems a straightforward matter to identify vocabulary-related research in a single journal, in practice this process is more subjective than it appears. Research practice between 1916 and 2010 is far from consistent. Few of the early papers contain any empirical data, for example, and many of these studies are little more than short opinion pieces, critical of other scholars' work, but not always providing good evidence for their authors' views. There are also a few pieces which deal with vocabulary acquisition alongside other questions – reading, vocabulary and grammar, for instance – and with these papers it was necessary to make a judgement call as to whether their vocabulary content was sufficient for them to be included or not. I eventually identified a total of 201 papers whose main focus appeared to be vocabulary, published in MLJ between 1916 and 2010. This works out to two or three papers per year – though it is worth noting that the distribution of papers is far from uniform, and there is a particularly striking fall-off in vocabulary research between 1940 and 1980 (see Figure 1). For reasons which

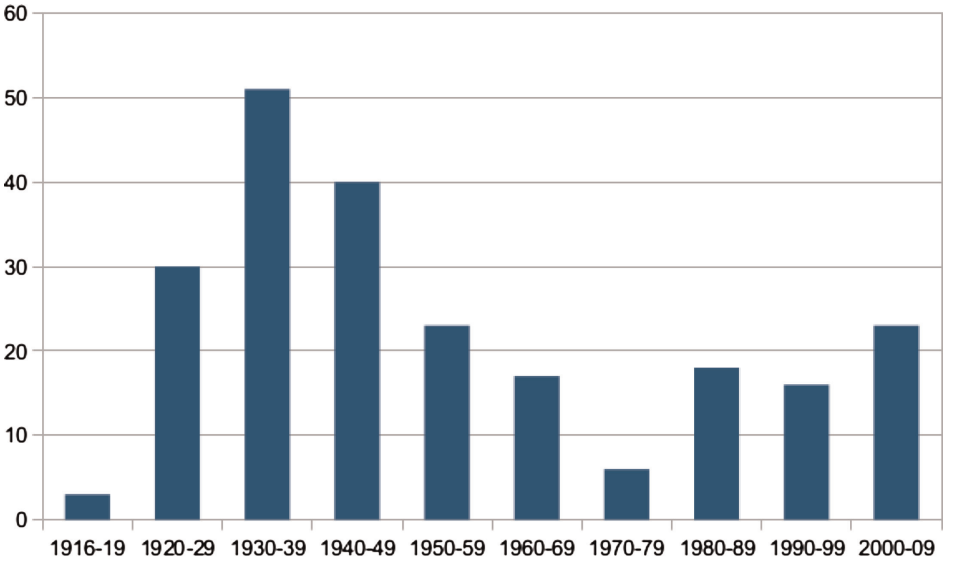

Figure 1. Papers whose main focus is vocabulary: The *Modern Language Journal* 1916–2010.

will become obvious as this paper progresses, the entire list of 201 sources is not included in the bibliography at the end of this paper but a complete listing is included as an Appendix to this paper.

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. Normally, this process is a purely mechanical one, which consists of extracting a machine-readable author list from a set of references at the end of the article. In the case of MLJ, this straightforward data collection turned out to be surprisingly difficult. Modern citation practices were not adopted by the journal until the 1970s, and earlier volumes often do not have proper bibliographies. A surprising number of early papers have no references at all, and where they do the references are sometimes cursory and incomplete. For instance, the text may refer to "a recent comment by a well-known Professor", without giving any bibliographical details. This meant that the only way to construct the complete author list was to read the papers and extract the data manually.

Once this work was complete, I developed a computer programme to generate the necessary co-citation lists. Numerate readers will realise that the amount of data generated in this way is very large, as the number of co-citations is an exponential function of the number of authors cited in a paper. An author list of only five people generates a set of 10 co-cited pairs ($5*4/2 = 10$), a list of 20 authors generates $20*19/2 = 190$ co-cited pairs, while a list of 100 authors generates $100*99/2 = 4950$ co-citations. Generally speaking, the early papers in MLJ cite only a few references, but the more recent articles have very extensive reference sets. The average number of citations over the entire set of 201 papers is 19, generating a set of 113,071 co-citations. Fortunately, there are available a number of network mapping programmes which allow data sets of this size to be processed relatively easily. In this paper, I have used a programme called Gephi (Bastian, Heymann, & Jacomy, 2009). Gephi performs a cluster analysis on the data and groups together authors who tend to be cited alongside each other in a number of papers. The clusters are taken to represent groups of researchers who share similar concerns.

3 Results

Gephi's output for the entire data set is shown in Figure 2.[2]

In this figure, the nodes represent individual authors, and the lines linking the nodes show the co-citations. The size of the nodes reflects the number of other sources each author is associated with, and the width of the links shows the frequency of the co-citations. The colour of the nodes indicates the clusters identified by Gephi. The raw data have been greatly simplified by the exclusion of authors who are cited in fewer than three texts or appear in fewer than three co-citations. However, even this simplified data is extremely complex: the graph contains 4833 co-citations linking 296 authors. It is immediately obvious that some of the authors cluster together into large clumps, while others are more peripheral or belong to detached clusters. In fact, Gephi has identified a total of six distinct

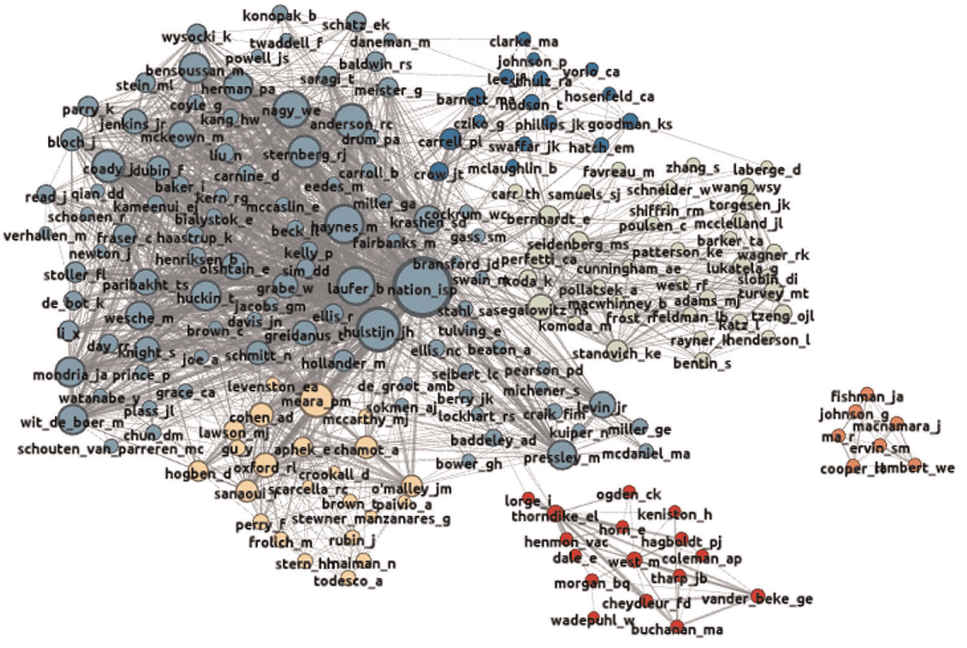

Figure 2. Co-citation analysis of vocabulary-related articles 1916–2010. Threshold for inclusion: three co-citations.

research clusters in this data. The complex nature of the data presented in Figure 2 makes it difficult to interpret these clusters, but we can simplify the task by extracting the data from a number of thin time-slices, before returning to the full picture at a later point. This is done in the following sections.

3.1 1916–1950

My original intention was to analyse the MLJ data in a set of 10-year windows, but this strategy was foiled by the very low level of citations in the early papers. In fact, the best we can do with the early papers is to group together all the work published between 1916 and 1950. This gives us a set of 102 papers – just over half of the data set. However, 25 of these papers contain no citations, and the average number of citations for the remaining 77 papers is a mere 5.8. Figure 3 shows Gephi's analysis of this data.

This map is made up of 134 co-citation links identifying 41 sources in three clusters. The clusters in Figure 3 can be easily described.

Cluster I, the central cluster focussed on West and Thorndike, is made up of 34 sources that have an interest in the development of word frequency counts for L2 learners. West is a very significant figure in English-language teaching and responsible for the influential General Service List of English Words published in 1953. Thorndike will be familiar to most readers for his work on the Thorndike and Lorge frequency count published in 1944, a text which was a standard reference tool in psycholinguistics well into the 1970s. However, Thorndike was not

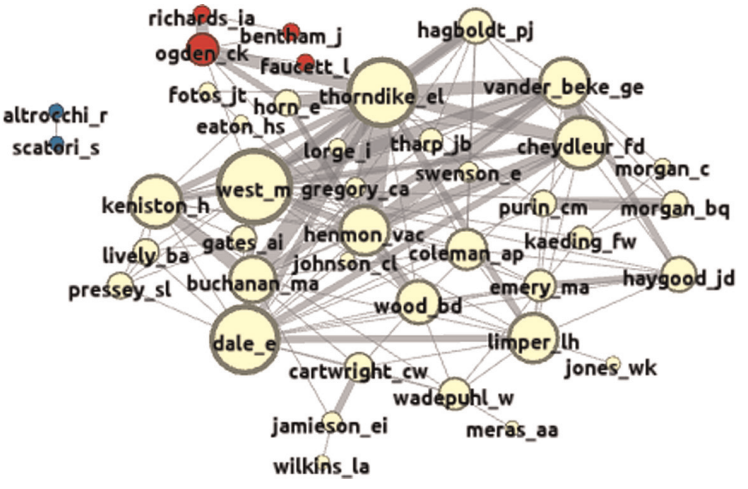

Figure 3. Co-citation analysis of vocabulary focussed articles 1916–1950. Threshold for inclusion: two co-citations. The Significant Influences 1916–1950: West, Thorndike, Dale, Keniston, Cheydleur, Henmon, Limper, Vander Beke, Buchanan, Wood, Coleman, Haygood, Hagboldt, Emery and Ogden.

principally a linguist: his main contribution to research at this time was in the psychology of learning and the measurement of cognitive skills, and a number of other sources cluster with Thorndike because of their work as educational psychologists interested in assessing vocabulary knowledge. Henmon published a frequency count for French, but like Thorndike he was basically a psychologist – he had worked with James McKean Cattell, one of the pioneers of psycholinguistic research. Horn was also an educational psychologist, particularly interested in children's vocabulary. He published a number of early word counts in the 1920s. Dale is mainly concerned with evaluating the difficulty of reading texts, and he worked with Irvin Lorge on a readability index for English. Dale's work on English is often cited in the context of L2 vocabulary acquisition because he is mainly interested in the implications of frequency-based word lists for teaching high-level reading courses, and this has some obvious implications for advanced L2 speakers. The other people in this cluster will probably be less familiar to readers. Keniston and Buchanan developed early word counts for Spanish, Vander Beke and Hagboldt worked on German, Cheydleur worked on French and was also influential in the development of language testing. The striking feature of this cluster is how closely the sources are interconnected. The cluster is basically a "theoretical core" of people working on the principles of word counts, and an extended periphery which tries to implement these principles in word counts for specific languages.

Cluster II, with four sources, is focussed on Ogden and Richards, the originators of BASIC English. BASIC addresses the same problems that form the focus of Cluster I, but solves these problems by developing a very small core vocabulary for English, which is largely independent of the available frequency counts.

Cluster III is a small detached cluster which comprises people who are interested in cognates and the role they might play in vocabulary learning.

A number of features are worth pointing out in these data. The first notable feature of this map is that almost all the work produced during this very long period falls into a single cluster that is predominantly concerned with word counts and their applications. This cluster consists of a densely connected core and a periphery of people who are less densely connected to this core or with each other. The periphery seems to consist of people who are actively involved in teaching languages, while the core of the cluster seems to consist of external reference points – people who work on vocabulary, but are not part of the ongoing debate about L2 vocabulary needs which characterises the periphery. However, the pattern of connections suggests that these external reference points are strongly linked to the rest of the cluster, and I interpret this to mean that the L2 researchers in Cluster I are very aware of work that is going on outside their own fields, and open to its influence. The second feature of this map is the way that the research on the main foreign languages – French, German, Spanish and to a lesser extent Italian – is not at all compartmentalised. People working on one language seem to be very much aware of developments in the other languages, and there is no evidence here of the split between Modern Languages and English Language Teaching (ELT) which is a common feature of more recent research. The third feature worth noting is the detached Cluster III. Given the importance of cognates for L2 learners, it is very surprising that Altrocchi and Scatori are not better connected to the main cluster. This separateness may be due to the fact that Altrocchi and Scatori are principally interested in Italian, a language which is not otherwise well represented in the main cluster.

The legend to Figure 3 lists 15 people who I have designated as "Significant Influences" in this period – the best connected nodes in the map. This list is surprising in a number of ways. The first surprising feature is that only a handful of these names will be familiar to modern readers: West, Thorndike and Ogden stand out in this respect. The other Significant Influences in this period are much less familiar – their work is rarely cited in modern research. This immediately raises the question of why this work has been forgotten. The second feature is the emergence of Michael West as the most Significant Influence during this period. West remains an important reference point for ELT researchers (Smith, 2003), but I suspect that his work might not be familiar to most teachers of modern languages. This raises some interesting questions about the relationship between ELT and modern language teaching during this early period, specifically why ELT and modern language teaching seem to have diverged in their approaches to vocabulary. The third surprising feature is the fact that two of the most Significant Influences for this period are primarily educational psychologists with only a tangential interest in L2 vocabulary, and that a number of other Significant Influences, notably Henmon and Limper, were also active researchers in this wider area. The implication of this is that vocabulary research during this period is not an inward-looking research activity. On the contrary, the field, small though it is, has strong connections to wider research interests.

3.2 1951–1980

Given this promising beginning, we might expect that vocabulary research after 1950 would have developed strongly, but, as we see in Figure 1, what we

actually get is a period of steady decline. In fact, between 1951 and 1980, MLJ published only 42 papers dealing with vocabulary issues – only three papers every two years. It is not clear whether this represents a much wider, general shift away from vocabulary research, or one that is specific to MLJ. Only further analysis of other sources can answer this question. In the meanwhile, the MLJ data will be taken at face value, and a co-citation analysis of this data is shown in Figure 4. Given the small number of papers, and the low level of citations in these papers (the average number of citations is around 10, but only 16 papers reach double figures), this analysis needs to be treated with appropriate caution.

The analysis has identified six clusters for this period.

Cluster I, the small cluster in the north-west corner of the map, will be familiar from Figure 3. This cluster is the remains of the large cluster that dominated the research in the 1916–1950 period. Significantly, this work has now become detached from the main lines of research.

Clusters II and **III**, the two small clusters at the south-west of the map focussed on Carroll, also provide a link with the earlier work. Carroll is a new figure, but all the other members of these clusters – Hagboldt, Morgan, Thorndike, Lorge, Fries and Borroughs – appeared in the earlier map. Their influence is much diminished, however, and only Carroll's links with the main cluster prevents these clusters from joining Cluster I in the wilderness. The three remaining clusters are all composed of new names that did not appear in the earlier map.

Cluster IV, the smaller cluster on the western edge of the map, is a group of psychologists focussed around Jenkins. This group is predominantly interested in word associations in L1 speakers. They are not interested in L2 vocabulary, but

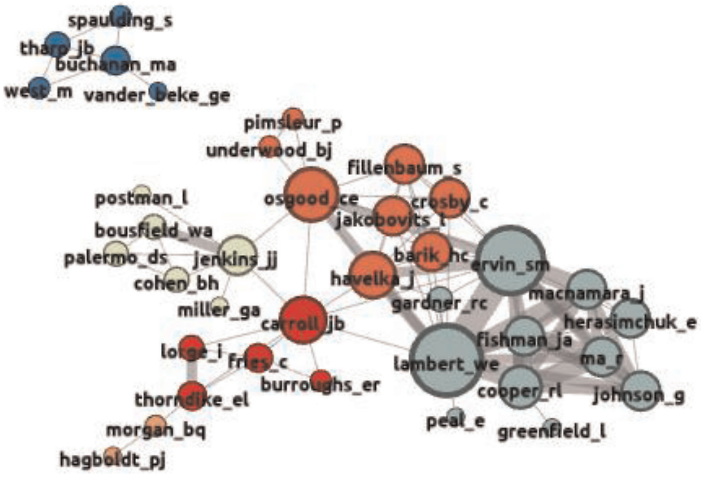

Figure 4. Co-citation analysis of vocabulary-focussed articles 1951–1980. Threshold for inclusion: two co-citations. The Significant Influences 1951–1980: Lambert, Ervin, Os-good, Havelka, Carroll, Cooper, Barik, Ma, Crosby, Johnson, Macnamara, Jenkins, Fishman, Herasimchuk, and Jakobovits.

their methods are often used by other researchers who use the word association methodology with L2 speakers and bilinguals.

The two remaining clusters represent an approach to L2 vocabulary which is very different from anything we see in Figure 3.

Cluster V, focussed on Lambert and Ervin, is a social psychology cluster. Lambert and his colleagues in Montreal were particularly interested in developing measures of language dominance in English–French bilinguals, and many of the test types that they explored were basically measures of vocabulary skills in these languages. This cluster also contains a group of sociolinguists centred around Fishman, and based in New York. This group was particularly interested in Hispanic speakers and used vocabulary data as a way of evaluating the way their subjects used English and Spanish in their everyday lives. This cluster, then, is really concerned with bilinguals in two very specific sociolinguistic contexts.

The final cluster, **Cluster VI**, is more difficult to characterise. Like Cluster V, this cluster is composed of psychologists, but the focus of this group is on cognitive psychology rather than social psychology. A particular concern of this cluster is the psychology of meaning.

Figure 4 suggests that the research published in MLJ during this period has not developed in the way we might have expected. There has been a significant shift away from word lists and frequency counts and a marked rise in psychological research. However, most of this work is not concerned with vocabulary acquisition: in the main clusters, only Pimsleur (Cluster VI) has an interest in language teaching. The other researchers are focussed on the behaviour of bilinguals. This interpretation throws a very stark light on vocabulary research during this period. We have already seen in Figure 1 that MLJ published very little vocabulary research between 1951 and 1980, and if we discount the papers by Lambert and Cook which furnish most of the co-citations that identify Clusters IV, V and VI, then we are left with two very small disconnected research islands. This shift in focus is reflected in the list of Significant Influences for this period (see the legend to Figure 4). None of the Significant Influences from the 1916 to 1950 period continues to be influential in the 1951–1980 – even major figures like Thorndike and West drop out of the Significant Influence table. More important, perhaps, is the fact that few of the Significant Influences for 1951–1980 will be familiar to modern researchers, and this provides a hint that the psycholinguistic and sociolinguistic approaches that dominate the 1951–1980 period have not left a lasting impression on L2 vocabulary acquisition research.

3.3 1981–2000

This conclusion is broadly confirmed by Figure 5, which shows an analysis of the 35 relevant papers published in MLJ between 1981 and 2000. Citation rates during this period are considerably higher than in the earlier periods – the average number of cited authors in these papers is 43.4. Consequently, I have adjusted the threshold for inclusion to take account of this. The map is based on a set of 418 co-citations which appear at least three times in the data set. These co-citations

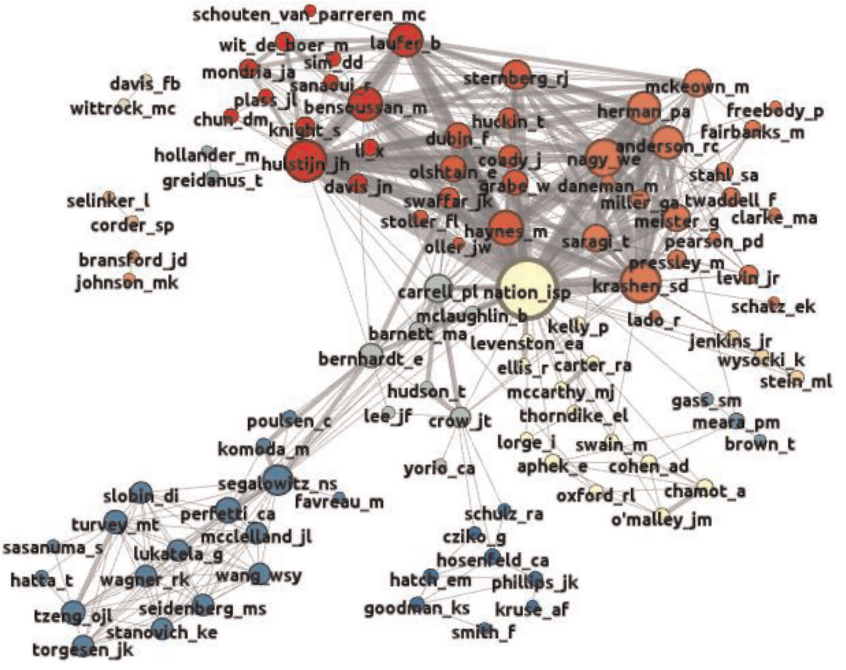

Figure 5. Co-citation analysis of vocabulary-focussed articles 1981–2000. Threshold for inclusion: three co-citations. The Significant Influences 1981–2000: Nation, Hulstijn, Krashen, Nagy, Haynes, Bensoussan, Laufer, Anderson, Herman, Segalowitz, Sternberg, Carrell, Meister, and McKeown.

identify 103 authors grouped into 13 clusters. Three of these clusters are detached from the main cluster grouping.

Again, the clusters are easily interpreted.

Cluster I, in the north-east corner of the map, is the largest cluster in this map. It contains six of the Significant Influences from this period (Anderson, Krashen, McKeown, Meister, Nagy and Herman). This cluster is mainly concerned with the differences between explicit vocabulary acquisition and incidental vocabulary acquisition from extended texts. Particularly striking is the large number of L1 reading researchers in this cluster. The cluster is very closely linked with Cluster IV.

Cluster II, the second largest cluster in the south-west corner of the map, is focussed on Segalowitz. This cluster is basically a group of psychologists with an interest in language processing. Most of these people do not work on L2 speakers, and the group as a whole is connected to the main clusters only through Segalowitz, Poulsen and Komoda's work on non-native speakers' performance.

Cluster III, the central cluster focussed on Nation, seems mainly to consist of people with an interest in Corpora and their implications for vocabulary acquisition, and a small group of people with an interest in vocabulary strategies. Nation's work on frequency counts explains his inclusion in this cluster. Nation is also massively cited alongside the members of Clusters I and IV, but the other members of this cluster have only weak links to other parts of the map.

Cluster IV, at the top edge of the map, contains three of the Significant Influences of this period (Hulstijn, Laufer and Bensoussan). This group is broadly interested in the factors which facilitate or inhibit the acquisition of L2 vocabulary. Knight, Davis, Chun and Plass represent a subgroup of researchers who deal with glosses.

Cluster V, focussed on Haynes and Sternberg, seems to be mainly concerned with reading in an L2, but this cluster is very tightly bound to the two large clusters: I and IV.

Cluster VI, at the southern edge of the map, is another reading cluster mainly composed of L1 reading theorists. It is notable that this cluster does not seem to have any links with L1 reading theorists who appear in Cluster I.

Cluster VII, focussed on Carrell and Bernhardt, is another reading cluster. I think this cluster is distinguished from the earlier clusters by an interest in Graded Readers. This cluster plays an important role in providing the links which prevent this map from fragmenting into disconnected clusters. The co-citations linking Segalowitz and Bernhardt, Crow and the L1-reading theorists in Cluster VI and Carrell's links with the clusters at the top of the map are particularly important in this regard.

Cluster VIII, Stein, Wysocki and Jenkins, has links with Nation and Krashen but is otherwise detached from the main cluster. This group were the authors of a 1984 paper dealing with the acquisition of L1 vocabulary through reading.

Cluster IX, Gass, Meara and Brown, does not have any single obvious theme. Meara and Gass both wrote overview articles during this period, while Brown authored a paper on vocabulary learning strategies.

Cluster X is loosely attached to Cluster IV; Hulstijn, Greidanus and Hollander jointly authored a frequently cited paper dealing with incidental vocabulary acquisition. This paper also analyses the effectiveness of glosses.

Cluster XI, Johnson and Bransford, points to a series of papers co-authored by these two authors. Their work is mainly focussed on sentence comprehension.

Cluster XII, Selinker and Corder are best known for their work on Interlanguage.

Cluster XIII, Wittrock and Davis, like many of the sources in Cluster IV, deals with the use of glosses in L2 reading.

There are a number of striking features in this map. The most obvious feature is the very large increase in the density of the map compared with the maps of the earlier periods. This feature partly reflects a dramatic change in citation practices at this time, but it also reflects a general rise in research output during the period 1981–2000, and a specific increase in vocabulary research as evidenced in Figure 1. The second feature worth noting is the very large number of clusters in this map. In the earlier maps, we have just a handful of clusters, but here we have a much larger number, indicating that we have a wide variety of different approaches to L2 vocabulary acquisition during this period, each with its own distinct set of emphases. However, two of these clusters, Clusters I and II, are very densely interconnected with many common reference points and this seems to indicate the

emergence of a very strong consensus about what questions L2 vocabulary research should address. Clusters I, II and III have no obvious antecedents in the earlier maps, and I interpret this to mean that the new consensus represents a sudden shift in the focus of vocabulary research. The theme that links these two clusters seems to be reading and comprehension – a slightly surprising feature, given the wider interest in communicative language learning during this period. In fact, the clusters located in the bottom half of the map, which include a distinctive group of L1 reading theorists in Cluster VI, do not appear to be involved in this consensus. They seem to be internally coherent in that the cluster members are frequently cited alongside each other, but these clusters have only loose co-citation links with the rest of the map and few links with each other.

The most striking feature of this map is the central role played by Paul Nation in linking the main cluster group together. Nation does not appear in the 1951–1980 map, though he was active during the latter part of this period. However, by 2000, he emerged as the biggest Significant Influence in L2 vocabulary research, with direct links to 8 of the 10 connected clusters in the map. Interestingly, the cluster analysis does not place Nation within either of the two largest clusters. Rather, he seems to belong to a cluster which contains people whose interests are less focused on L2 reading, and more concerned with strategies or with applications of corpus linguistics in language teaching. The fact that these people are frequently co-cited with Nation, but never cited with other vocabulary researchers is odd, but may simply reflect the fact that Nation's own work is much more wide ranging than some of the narrowly focussed work in Clusters I and II. Nation's work also tends to have a broader historical grounding than some of the other writers publishing in this period. His links with Cluster VII indicate that researchers outside the field are citing his work as prototypical of L2 vocabulary research.

In Figure 4, we note that some of the clusters are short-lived and the same feature can be found in Figure 5. The best example of this is the research strand exemplified by Clusters V and VI. This strand (another reading group) was very active in the earlier part of the 1981–2000 period, and its influence seems to have lessened after 1990. However, this group provides the only links between the main cluster grouping and the important cluster VII, the group focussed on Segalowitz. Cluster VII, which is strikingly well connected within itself, seems to consist of some heavyweight psychologists whose theoretical work is of great relevance to L2 vocabulary acquisition. The map suggests, however, that this work may have had only a limited influence on mainstream research as represented by Clusters I and II, and that this work is mediated through a small number of pivotal connections.

The legend to Figure 5 lists the most Significant Influences during this period. None of the Significant Influences in the previous period appears in this list, reinforcing the view that the new work represents something of a paradigm shift in L2 vocabulary research. For the first time here, we have a number of names which will be familiar to most researchers in the field. Nation, as we have already mentioned, has emerged as the outstanding Significant Influence during this period. This is most likely due to the seminal influence of his books (Nation, 1990, 2001), which summarised huge amounts of earlier research and rapidly became an

essential source for vocabulary researchers. The list also includes a number of people who are not based in North America – Hulstijn, Bensoussan and Laufer illustrate this trend. Particularly striking, however, is the presence of Anderson and Herman, and Sternberg and Nagy in the list. These people are predominantly L1 reading researchers, and this indicates that L1 reading research was a particularly important influence during this period. Two other Significant Influences are worth commenting on in more detail. Krashen is best known for his work on the Monitor Model, a general theory of L2 performance which was very influential in the 1980s. Krashen published a particularly important paper in MLJ in 1989, arguing that most vocabulary is acquired by reading and not by direct instruction. This massively cited paper encapsulated an attractive idea for teachers, and influenced the direction research was to take out of all proportion to the strength of the data it reported. A much more interesting Significant Influence is Segalowitz who stands outside from the main cluster grouping. Segalowitz was an active member of the Montreal research group which featured in Figure 4, and should perhaps be seen as the successor to Lambert, the most Significant Influence in the 1950–1980 period. Segalowitz is particularly important because he provides the only link between research in L2 vocabulary acquisition and some important research in cognitive psychology which has implications for L2 vocabulary research. Segalowitz continues to work in the area of bilinguals' lexical performance, with a particular emphasis on lexical fluency. This is a complex topic with obvious implications for models of L2 lexical development, but on the evidence of Figure 5, its impact on mainstream research in L2 vocabulary seems to be slight.

To summarise, then, during the 1981–2000 period, MLJ published a lot of work on L2 vocabulary acquisition, but this work largely ignores all the previous work published in the journal. Instead, we find a marked focus on L2 reading skills, with a number of different approaches addressing this question. The main clusters rely heavily on Paul Nation's work, and a dense network of co-citations links Nation and the other members of the two main clusters. A few smaller clusters are loosely connected to the main cluster, but the members of these clusters are rarely cited alongside the members of the main clusters. The earlier work of this period is strongly influenced by the work of cognitive psychologists, but there is a suggestion that the more recent L2 research is becoming more autonomous and less outward-looking.

3.4 2001–2010

Figure 6, which shows a map of the research published in the journal between 2001 and 2010, broadly confirms the trends we note in Figure 5. This map, which can be thought of as a map of current trends in L2 vocabulary research, needs to be treated with some caution, since it covers only 10 years of research and only 20 papers were published during this period. In spite of this, the number of co-citations in Figure 6 is about the same as the number of co-citations that Figure 5 is based on – the average number of sources cited in a paper published in this period is 71. Setting a threshold of three co-citations for inclusion gives us a map consisting of 464 co-citations linking 68 authors in 7 clusters, one of which is detached.

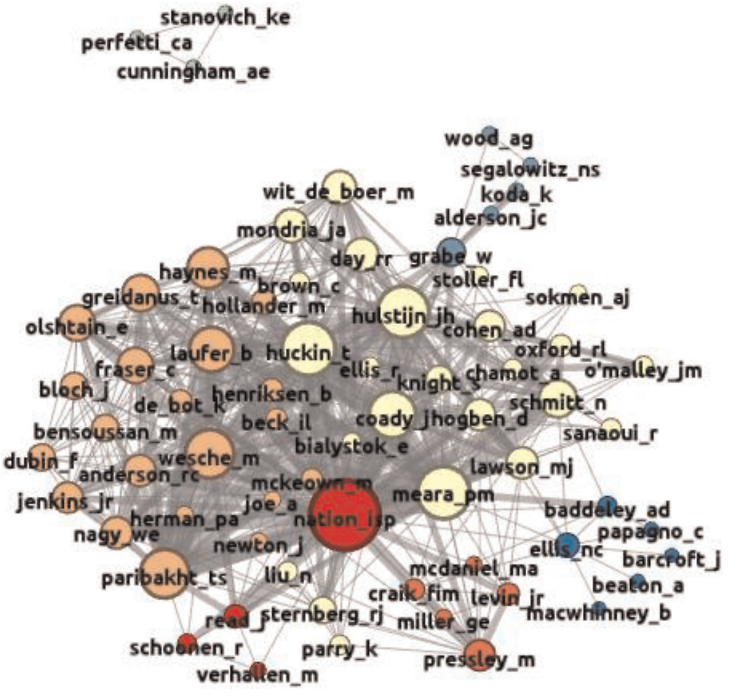

Figure 6. Co-citation analysis of vocabulary-focussed articles 2001-2010. Threshold for inclusion: three co-citations. The Significant Influences 2001–2010: Nation, Huckin, Meara, Hulstijn, Paribakht, Wesche, Coady, Laufer, Haynes, Schmitt, Greidanus, Fraser, Olshtain, Day and Mondria.

Gephi identifies seven clusters in this data.

Cluster I, with Nation, Wesche, Paribakht and Laufer as foci, is mainly concerned with L2 reading and particularly concerned with the uptake of unknown words through inferencing.

Cluster II, with Huckin, Hulstijn and Meara as foci, seems to consist mainly of a group of people who have written extensively about L2 vocabulary acquisition and use a wide variety of different approaches to this topic in their own research.

Cluster III, focussed on Nick Ellis, is a group of cognitive psychologists concerned with models of memory.

Cluster IV, the small cluster focussed on Grabe, mainly researches the mechanics of L2 reading.

Cluster V, focussed on Pressley, is a group whose work mainly deals with memory for words, with a particular emphasis on the keyword method.

Cluster VI, Read, Schoonen and Verhallen, is basically a testing group.

Cluster VII, the detached cluster comprising Stanovich, Perfetti and Cunningham, is another group of psychologists who work on L1 reading.

Again, this map has a number of features which need comment. The first feature is that all of the Significant Influences identified fit into the two main

clusters. This suggests that the consolidation we noted in the 1981–2000 period seems to have intensified. The second feature is that the influence from Cognitive Psychology seems to have shifted. In this map, we have three separate groups of this sort. Connections with the rest of the network seem to be stronger than was the case in Figure 5, though Cluster VII is detached from the main clusters, and the other two clusters (Clusters III and IV) link mainly to specific individuals in Cluster II. Cluster II seems to be distinguished from Cluster I mainly by a surprising lack of co-citations with Cognitive Psychologists. The third feature is that, despite these changes, the clusters in Figure 6 can be readily identified with the clusters in Figure 5. This is the first time that we have seen any kind of continuity in these maps.

This continuity also appears in the list of Significant Influences in Figure 6. Four of the Significant Influences identified in Figure 5 – Nation, Hulstijn, Laufer and Haynes – continue to be Significant Influences in Figure 6, with Nation maintaining his position as by far the most co-cited source in the field. A number of people who played only a small role in Figure 5 have become Significant Influences in Figure 6 – notably Paribakht, Wesche, Meara and Schmitt. Fraser appears as a Significant Influence, despite not having a presence in the 1981–2000 map. Four Significant Influences in the 1981–2000 map do not appear in this map: Krashen, Carrell, Saragi and Meister.

3.5 Overview

We can now return briefly to the overview of MLJ research in Figure 2, which, in the light of the preceding analyses is much easier to interpret. Surprisingly, this map now appears to be largely organised along historical lines and it strongly emphasises the lack of continuity in the research published in MLJ.

The large central clusters, focussed on Nation and Hulstijn, consist mainly of people who are currently research-active. These clusters correspond to the central clusters in Figures 5 and 6. The large cluster at the eastern edge of the map, dominated by Segalowitz, captures the psychological theory that underpins a lot of the recent research. The small cluster that includes Carrell and Goodman captures the reading research that was characteristic of the 1970–1980 period, but is referred to only rarely in the most recent research. The cluster focussed on Thorndike and West at the south-east of the main clusters identifies the early research covering the period up to 1950. This cluster is only tenuously attached to the main clusters. The large cluster at the south-east corner of the map contains the psycholinguists and sociolinguists from the earlier 1960 to 1970 period. This group now forms a detached island, indicating that though it was once important, it retains only limited relevance for current research.

4 Discussion

The question we posed in the introduction to this paper was whether an analysis of the vocabulary research in a single journal might provide a skeleton for a more detailed and in-depth study of how L2 vocabulary research has developed over the last 90 years. The answer to this question seems to be a tentative yes, but

with some reservations. The analysis I have presented suggests that we can identify four broad periods of activity. 1916–1950 defines a period when most research was concerned with the development of reliable word lists for modern language teaching. 1951–1980 signals a short-lived shift away from word lists towards an approach that is more driven by cognitive psychology and sociolinguistics. 1981–2000 represents the start of modern research into L2 vocabulary acquisition, and the development of a broad consensus about what questions this research should be addressing. 2001–2010 sees the further consolidation of this consensus and the emergence of a new orthodoxy centred on Paul Nation's work. Most readers of this paper will find this is a plausible, if incomplete, account of how the field has developed since 1916. The central theme which seems to run through all the work reported here is a concern with reading in a second language – how to provide books that learners can usefully read, how reading skills can serve as an index of proficiency, how reading is an essential source of new vocabulary and how learners infer the meaning of new words they encounter in their reading. This emphasis, and its enduring role in the research, came as something of a surprise to me.

The question we have to ask now is whether the historical picture which emerges from MLJ is indeed an accurate portrayal of the way vocabulary research developed in the period under review. This question is surprisingly difficult to answer, as we do not have any similar studies based on other long-running journals against which this account could be compared. We might ask, for example, whether the apparent decline in vocabulary research between 1950 and 1980 really reflects what was going on in the research, or is it just the result of MLJ publishing fewer, longer articles? Was vocabulary research being published somewhere else because MLJ was hostile towards this sort of research? What was the editorial politics of MLJ during this period? Questions of this sort are obviously important, but need to be looked at in a wider context which is not currently available to us. Furthermore, even if we accept the raw figures from MLJ at face value there are a number of features which make it difficult to interpret the data. Most modern accounts of vocabulary research, for instance, do not make much reference to very early work in the field – Schmitt (2000), for example, cites in a very extensive bibliography only six papers published before 1945, two very early papers published in 1885 and 1889, and four later papers published in 1936, 1941, 1942 and 1946. Clearly, some further work is needed to establish how far this early work was influential outside the confines of MLJ. It is noticeable that some of the early papers published in MLJ refer to work which appeared in the *Journal of Educational Psychology* (JEP). The JEP papers are considerably more technical than the work appearing in MLJ and typically deal with questions of vocabulary assessment. This suggests that there might have been separate parallel developments in vocabulary research in other important journals appearing at this time. A preliminary trawl suggests that the JEP published a substantial number of papers dealing with aspects of vocabulary in a second language between 1916 and 2010, but the publication of these papers was heavily influenced by an explicit editorial policy. A co-citation analysis of this material might be able to put the MLJ papers in a wider context.

The North American focus provided by the Journal becomes less striking as we move away from the very early volumes, and by the time we get to the map

shown in Figure 5, North Americans no longer have a monopoly of the places in the Significant Influences list. Nation is based in New Zealand, Hulstijn in the Netherlands, Bensoussan and Laufer in Israel. This internationalising trend becomes even more noticeable in the very recent research shown in Figure 6, where only a minority of the Significant Influences are based in the USA. Presumably, this trend reflects the growing importance of cheap air travel, the influence of better telecommunications and the growth of the internet on the way research develops. However, it is also possible that what we are picking up here is a move by US researchers away from vocabulary issues into some other area leaving a vacuum to be filled by researchers from elsewhere. Again, we cannot really assess this analysis without looking at comparable data in other journals.

Another feature of the maps based on the MLJ is the complete absence of any reference to some research initiatives that we might have expected to play a role in a more general history of research into vocabulary in an L2. There are three very obvious examples of this. Firstly, it is surprising that special purpose vocabularies do not appear as a cluster in any of the maps, as we might have expected this sort of work to be an area of some importance in the Journal. MLJ did in fact publish some work of this sort in the 1940s and early 1950s, but it appears to be practically rather than theoretically motivated, and the papers tend not to cite other work. In the 1940s, for example, the Journal published a number of "air vocabularies" – word lists that might be of use to military aviators: *fuel, take-off, landing strip, bomb, machine-gun, prisoner of war* and so on – but these lists seem to be based on an ad hoc understanding of the situations involved, rather than something more principled that might have developed into a research theme. Secondly, and rather surprisingly given the early concerns of the journal, we find few signs of any research related to corpus linguistics, with the result that major initiatives like the COBUILD dictionary do not figure in the maps. In a comprehensive history, the Birmingham group and the Nottingham group, both based in the UK, might have been expected to play a larger role than they do here. Thirdly, there is a serious lack of engagement with some important strands of European research: again this is surprising given that the Journal is ostensibly concerned with the teaching of modern European languages, and that some early papers in the Journal report research visits to European centres where innovative work was taking place. A considerable amount of high-quality research was being undertaken, particularly in the Netherlands in the 1960s and 1970s. Both Spain and France also have a long tradition of researching vocabulary, with a particular emphasis on accessibility, and a significant omission from the maps is any mention of the *Français fondamental* project directed by Gougenheim (Gougenheim, Michéa, Rivenc, & Sauvageot, 1956). Belgian and Dutch researchers were actively involved in developing word counts for the major European languages during the 1960s, and though this work was influential on the teaching of English, its impact on the foreign language teaching reported in MLJ seems to be negligible. This suggests that an analysis of a journal with professional aims similar to those of MLJ but with a different geographical base might generate a very different set of maps and provide a different looking history for the field.

Other topics which one might also expect to appear in a more comprehensive history are also missing from these maps. These topics include vocabulary testing

and assessment, which we might have expected to be core areas for the MLJ, and the use of brain imaging methods to examine how L2 speakers perform, an area which we might have expected to be a feature of the more recent research. In general, the maps suggest that "outside influences" have become less important in the later issues of the journal. By the time we reach the more recent research, shown in Figure 6, all the Significant Influences are people who would primarily describe themselves as vocabulary researchers, and the input from other disciplines – L1 acquisition, psycholinguistics, language loss and cognitive psychology in particular – has fallen to alarmingly low levels. Where this work does appear in the maps, it tends to be channelled through a small number of people. Segalowitz, in Figure 5, is a particularly striking example of this role. Certainly, the picture which emerges in Figure 6 suggests that the vocabulary research published in MLJ has become increasingly monolithic, with very few points of reference outside itself. Whether this reflects a general characteristic of research or a more specific narrowing of horizons in the field at large remains to be seen.

It would be nice to be able to say that this analysis of Vocabulary research published in MLJ provides a reliable outline sketch of the vocabulary research work that was being undertaken throughout the twentieth century, and that all that remains to be done is to fill in some of the details. Certainly, the structure which emerges from the map in Figure 3 is a plausible account of how North American – or more specifically USA – L2 vocabulary research developed during the twentieth century, and the broad lines of this map will be familiar to most researchers working in the field. However, I have also signalled some striking gaps in the maps, and most readers will be aware of other areas of vocabulary research that might have been added to that list. The best interpretation of these features seems to be that a single journal is perhaps not very representative of an entire field, but captures only a part of that work. This, of course, raises some interesting questions about the extent to which one journal's coverage differs from that of another, which journals best capture the available research and how many journals one would need to analyse in order to get a stable picture identifying most of the important features of research in a given time span. However, the earlier work reported in MLJ does not lend itself well to the co-citation analysis methods used in this paper, as the small number of citations means that we need to use very large timespans to get enough data, and this suggests that the co-citation approach may work well only with more recent sources. Nevertheless, even with the later research, the extended time spans are a problem. The specific time spans that I have used are motivated primarily by the need to group papers into similar sized data sets. To some extent, the divisions are arbitrary, and slightly different cluster analyses emerge when we divide the data up in other ways. Clearly, we need to ask whether extended time spans of the sort analysed here are best suited to capturing new developments in a field, or whether thinner slices from more journals might have provided us with better data. The maps do strongly suggest that there is a massive shift in research culture somewhere between 1980 and 1990. We cannot really pin these shifts down with any confidence but the MLJ data suggests that it would be worth looking at this period in much more detail.

5 Conclusions

This paper has presented a co-citation analysis of the vocabulary research published in *Modern Language Journal* between 1916 and 2010, and has argued that the co-citation approach is able to throw some interesting light on the way the field has developed during this period. Using a single journal as a data source is not entirely satisfactory, as the picture which emerges from the co-citation maps is less complete than we might have hoped. However, it ought to be possible to supplement these maps with data from other journals and to establish how much data we need before a stable mapping emerges. Some problems with early citation practice were also noted. This suggests that the co-citation method might not produce reliable data from sources published before 1960, though data from later periods do seem to be very rich and capable of being mined in a number of different ways.

At first sight, the issues addressed here might appear to be rather academic. However, readers will be aware that bibliometric data of the sort reported in this paper are increasingly being used by research agencies to evaluate research outputs, and these evaluations have very significant impacts on the way we work. It would be reassuring to feel that these agencies are using data that we can have some confidence in. This is particularly important for research areas like L2 vocabulary acquisition, which historically have not had a single natural outlet.

Notes

1. Another comprehensive database is maintained by Paul Nation and can be found at http://www.victoria.ac.nz/lals/resources/vocrefs/bibliography. It is difficult to make direct comparisons between these two sources as they are differently structured. Nation's list is very comprehensive and organised by topic. However, it lacks the search facilities provided by the VARGA site. Readers interested in following up this comparison might find useful a paper by Murray, Ke, and Börner (2006), which explores the use of bibliometric analyses with personal bibliography files.
2. From Figure 3 onwards, the maps are all standardised using the same settings in the Gephi programme. Nodes are sized according to the number of connections they have with other nodes (their "degree"). It would have been possible to draw the maps so that the size of a node was directly determined by its degree. In this case, a node with 100° would be 20 times larger than a node with 5°. This approach generated very large maps, which were impossible to reproduce legibly in the space available. As a compromise, the nodes in Figures 3–6 are scaled in relative terms. The nodes with the highest degree in each map are the same large size, and the nodes with the lowest degree are the same small size. Other nodes between these extremes are sized on a proportional scale. Since degree scores are not evenly spread out over the entire interval, we get rather a lot of smaller nodes, particularly in the larger maps. Full colour versions of these maps are available at http://www.lognostics/co/uk/maps.

References

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

Gougenheim, G., Michéa, R., Rivenc, P., & Sauvageot, A. (1956). *L'élaboration du français fondamental (Ier degré)* [The development of Français fondamental (Level 1)]. Paris: Didier.

Hellqvist, B. (2010). Referencing in the Humanities and its implications for citation analysis. *Journal of the American Society for Information Science and Technology, 61* (2), 310–318. doi:10.1002/asi.21256

Meara, P.M. (n.d.). *The vocabulary acquisition research group archive (VARGA)*. Retrieved from http: //www.lognostics.co.uk/varga/

Meara, P.M. (2012). The bibliometrics of vocabulary acquisition: An exploratory study. *RELC Journal, 43* (1), 7–22. doi:10.1177/0033688212439339

Meara, P.M. (forthcoming). *Life before nation: Bibliometrics and L2 vocabulary studies in 1982*.

Murray, C., Ke, W., & Börner, K. (2006). Mapping scientific disciplines and author expertise based on personal bibliography files. In: *Proceedings of the Conference on Information Visualization* (pp. 258–263). Washington, DC: IEEE Computer Society.

Nation, I.S.P. (1990). *Teaching and learning vocabulary.* Boston, MA: Heinle and Heinle.

Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139524759

Price, D. (1965). Networks of scientific papers. *Science 149* (3683), 510–515. doi:10.1126/science.149.3683.510

Price, D. (1986). *Little science, big science… and beyond*. New York: Columbia University Press.

Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.

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. doi:10.1002/asi.4630240406

Smith, R.C. (Ed.). (2003). *Teaching English as a foreign language, 1912–1936: Pioneers of ELT* (Vol. 3). London: Routledge.

White, H.D., & Griffith, B.C. (1981). Author cocitation: A literature measure of intellectual structure. *Journal of the American Society for Information Science, 32*, 163–171. doi:10.1002/asi.4630320302

Zuccala, A. (2006). Modeling the invisible college. *Journal of the American Society for Information Science and Technology, 57* (2), 152–168. doi:10.1002/asi.20256

Appendix: The 201 papers used as source material for this report (listed by date).

Bagster-Collins, E.W. (1918). A brief study showing the relation between the vocabulary and treatment of the annotated reading text. *MLJ, 2* (8), 341–345.

Weigel, J.C. (1919). The acquisition of a vocabulary. *MLJ*, *3* (8), 339–360.

Bovée, A.G. (1919). Teaching vocabulary by the direct method. *MLJ, 4* (1), 63–72.

Wadepuhl, W. (1923). A standardized vocabulary for elementary German. *MLJ, 8* (1), 23–30.

Keller, M.V. (1923). The necessity of teaching a basic vocabulary in modern language work. *MLJ, 8* (1), 35–40.

Jamieson, E.I. (1924). A standardized vocabulary for elementary Spanish. *MLJ, 8* (6), 325–333.

Hubman, S. (1924). The business of getting a vocabulary. *MLJ, 9* (3), 159–166.

Wilkins, E.H. (1924). Suggestions as to method in making a vocabulary. *MLJ, 9* (3), 167–169.

Morgan, B.Q. (1925). The Chicago MLT adopts a German vocabulary. *MLJ, 9* (7), 397–457.

Cartwright, C.W. (1925). A study of the vocabularies of eleven Spanish grammars and fifteen Spanish reading texts. *MLJ, 10* (1), 1–14.

Eddy, H.M. (1926). The French element in English. *MLJ, 10* (5), 271–280.

Morgan, C. (1926). Vocabulary analysis of a second year Spanish text. *MLJ, 10* (7), 427–430.

Wood, B.D. (1927). A comparative study of the vocabularies of sixteen French textbooks. *MLJ, 11* (5), 265–289.

Johnson, C.L. (1927). Vocabulary difficulty and textbook selection. *MLJ, 11* (5), 290–297.

Broom, E., & Contreras, M.S. (1926). A background vocabulary list in Spanish. *MLJ, 11* (7), 459–463.

Dexter, E.F. (1928). An analysis of a first year Spanish vocabulary. *MLJ, 12* (2), 272–278.

Gay, L.M. (1928). Reflections on the Henmon Word Book. *MLJ, 12* (5), 365–369.

Roulston, R.B. (1929). A plea for more and better special vocabularies. *MLJ, 13* (4), 304–307.

Simmons, L.V. T. (1929). A vocabulary count based on three German dramas. *MLJ, 14* (1), 33–36.

Blackburn, M. (1930). An analysis of the vocabulary of two recent French grammars. *MLJ, 14* (6), 431–441.

West, M. (1930). Speaking vocabulary in a foreign language. *MLJ, 14* (7), 509–521.

Maronpot, R.P. (1930). Teaching and testing vocabulary on a one-language basis. *MLJ, 14* (7), 554–560.

Talmey, M. (1930). Word composition in a logical language. *MLJ, 15* (3), 200–214.

Engel, E.F. (1931). The use of a standardized vocabulary in beginning German. *MLJ, 15* (4), 281–291.

Breageale, E. (1931). The value and use of the new word lists. *MLJ, 15* (4), 298–299.

Fotos, J.T. (1931). Word and idiom frequency counts in French and their value. *MLJ, 15* (5), 344–353.

FitzGerald, T. (1931). Textbook vocabularies. *MLJ, 16* (2), 135–139.

Coleman, A.P. (1931). The basic vocabulary in Polish. *MLJ, 16* (2), 140–146.

Scatori, S. (1932). Deceptive cognates in Spanish. *MLJ, 16* (5), 396–401.

Hagboldt, P.J. (1932). Types of memory in language study. *MLJ, 16* (7), 583–590.

Arnold, H.H. (1932). A list of graded vocabularies and a method of grading. *MLJ, 16* (8), 644–655.

Handschin, C.H. (1932). The question of the most economical learning of the German vocabulary. *MLJ, 17* (3), 195–199.

Rose, E. (1933). Vocabularies in German textbooks. *MLJ, 17* (5), 334–341.

Bement, N.S., & Smith, E.M. (1933). Standardizing French vocabulary for beginners. *MLJ, 17* (8), 579–591.

Grimes, M.E. (1933). French noun endings and vocabulary building. *MLJ, 18* (2), 93–99.

Morgan, B.Q. (1933). A minimum standard vocabulary for German. *MLJ, 18* (3), 145–152.

Haygood, J.D. (1933). The amount and composition of a minimum essential French reading vocabulary. *MLJ, 18* (3), 177–189.

Bovée, A.G., Coleman, A., Eddy, H.M., Russell, P., & Tharp, J.B. (1934). A basic French vocabulary. *MLJ, 18* (4), 238–274.

Porterfield, A.W. (1934). The treatment of vocabularies in textbooks. *MLJ, 18* (7), 451–459.

Morgan, B.Q. (1934). Derivatives in the new basic French vocabulary. *MLJ, 18* (7), 460–462.

Schobinger, E. (1934). Adventures with words. *MLJ, 19* (1), 9–14.

Tharp, J.B. (1934). The basic French vocabulary and its use. *MLJ, 19* (2), 123–131.

Gilbert, L.C., & Loobourow, G.C. (1935). High-school spelling of English and foreign language words. *MLJ, 19* (4), 266–270.

Wilson, W.E. (1935). A supplementary vocabulary list for Spanish composition. *MLJ, 19* (6), 439–440.

Jones, W.K. (1935). Deceptive doubles: Twin words--With a difference. *MLJ, 20* (1), 15–18.

Skinner, L.H. (1935). A comparative study of the vocabularies of forty-five Italian textbooks. *MLJ, 20* (2), 67–84.

Kaulfers, W.V. (1936). Interpretative vocabulary-exercises for beginners in French. *MLJ, 20* (7), 396–402.

Skinner, L.H. (1936). An analysis of the vocabulary-difficulty of forty-five Italian textbooks. *MLJ, 20* (7), 411–415.

Vowles, G.R. (1936). Norwegian loan-words. *MLJ, 21* (1), 16.

Kurath, W., & Stalnaker, J.M. (1936). Two German vocabulary tests. *MLJ, 21* (2), 95–102.

Sharp, S.L. (1936). German textbook vocabularies and a supplementary dictionary. *MLJ, 21* (3), 157–161.

West, M. (1937). The present position in vocabulary selection for foreign language teaching. *MLJ, 21* (6), 433–437.

Coleman, A. (1937). A minimum French idiom list. *MLJ, 21* (8), 569–576.

Kennedy, G.A. (1937). A minimum vocabulary in modern Chinese. *MLJ, 21* (8), 587–592.

Walpole, H. (1937). The theory of definition and its application to vocabulary limitation. *MLJ, 21* (6), 398–402.

Lorenz, C.M. (1937). Foreign language words and phrases in American periodicals. *MLJ, 22* (2), 92–103.

West, M. (1937). ‘‘The reading approach’’ and ‘‘the new method system’’. *MLJ, 22* (3), 220–222.

Hagboldt, P. (1938). The work of Michael West. *MLJ, 22* (6), 411–419.

Wade, G.E. (1938). A note on Spanish vocabulary. *MLJ, 22* (6), 433–436.

Smith, H.E. (1938). Increase your vocabulary. *MLJ, 23* (1), 41–45.

Tate, H.L. (1938). Two experiments in reading-vocabulary building. *MLJ, 23* (3), 214–218.

Gibson, A.D. (1939). The vocabulary of sports in French. *MLJ, 23* (5), 350–356.

Tharp, J.B. (1939). The measurement of vocabulary difficulty. *MLJ, 24* (3), 169–178.

Miller, M.M., Farr, G. (1939). Student recognition of some Spanish-English cognates. *MLJ, 24* (3), 216–220.

Delcourt, J. (1940). On the impressive possibilities of the English vocabulary. *MLJ, 25* (1), 4–13.

Morgan, B.Q. (1940). Cognates and the minimum standard German vocabulary. *MLJ, 25* (1), 26–30.

Lembi, D.P., & Kaulfers, W.V. (1940). French expressions used in English: An exploratory vocabulary unit for the first week of beginning French or General Language. *MLJ, 25* (3), 174–186.

Johnston, M.C. (1941). Spanish-English cognates of high frequency. *MLJ, 25* (5), 405–417.

Stevens, L.C. (1941). The problem of vocabulary levels in teaching French. *MLJ, 25* (8), 633–636.

Henninger, G.A. (1942). Some questions for a language-vocabulary course. *MLJ, 26* (1), 27–34.

Morgan, E.A. (1942). Teaching the gender of nouns. *MLJ, 26* (5), 352–357.

Moore, A.Z. (1942). An experiment designed to measure the comparative achievement in vocabulary and reading ability of second year classes in French and Spanish. *MLJ, 26* (5), 358–360.

Scanlon, C. (1942). The purpose of graded texts. *MLJ, 26* (6), 421–423.

Redick, J. (1942). Another aid in teaching vocabulary. *MLJ, 26* (8), 587–588.

Peacock, V.L. (1942). Foreign words in everyday reading matter. *MLJ, 26* (8), 589–591.

Dummer, E.H. (1943). A new vocabulary test. *MLJ, 27* (1), 21–24.

Stevens, L.C. (1943). Textbook vocabularies and deceptive cognates in Spanish. *MLJ, 27* (2), 116–118.

Huebner, T. (1943). An air vocabulary of 100 Words. *MLJ, 27* (5) 353–355.

Lindgren, H.C. (1943). Understanding foreign terms used in English. *MLJ, 27* (7), 505–507.

Eaton, H.S. (1943). A new course in Basic Language. *MLJ, 27* (8), 578–581.

Henninger, G.A. (1944). In defense of dictionaries and definitions. *MLJ, 28* (1), 29–39.

Liebesny, H.J. (1944). Vocabulary learning enjoyable: Making a proper use of proper dictionaries. *MLJ*, *28* (2), 182–189.

Richards, I.A. (1945). BASIC English. *MLJ, 29* (1), 60–65.

Willis, E. (1944). Environmental vocabulary. *MLJ, 28* (3), 239–245.

Seibert, L.C. (1945). A study of the practice of guessing word meanings from a context. *MLJ, 29* (4), 296–322.

Cook, P. (1945). Iowa place names of foreign origin. *MLJ, 29* (7), 617–628.

Steinbugler, J.L. (1945). English–German vocabulary. *MLJ, 29* (8), 698–704.

Vance, D.G. (1946). Correlation of Latin and Spanish. *MLJ, 30* (4), 208–216.

Greiner, O.A. (1946). Words, not grammar. *MLJ, 30* (6), 347–351.

Vittorini, D. (1946). Formation of Italian nouns from English equivalents. *MLJ, 30* (8), 604–611.

Fitz Gerald, T.A. (1947). A plea for graded reading texts. *MLJ, 31* (6), 359–362.

Vittorini, D. (1947). Formation of Italian adjectives and verbs from English equivalents. *MLJ, 31* (3), 138–146.

Russo, G.A. (1947). A combined Italian word list. *MLJ*, *31* (4), 218–240.

Josselson, H.H. (1947). Russian word count. *MLJ, 31* (7), 422–425.

Vittorini, D. (1947). Cultural material and vocabulary in the study of romance languages. *MLJ, 31* (8), 486–488.

Eoff, S., & Bull, W.E. (1948). A semantic approach to the teaching of foreign languages. *MLJ, 32* (1), 3–13.

Iacobelli, M. (1949). The semantic discipline. *MLJ, 33* (1), 16–22.

Wuest, A. (1949). French vocabulary aids. *MLJ, 33* (2), 122–123.

Bull, W.E. (1950). Spanish word counts: Theory and practice. *MLJ, 34* (1), 18–26.

Abraham, R.D. (1950). Fixed order of coordinates: A study in comparative lexicography. *MLJ, 34* (4), 276–287.

Resnick, S. (1950). English words in Spanish literature. *MLJ, 34* (3), 204–208.

Russell, H.J. (1950). What is readable and what is not readable in a foreign language? *MLJ, 34* (4), 270–275.

Parker, F. (1952). The teaching and development of Russian vocabulary. *MLJ, 36* (3), 135–136.

Shaewitz, L. (1953). Semantics for modern languages. *MLJ, 37* (4), 177–180.

Spaulding, S. (1953). Three-dimensional word repetition in reading material. *MLJ, 37* (5), 226–230.

Lockwood, E., & Lockwood, W.B. (1953). Vocabulary difficulties in English and German. *MLJ, 37* (6), 287–292.

Nabholz, N. (1953). Applied philology. *MLJ, 37* (7), 347–350.

Prince, J.R. (1954). Philology at work. *MLJ, 38* (2), 75–79.

Kownacki, S. (1955). Teaching foreign languages in specialized fields. *MLJ, 39* (7), 351–352.

Richards, E.S., & Appel, J.E. (1956). The effects of written words in beginning Spanish. *MLJ, 40* (3), 29–133.

Crawford, P., & Vittorini, D. (1956). English cognates and Spanish masculine nouns. *MLJ, 40* (7), 391–395.

Spaulding, S. (1956). A Spanish readability formula. *MLJ, 40* (8), 433–441.

Keppler, K. (1957). Problems of German-English dictionary making. *MLJ, 41* (1), 26–29.

Hammer, C. (1957). Stress the German–English Cognates! *MLJ, 41* (4), 177–182.

Ianucci, J.E. (1957). Meaning discrimination in bilingual dictionaries: A new lexicographical technique. *MLJ*, *41* (6), 272–281.

Hietsch, O. (1958). Meaning discrimination in modern lexicography. *MLJ, 42* (5), 232–234.

Salling, A. (1958). An essay in comparative vocabulary study. *MLJ, 42* (5), 222–225.

Chandler, R.E. (1958). Spanish–Portuguese transfer. *MLJ, 42* (7), 329–336.

Beberfall, L. (1959). The verb and the beginning student of Spanish. *MLJ, 43* (1), 26–28.

Volman, R.J. (1959). Similarities of gender in basic Spanish and French vocabularies. *MLJ, 43* (1), 21.

Spotts, L.H. (1959). Foundations of vocabulary selection for the teaching of Hebrew in America. *MLJ, 43* (6), 281–287.

Di Blasi, S. (1959). Problems in Italian lexicography. *MLJ, 43* (7), 316–320.

Lowe, R. (1960). French name words in the classroom. *MLJ, 44* (2), 71–74.

Meyersteen, R.S. (1960). Bilingual dictionaries and applied linguistics. *MLJ, 44* (3), 163–167.

Russo, J.A. (1962). An Italian idiom list. *MLJ, 46* (1), 14–21.

Lambert, W.E. (1963). Psychological approaches to the study of language: Part I: On learning, thinking and human abilities. *MLJ, 47* (2), 51–62.

Lambert, W.E. (1963). Psychological approaches to the study of language: Part II: On second-language learning and bilingualism. *MLJ, 47* (3), 114–121.

Hakes, D.T. (1965). Psychological aspects of bilingualism. *MLJ, 49* (4), 220–227.

Hadlich, R.L. (1965). Lexical contrastive analysis. *MLJ, 49* (7), 426–429.

Pimsleur, P. (1967). A memory schedule. *MLJ, 51* (2), 73–75.

Preller, A.G. (1967). Some problems involved in compiling word frequency lists. *MLJ, 51* (7), 399–402.

Adkins, P.G. (1968). Teaching idioms and figures of speech to non-native speakers of English. *MLJ, 52* (3), 148–152.

Cooper, R.L., & Greenfield, L. (1969). Word frequency estimation as a measure of degree of bilingualism. *MLJ, 53* (3), 163–166.

Cooper, R.L. (1969). Two contextualized measures of degree of bilingualism. *MLJ, 53* (3), 172–178.

Berney, T.D., & Cooper, R.L. (1969). Semantic independence and degree of bilingualism in two communities. *MLJ, 53* (3), 182–185.

Fertig, S., & Fishman, J.A. (1969). Some measures of the interaction between language, domain and semantic dimension in bilinguals. *MLJ, 53* (4), 244–249.

Ronch, J., Cooper, R.L., & Fishman, J.A. (1969). Word naming and usage scores for a sample of Yiddish–English bilinguals. *MLJ, 53* (4), 232–235.

Darian, S. (1969). Backgrounds of modern language teaching: Sweet, Jespersen, and Palmer. *MLJ, 53* (8), 545–550.

Smith, P. (1969). Teaching vocabulary. *MLJ, 53* (8), 531–553.

Leong, C.K. (1972). A study of written Chinese vocabulary. *MLJ, 56* (4), 230–234.

Baetens Beardsmore, H., & Lee, E.J. (1973). Teaching English as a foreign language for special purposes. *MLJ*, 57(7), 343–348.

Reinert, H. (1976). One picture is worth a thousand words? Not necessarily! *MLJ, 60* (4), 160–168.

Cummins, J. (1977). Cognitive factors associated with the attainment of intermediate levels of bilingual skills. *MLJ, 61* (1/2), 3–12.

Cornu, A.-M. (1979). The first step in vocabulary teaching. *MLJ, 63* (5/6), 262–272.

Cartwright, D.S., & Nickerson, C.A. (1979). An empirical thesaurus: Meaning norms for ninety common words. *MLJ*, *63* (8), 442–447.

Sugano, M.Z. (1981). The idiom in Spanish language teaching. *MLJ, 65* (1), 59–66.

Banta, F.G. (1981). Teaching German vocabulary: The use of English cognates and common loan words. *MLJ, 65* (2), 129–136.

Vari-Cartier, P. (1981). Development and validation of a new instrument to assess the readability of Spanish prose. *MLJ, 65* (2), 141–148.

Odlin, T., & Natalicio, D. (1982). Some characteristics of word classification in a second language. *MLJ, 66* (1), 34–38.

Adams, S.J. (1982). Scripts and the recognition of unfamiliar vocabulary: Enhancing second language reading skills. *MLJ, 66* (2), 155–159.

Schulz, R.A. (1983). From word to meaning: Foreign language reading instruction after the elementary course. *MLJ, 67* (2), 127–134.

Bahrick, H.P. (1984). Fifty years of second language attrition: Implications for programmatic research. *MLJ, 68* (2), 105–118.

Martin, M. (1984). Advanced vocabulary teaching: The problem of synonyms. *MLJ, 68* (2), 130–137.

Ostyn, P., & Godin, P. (1985). RALEX: An alternative approach to language teaching. *MLJ, 69* (4), 346–355.

Crow, J.T. (1986). Receptive vocabulary acquisition for reading comprehension. *MLJ, 70* (3), 242–250.

Murphy, D.T. (1986). Approaches to the verb and Russian word counts. *MLJ, 70* (3), 258–262.

Barnett, M.A. (1986). Syntactic and lexical/semantic skill in foreign language reading: Importance and interaction. *MLJ, 70* (4), 343–349.

Keller, H. (1987). Pedagogical wishes for a machine dictionary: An example from Russian. *MLJ, 71* (1), 12–17.

Hayes, E.B. (1988). Encoding strategies used by native and non-native readers of Chinese Mandarin. *MLJ, 72* (2), 188–195.

Bolton, S. (1988). A comparative study of basic German vocabulary lists. *MLJ, 72* (2), 196–200.

Davis, J.N. (1989). Facilitating effects of marginal glosses on foreign language reading. *MLJ, 73* (1), 41–48.

Koda, K. (1989). Development of L2 intraword orthographic sensitivity and decoding skills. *MLJ, 73* (1), 51–64.

Kern, R.G. (1989). Second language reading strategy instruction: Its effects on comprehension and word inference ability. *MLJ, 73* (2), 135–149.

Krashen, S.D. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *MLJ, 73* (4), 440–464.

Bland, S.K., Noblitt, J.S., Armington, S., & Gay, G. (1990). The naive lexical hypothesis: Evidence from computer-assisted language learning. *MLJ, 74* (4), 44–450.

Laufer, B. (1991). The development of L2 lexis in the expression of the advanced learner. *MLJ*, *75* (4), 440–448.

Hulstijn, J.H. (1993). When do foreign-language readers look up the meaning of unfamiliar words? The influence of task and learner variables. *MLJ, 77* (2), 139–147.

Knight, S. (1994). Dictionary use while reading: The effects on comprehension and vocabulary acquisition for students of different verbal abilities. *MLJ, 78* (3), 285–299.

Sanaoui, R. (1995). Adult learners' approaches to learning vocabulary in second languages. *MLJ, 79* (1), 15–28.

Chun, D.L., & Plass, J.L. (1996). Effects of multimedia annotations on vocabulary acquisition. *MLJ, 80* (2), 183–198.

Hulstijn, J.H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *MLJ, 80* (3), 327–339.

Ke, C.R. (1996). An empirical study on the relationship between Chinese character recognition and production. *MLJ, 80* (3), 340–349.

Koda, K. (1996). L2 word recognition research: A critical review. *MLJ, 80* (4), 450–460.

Prince, P. (1996). Second language vocabulary learning: The role of context versus translations as a function of proficiency. *MLJ, 80* (4), 478–493.

Laufer, B., & Hadar, L. (1997). Assessing the effectiveness of monolingual, bilingual, and "bilingualised" dictionaries in the comprehension and production of new words. *MLJ, 81* (2), 189–196.

Mori, Y. (1998). Effects of first language and phonological accessibility on kanji recognition. *MLJ, 82* (1), 69–82.

Grace, C. (1998). Retention of word meanings inferred from context and sentence-level translations: Implications for the design of beginning-level CALL software. *MLJ, 82* (4), 533–544.

Kojic-Sabo, I., & Lightbown, P.M. (1999). Students' approaches to vocabulary learning and their relationship to success. *MLJ, 83* (2), 176–192.

Mori, Y. (1999). Beliefs about language learning and their relationship to the ability to integrate information from word parts and context in interpreting novel kanji words. *MLJ, 83* (4), 534–554.

Wesche, M., & Paribakht, T.S. (2000). Reading-based exercises in second language vocabulary learning: An introspective study. *MLJ, 84* (2), 196–213.

Grace, C.A. (2000). Gender differences: Vocabulary retention and access to translations for beginning language learners in CALL. *MLJ, 84* (2), 214–224.

Khaldieh, S.A. (2001). The relationship between knowledge of I'raab, lexical knowledge, and reading comprehension of nonnative readers of Arabic. *MLJ, 85* (3), 416–431.

Greidanus, T., & Nienhuis, L. (2001). Testing the quality of word knowledge in a second language by means of word associations: Types of distractors and types of associations. *MLJ, 85* (4), 567–577.

Jones, L.C., & Plass, J.L. (2002). Supporting listening comprehension and vocabulary acquisition in French with multimedia annotations. *MLJ, 86* (4), 546–561.

Barcroft, J. (2003). Effects of questions about word meaning during L2 Spanish lexical learning. *MLJ, 87* (4), 546–561.

Frantzen, D. (2003). Factors affecting how second language Spanish students derive meaning from context. *MLJ, 87* (2), 168–199.

Fan, M.Y. (2003). Frequency of use, perceived usefulness, and actual usefulness of second language vocabulary strategies: A study of Hong Kong learners. *MLJ, 87* (2), 222–241.

Morin, R. (2003). Derivational morphological analysis as a strategy for vocabulary acquisition in Spanish. *MLJ, 87* (2), 200–221.

Mori, Y. (2003). The roles of context and word morphology in learning new kanji words. *MLJ, 87* (3), 404–420.

Jiang, N. (2004). Semantic transfer and its implications for vocabulary teaching in a second language. *MLJ, 88* (3), 416–432.

Fukkink, R.G., Hulstijn, J.H., & Simis, A. (2005). Does training in second-language word recognition skills affect reading comprehension? An experimental study. *MLJ, 89* (1), 54–75.

Greidanus, T., Beks, B., & Wakely, R. (2005). Testing the development of French word knowledge by advanced Dutch- and English-speaking learners and native speakers. *MLJ, 89* (2), 221–233.

Chikamatsu, N. (2006). Developmental word recognition: A study of L1 English readers of L2 Japanese. *MLJ, 90* (1), 67–85.

Sagarra, N., & Alba, M. (2006). The key is in the keyword: L2 vocabulary learning methods with beginning learners of Spanish. *MLJ, 90* (2), 228–243.

Nassaji, H. (2006). The relationship between depth of vocabulary knowledge and L2 learners' lexical inferencing strategy use and success. *MLJ, 90* (3), 387–401.

Trofimovich, P., & Gatbonton, E. (2006). Repetition and focus on form in processing L2 Spanish words: Implications for pronunciation instruction. *MLJ, 90* (4), 519–535.

Shen, H.H., & Ke, C.R. (2007). Radical awareness and word acquisition among nonnative learners of Chinese. *MLJ, 91* (1), 97–111.

Pavlenko, A., & Driagina, V. (2007). Russian emotion vocabulary in American learners' narratives. *MLJ, 91* (2), 213–234.

Chieh-Fang, H. (2008). Rate of acquiring and processing L2 color words in relation to L1 phonological awareness. *MLJ, 92* (1), 39–52.

Kim, Y.J. (2008). The contribution of collaborative and individual tasks to the acquisition of L2 vocabulary. *MLJ, 92* (1), 114–130.

Hansen, G.F. (2010). Word recognition in Arabic. *MLJ, 94* (4), 567–581.

Vocabulary Learning and Instruction
Volume 3, Issue 1, December 2014
http://vli-journal.org

Article

Do Japanese Students Overestimate or Underestimate Their Knowledge of English Loanwords More than Non-loanwords on Yes–No Vocabulary Tests?

Raymond Stubbe
Kyushu Sangyo University
doi: http://dx.doi.org/10.7820/vli.v03.1.stubbe

Abstract

English loanwords (LWs), *gairaigo* in Japanese, make up a much greater percentage of the Japanese language than many university English teachers realize, especially if their native language is not Japanese. Unfortunately, a *gairaigo bias* exists which has made these LWs unpopular amongst teachers and researchers. The aim of this study is to compare student over-estimation and under-estimation of their knowledge of English LWs on yes–no vocabulary tests with an equal number of non-loanwords (NLWs). Undergraduate students from four Japanese universities ($n = 455$) took two vocabulary tests of their receptive and passive recall knowledge of LWs and NLWs. Six LWs and six NLWs from each of the eight JACET 8000 levels were tested in a self-report yes–no test followed by a passive recall translation test (English to Japanese) of the same 96 items. Overall, over-estimation rates were nearly equal at 24.6% for LWs and 25.8% for NLWs. Additionally, over-estimation was more prevalent for NLWs at the higher three frequency levels (1K–3K), nearly equal with LWs at the 4K level and then more prevalent for the LWs at the lower four frequency levels (5K–8K), suggesting that student knowledge of NLWs is weak even at the higher frequency levels. Under-estimation, on the other hand, was much more prevalent for LWs (4.4% versus 0.7%). Six of the 48 LWs actually had higher passive recall test scores than yes–no test scores. These results suggest that although students do not over-estimate their knowledge of LWs more than NLWs on yes–no vocabulary tests, they do under-estimate their LW knowledge much more than NLWs.

Keywords: over-estimation of lexical knowledge; loanwords; yes-no vocabulary tests; passive recall knowledge; JACET 8000.

1 Background

1.1 English LWs in Japanese

The Japanese language contains thousands of English loanwords (LWs), "many of which are well-established and in universal use" (Kay, 1995). It has been estimated that about half of the most common 3,000 words of English have some borrowed form in Japanese (Daulton, 1998). Of a random selection of words

Please cite this article as: Stubbe, R. (2014). Do Japanese students overestimate or underestimate their knowledge of English loanwords more than non-loanwords on yes–no vocabulary tests? *Vocabulary Learning and Instruction*, *3* (1), 29–43. doi: 10.7820/vli.v03.1.stubbe

contained in the reading sections (parts four and five) of two official TOEIC Bridge Practice Tests (Ashmore et al., 2007), 53% were English LWs in Japanese (Stubbe, 2010). In that study, LW recognition was found to be significantly better than NLW recognition, especially among lower-level students.

A distinction between LWs which appear in Japanese and the term *cognates* will be made for this study. Whereas LWs are words borrowed into an L1 from another language (Daulton, 2008), cognates can be considered as lexical items shared between two languages, often occurring when "there is a strong lexical resemblance between the target language and the learner's mother tongue" (Eyckmans, 2004, p. 24). Naturally, no such resemblance exists between Japanese and English.

Although LWs appear to make up a sizable portion of students' English vocabulary, some of their university teachers may be unaware of this. A focus group consisting of 18 university English teachers (14 native English speakers and 4 Japanese teachers), most of whom had been teaching the same vocabulary lists to students for at least two years, considered English LWs in Japanese. The members were asked what percentage of the first year word list they thought were LWs. Responses ranged from 17% through 35%, plus one outlier of 60% (from a vocabulary testing researcher). The actual LW rate of 53% was far above the focus group's mean estimate of 29% (including the outlier), suggesting that many of these English teachers were not aware of the prevalence of English LWs in Japanese.

There also appears to be a prevalent bias against LWs, known as gairaigo in Japanese meaning "words from abroad." Daulton (2011) refers to this as a "gairaigo bias." Uchida (2007) and Daulton (2011) both explain that university students overwhelmingly reported that their junior or senior high-school English teachers either avoided references to gairaigo in the classroom or that if they did refer to it, it was spoken of negatively. In addition to secondary school teachers, some university English teachers from abroad also seem to suffer from this gairaigo bias. Shepherd (1996) suggested that gairaigo should be avoided in the classroom. Simon-Maeda (1995) states that LWs having meanings in Japanese which differ from those in English, "provides the most headaches for EFL teachers." More recently, Masson (2013) reported that her students overestimated their knowledge of word meaning for three LWs (10% of the LWs tested). These three LWs were labeled as "distant false friends" (see Uchida, 2001), which are LWs having L1 meanings differing from L2. According to Masson (2013, p. 10), these LWs "are particularly problematic for students who are lulled into a false sense of familiarity, but do not notice that the meaning has changed." Overestimation was not the only problem with LWs identified in that study, a second was usage. On five other LWs, all having meanings matching or close to their L2 equivalents, the students made collocational or word-placement errors. Five of 30 (17%) does not appear to be overly excessive. Unfortunately, NLWs were not included in that study to provide a base for comparison of the meaning and usage errors. This omission coupled with the fact that only 27% of the LWs (8 of 30) formed the basis for the entire article may suggest that Masson also suffers from gairaigo bias.

One study that did investigate differences in student knowledge of LWs versus NLWs is the pilot to the present study (Stubbe & Yokomitsu, 2012). In that study,

students ($n = 71$) were given two vocabulary tests: a yes–no (passive recognition) test of 60 LWs and 60 NLWs from all eight levels of the *JACET List of 8000 Basic Words (2003)* (hereinafter J8000); followed by a passive recall (L1 to L2 translation) test of the same 120 words. This pilot study found that on the yes–no test Japanese university students' ($n = 71$) LW knowledge was almost 70% greater than their knowledge of NLWs (LW $M = 77.5\%$, standard deviation [SD] = 30.5%; NLW $M = 45.9\%$, SD = 36.2%). That investigation also found that passive recall knowledge as measured by an English to Japanese (L2 to L1) translation test of the same LWs was on average three times greater than their passive recall knowledge of the same NLWs (LW $M = 40.0\%$, SD = 37.1%; NLW $M = 13.1\%$, SD = 25.4%). These greater yes–no scores suggest that the participants were simply over-estimating their lexical knowledge, especially on the NLWs. A review of the literature suggests different explanations for the gaps between yes–no test results and passive recall test scores.

1.2 Gap between Yes–No Test and Passive Recall Tests Results

The decrease in scores between the yes–no test and the passive recall test reported in the pilot study has been represented in the literature as students simply overestimating their lexical knowledge on the yes–no test (discussed below). In a study using a similar protocol to the pilot, Waring and Takaki (2003) reported a nearly 70% decrease in mean scores between a similar recognition checklist test (61.2%) and an L2 to L1 passive recall translation test (18.4%; same 25 items). In that study, 25 words in a graded reader were replaced with non-real words (for example, "sun" became "blund"), and students' ability to passively recognize and recall those 25 non-words was tested. It is possible that students who took the Stubbe and Yokomitsu yes–no test and the Waring and Takaki recognition checklist test were checking words which they thought they recognized and believed they knew a meaning of, but both studies' passive recall results showed that translations were often lacking or faulty. In other words, there appears to be a considerable gap between thinking one knows a word and actually being able to produce a correct translation for that word. In the pilot study, only 45.6% of the 120 items were attempted on the passive recall test, with 45.3% of these being incorrect (Stubbe & Yokomitsu, 2012). In their study, Waring and Takaki (2003) also conducted a third, multiple choice, test of the same 25 non-words and reported a mean of 42.4%. Discussing Waring and Takaki (2003), Nation and Webb (2011, p. 282) wrote:

> Thus only a small number of words were learned well (per the results of the translation test), but quite a large number were learned at least partially. If only the translation test had been given, the amount of vocabulary learning from the reading would have been greatly underestimated.

It is possible that the students involved in the pilot study were checking yes–no items which they had partial knowledge of, and this could account for a portion of the gap between those yes–no and passive recall test scores.

Milton (2009) offers another possible explanation for this *recognition–translation* gap. Words encountered repeatedly can become easy to recognize through implicit (incidental) learning. "Learning the meaning of new words, however, would require conscious processing at the semantic and conceptual levels (explicit learning)" (Milton, 2009, p. 219). Accordingly, recognizing items on a yes–no test should be much easier than providing correct L1 translations of those same items. Similarly, LWs, which are encountered in the L1 and thus can be learned implicitly, would be expected to have higher test scores than NLWs which are only encountered in the L2, and thus required explicit learning.

The "strength of vocabulary knowledge" hypothesis discussed in Laufer and Goldstein (2004) offers still another plausible explanation for the disparity between yes–no and translation test results. That study considered four types of vocabulary knowledge: *active recall*, where students can provide (recall) an L2 word following a prompt of the word in the L1; *passive recall*, where students can provide an L1 translation following an L2 prompt; *active recognition*, where students can select the correct L2 word out of a set of L2 distractors based on an L1 prompt; and *passive recognition*, where students can select the correct L1 translation out of a set of L1 distractors based on an L2 prompt. Laufer and Goldstein (2004) demonstrated that the order presented above (*active recall* through *passive recognition*) represented a hierarchy of learning difficulty in descending (hardest to easiest) order. Considering the recall versus recognition dichotomy, Laufer and Goldstein (2004, p. 408) explained that "recalling a word's meaning or form can be considered a more advanced degree of knowledge than recognizing it." Thus, scores on a passive yes–no recognition test would be expected to be higher than scores on an L2 to L1 translation (passive recall) test of the same words.

1.3 Over-estimation and Under-estimation of Lexical Knowledge on Yes–No Tests

Yes–no checklist vocabulary tests rely on students self-reporting their lexical knowledge. Unfortunately, self-reporting often results in over-estimation of lexical knowledge by test-takers claiming words they do not really know. Pseudowords were introduced to the yes–no test format by Anderson and Freebody (1983) as a means of checking for evidence of such over-estimation. Pseudowords were introduced to the field of second language acquisition by Meara and Buxton (1987). In such tests, knowledge of a real word is known as a "hit," while claiming knowledge of a pseudoword is a "false alarm." Not claiming knowledge of a real word is labeled a "miss" and not claiming knowledge of a pseudoword is a "correct rejection" (Anderson & Freebody, 1983). In this paper, the terms *hit* and *miss* will be used frequently.

A number of other studies have reported yes–no test over-estimation as well as under-estimation. The degree of both varies widely from study to study depending, in part, on the type of test used to confirm actual lexical knowledge on the yes–no test, as well as the English proficiency levels of the participants. Mochida and Harrington (2003) report over-estimation and under-estimation rates of 5.4% and

6.4%, respectively, for high-proficiency English learners, when yes–no test results were compared to a subsequently taken *Vocabulary Levels Test* (VLT; Nation, 1990; Schmidt, Schmidt & Clapham, 2001). Stubbe, Stewart and Pritchard (2010) studying low-proficiency English learners reported rates of 14.5% and 33.1% respectively, when yes–no test results were compared to a multiple choice test of the same items. Stubbe (2012), also comparing yes–no test results with a multiple choice test of the same items, reported over- and under-estimation rates of 4.5% and 21.23%, respectively. That study also reported significantly higher over-estimation rates for the lower-ability students. In a series of yes–no test experiments, Eyckmans (2004) reported yes–no test scores which were between 4.9% and 38.7% higher than subsequent L2 to L1 (Dutch to French) translation scores of the same items. Directly comparing each participant's score on each word tested in the yes–no tests with the same student/word score on the translation test in that study, Eyckmanns also reported over-estimation (yes–no hit matched with an incorrect translation) rates of between 30.6% and 48%, with under-estimation (yes–no miss matched with a correct translation) rates ranging from 12.0% through 25.4%. Perhaps surprisingly, over-estimation rates were higher when cognates were excluded from the test results. Eyckmans found "that the cognates in the test were not responsible for the participants' tendency to overestimate their word knowledge" (2004, p. 85). It was also found that under-estimation rates were lower for non-cognate test items, leading Eyckmans to speculate that the students did not trust the cognates on the yes–no test likely because of a warning about the inclusion of pseudowords in the test instruction. However, upon seeing them again on the translation test they attempted translations.

2 Aims

This study aims to determine whether LW status significantly affects over-estimation rates on a yes–no vocabulary test compared with a passive recall (L2 to L1) translation test of the same words.

3 Method

3.1 Item Selection and Test Preparation

For this study, an item pool was prepared consisting of 96 real-words (six LWs and six NLWs from each of the eight JACET 8000 (2003) levels) plus 32 pseudowords, for a total of 128 items. Included in this item pool were the nine best pseudowords as well as the top 40 real-words identified in Stubbe and Stewart (2012), plus four more NLWs from the pilot to this study (Stubbe & Yokomitsu, 2012). The nine pseudowords were chosen by examining "the degree to which pseudowords have *negative* point-biserial correlations with total scores on the translation tests." The 40 real-words were selected "by examining phi (dichotomous) correlations between students' self-reports on given words and whether or not the same word was confirmed as known on the subsequent translation test" (Stubbe & Stewart, 2012, p.5). The remaining 52 real-words were randomly selected as required from the eight levels of the *JACET 8000* (2003). In selecting LWs and NLWs, part of speech was not taken into account.

In selecting LWs no distinction was made concerning the type of LW encountered. Uchida (2001) discusses six types of Japanese–English cognates of varying degrees of difficulty, into which LWs can be classified (see Masson, 2013 for a concise overview). It was decided to ignore LW type, because determining LW type would be cumbersome for all 48 LWs; plus selecting some types of LWs while rejecting others would inevitably skew results in favor of either LWs or NLWs. All 48 randomly selected LWs were reviewed by a Japanese teacher of English to ensure that they all were in fact LWs. The same was done for the 48 randomly selected NLWs to make sure that none of them were LWs. Three LWs and five NLWs required replacement and these eight replacements were again checked by the same teacher to ensure they were of the proper LW status.

These 96 items (see Appendix) were used to create two vocabulary tests, the first being a passive recognition yes–no vocabulary test, which also included 32 pseudowords. These 128 real-words and pseudowords were all assigned random numbers (1–128) and ordered into the yes–no test accordingly. The second vocabulary test was a passive recall test of the same 96 items from English into Japanese. A passive recall test was chosen because translation ability is a strong indicator of which words students can actually understand while reading (Waring & Takaki, 2003) and "asking participants to provide mother-tongue equivalents of the target language words was the most univocal way of verifying recognition" (Eyckmans, 2004, p. 77). The 96 items in this test were also assigned random numbers (1–96) for ordering. Participants were given the yes–no test at the beginning of one class in July or August 2012 and received the passive recall test toward the end of that same class. Tests were given in the same class in order to avoid any *loss of yes–no – passive recall test pairings* due to student absences in a subsequent class. The former test was a paper yes–no test in which the students signaled whether they knew a word or not by filling in either a *yes* bubble or a *no* bubble beside each item. Similar to Eyckmans (2004), test-takers were advised of the existence of pseudowords in the yes–no test instructions. Yes–no test forms were subsequently marked by running them through an optical scanner and the resulting data were converted into an Excel file for analysis.

The passive recall test forms, on the other hand, were hand-marked by three markers: one Japanese English instructor and two third-year students. Inter-rater reliability among these three markers was .92, and *FACETS* analysis (Linacre, 2012) revealed that the raters were basically equal with overall measures of .02, .02 and −.04 logits. One weakness of decontextualized passive recall tests deserves attention here. As many of the 96 words used in this study have multiple meanings, any one of those was considered correct, including katakana transcriptions of the LWs as long as the usage in Japanese also exists in English. Naturally, this would make LWs far easier to translate, but also easier to learn, at least on the single meaning level. Finally, informed consent was obtained from all participants.

3.2 Participants

Undergraduate students from 23 classes in four Japanese universities ($n = 455$) covering a wide range of English proficiency levels participated in this study.

4 Results and Discussion

Similar to the pilot study, the yes–no test mean was nearly double that of the passive recall test (50.9% versus 28.2% of the 96 words). SDs were 17.9% and 12.7%, respectively. Test reliabilities (Cronbach's alpha) were high at .96 and .92, respectively (see Table 1).

Table 1. Descriptive Statistics for Yes–No and Passive Recall Tests (n = 455, k = 96)

Test	Mean (%)	SD (%)	Reliability
YN	50.9	17.9	.96
PR	28.2	12.7	.92

Note: YN and PR refer to yes–no and passive recall tests, respectively, for all Tables.

The second table breaks down yes–no and passive recall test results by LW status. On the yes–no test, LWs had 49.2% more reports than NLWs (61.8% and 39.9%, respectively). On the passive recall test, however, LWs had 166.7% more correct translations than NLWs (41.6% and 14.9%, respectively).

Table 2. Descriptive Statistics for Yes–No and Passive Recall Tests by Loanword Status (n = 455, k = 48)

Test	Mean (%)	SD (%)	Range (%)	Low–High (%)
YN LWs	61.8	9.8	89.6	6.25–95.8
YN NLWs	39.9	9.0	85.4	0–85.4
PR LWs	41.6	8.1	75.0	6.25–81.3
PR NLWs	14.9	5.4	54.2	0–54.2

For purposes of comparison, Table 3 presents the LW means on both tests as well as the NLW means in separate columns. The difference in means was smaller for the LWs (20.2% versus 25.0% for NLWs). The yes–no LW mean was 0.49 times greater than on the passive recall test, while the yes–no NLW mean was 1.73 times that of the passive recall test. It appears as if these participants did not over-estimate their LW knowledge more than their NLW knowledge.

Table 3. LWs versus NLWs Means by Both Tests (YN and PR)

Test	LW mean (%)	NLW mean (%)
YN	61.8	39.9
PR	41.6	14.9
Diff: YN–PR	20.2	25.0

Table 4 displays the LW and NLW means for the yes–no and passive recall tests in each of the eight J8000 levels. With the exception of the 3K level on the yes–no test, LW scores exceeded those of NLWs. LW scores at the 7K and 8K levels were higher than NLW scores at the 2K and 3K levels (24.5% and 38.3% versus 34.5% and 19.3%). It can also be seen that decreases between yes–no scores and passive recall scores were much greater for the NLWs at the 1K through 4K levels. As these higher frequency words are far more critical for students to master (Milton, 2009), NLW over-estimation is arguably a far more serious problem than that of LWs. However, if any under-estimation of knowledge also existed in the data, as with the cognates in Eyckmans (2004), over-estimation rates would be artificially reduced. To calculate accurate over-estimation and under-estimation rates, direct comparison of the two test forms was required (see Table 5).

Table 4. LW and NLW Means for Yes–No and Passive Recall Tests by J8000 Level ($k = 6$)

J8000 level	YN LW (%)	PR LW (%)	YN NLW (%)	PR NLW (%)
1K	88.2	75.0	82.5	40.5
2K	83.3	63.3	71.7	34.5
3K	51.8	28.5	54.2	19.3
4K	64.2	46.3	29.8	6.7
5K	48.8	31.7	17.8	2.5
6K	42.7	24.5	24.3	7.7
7K	49.0	24.5	18.0	2.2
8K	66.3	38.3	21.0	5.2
Overall	61.8	41.5	40.0	14.8

Table 5. Direct Comparison of Yes–No Results and Passive Recall Test Scores, by Participant and Item (LW Result/NLW Result)

		Passive recall test	
		correct	incorrect
Yes–No test	hit	Confirmed correct 37.2% / *14.2%*	Over-estimation 24.6% / *25.8%*
	miss	Under-estimation 4.4% / *0.7%*	Confirmed unknown 33.9% / *59.4%*

Note: Incorrect includes blanks; percentages are of total incidents (21,840 = 455 students × 48 LWs or 48 NLWs), LW matches precede *NLW* matches, which are italicized.

4.1 Direct Comparison of Yes–No Test and Passive Recall Test Results

Similar to Stubbe (2012), Mochida and Harrington (2006) and Eyckmans (2004), this final analysis matched each student's response for each word on the yes–no test ($n = 455$, $k = 48$ LWs and 48 NLWs) with his/her answer to that word

on the passive recall test (answers left blank on the passive recall test were combined with incorrect responses for this analysis). Table 5 presents the results of this analysis, which resulted in four possible matches: (a) a yes–no *hit* matched with a passive recall test *correct* answer which was labeled "confirmed known"; (b) a yes–no *hit* matched with a passive recall test *incorrect*, labeled "over-estimation"; (c) a yes–no *miss* and a passive recall test *correct* was labeled "under-estimation"; and (d) yes–no *miss* and passive recall test *incorrect* was labeled "confirmed unknown." In this table, LW matches precede NLW matches, with the later also *italicized*. For example, 37.2% / *14.2%* are the percentages of confirmed known matches for LWs and NLWs, respectively. As expected, confirmed known heavily favored the LWs, with the opposite for confirmed unknown. Together the confirmed known and confirmed unknown categories account for 71.08% and 73.58% of possible matches for LWs and NLWs, respectively. Perhaps not surprisingly, the NLWs were slightly more reliable. The remaining responses, over-estimation, and under-estimation are discussed below.

As mentioned, a hit matched with a non-correct translation (including blanks) was labeled over-estimation. Perhaps contrary to expectations, the LWs' percentage is slightly lower than the NLWs' (24.6% versus 25.8%). It appears as if LW status (similar to cognate status) did not significantly effect over-estimation rates on the yes–no vocabulary test. If blanks are excluded from the analysis leaving actual translation attempts, the LWs had an error rate of 19.3% while the NLWs had a rate of 20.4%.

It may be remembered that Eyckmans (2004) also found that under-estimation occurred much more often with cognates than with non-cognates. In the final quadrant of Table 5, under-estimation, the LW results replicated this finding by exceeding the amount of the NLWs' under-estimation by a considerable margin (4.4% versus 0.7%). It may also be possible that the participants in this study did not trust some of the LWs on this yes–no test because of a warning about pseudowords in the test instructions, similar to Eyckmans (2004). To illustrate, the 4K LW *helicopter* had the greatest amount of "under-estimation" at 21.8% of all students. Five other LWs also had higher passive recall test scores than yes–no test scores, while none of the NLWs did (see Appendix).

Table 6 breaks down the over-estimation quadrant in Table 5 by J8000 level. As mentioned above, only in this quadrant was the difference between LW and NLW results slight, suggesting that students may be able to *recognize* without being able to *recall the meaning* of LWs and NLWs in near equal measure. However, Table 6 suggests that over-estimation was more prevalent for NLWs at the higher three frequency levels, nearly equal with LWs at the 4K level and then more prevalent for

Table 6. Over-estimation of LW and NLW Knowledge on the Yes–No Test Broken Down by J8000 Level (n = 455)

	J8000 level							
Category	1K (%)	2K (%)	3K (%)	4K (%)	5K (%)	6K (%)	7K (%)	8K (%)
LWs	14.6	23.6	28.2	26.4	22.5	21.9	27.0	32.4
NLWs	42.5	37.8	36.2	23.7	15.8	17.2	16.3	16.7

the LWs at the lower four frequency levels. It appears that over-estimation or *recognition* without *meaning recall* was strongest for NLWs at the highest frequency level (ranging from 20.9% to 59.3% of students overestimating the 1K NLWs, see Appendix), decreasing steadily through 5K, before basically leveling off. For the LWs, on the other hand, over-estimation started low at 1K, increased steadily through 3K, decreased from 4K through 6K, before increasing to its highest value at 8K. Viewing lexical development as a process of moving from passive recognition to passive recall of meaning, we can see that for NLWs the gap between these two stages is greatest at the highest frequency level suggesting that these students still need to develop their *passive recall of meaning* ability for many of the most frequent English NLWs. As knowledge of NLWs decreased following word frequency level expectations, over-estimation decreased through 5K for NLWs, then basically leveled off thereafter. For LWs, however, it appears as if the opposite occurred. With the highest frequency items, over-estimation was the least, while increasing (albeit unevenly) through to its highest level at 8K.

5 Conclusion

This study has investigated whether LW status significantly affects over-estimation rates on a yes–no vocabulary test. Student recognition and translation abilities of 48 LWs and 48 NLWs, randomly selected from all levels of the JACET8000 list, were tested. Overall, it was found that the NLWs had slightly more over-estimation than the LWs. A breakdown of actual instances of over-estimation by J8000 frequency level revealed that, at the highest frequency levels, students over-estimated their knowledge of the NLWs considerably more than the LWs. As high frequency vocabulary is used much more often than low frequency, especially by L2 learners, it seems concerns regarding over-estimation might better be focused on NLWs. It was also found that lower frequency 7K and 8K LWs were better known than NLWs at the 2K and 3K levels. Additionally, students made more actual translation errors on the NLWs. The LWs, however, did attract considerably more under-estimation on the yes–no test.

This study does suffer from a number of limitations. The 92% inter-rater reliability amongst the three markers could be higher. Also, the selection of only six LWs and six NLWs from each J8000 could be considered too small to capture a truly representative sampling and thus allow for the skewing of the results. For instance, the LW jump at the 4K level was possibly due to the influence of the single word *helicopter*. Sampling a greater number of words from fewer J8000 levels could help alleviate this weakness. Additionally, a limitation of single word (decontextualized) passive recall tests must be acknowledged. As many of the 96 words used in this study have multiple meanings, any one of those was considered correct, including katakana transcriptions of the LWs as long as the usage in Japanese also exists in English. Naturally, this would make LWs far easier to translate, but also easier to learn, at least on a single meaning level. Finally, it has been suggested in the literature that nouns are the easiest part of speech to learn (see Hirsh-Pasek & Golinkoff, 2006 for a good discussion). As LWs are usually nouns, this could further contribute to their ease of learning. Not controlling the balance of *parts of*

speech between the LWs and NLWs may have contributed to the dominance of LWs in these results.

Despite these weaknesses, these results may have implications for vocabulary testers and teachers. For testers of vocabulary, knowing which items are LWs while developing a test should help to better predict item performance. As LW scores at the 7K and 8K levels were higher than NLW scores at the 2K and 3K levels, not knowing which items in a test are LWs and building a test for Japanese students based only on word frequency levels could lead to some unexpected results. The lower level students could do relatively well compared to the high-level students based solely on their LW knowledge. It was found that although over-estimation rates are comparable to NLWs, students may also under-estimate their knowledge of some of the LWs when completing *yes–no* checklist tests.

For teachers, knowing the LWs in a list of vocabulary to be taught or used in a classroom could help them to better assist students in their lexical development. The LWs in the list sharing meanings and usages in both languages could be quickly reviewed before teaching the NLWs, which students tended to over-estimate their knowledge of at the highest frequency levels. Any LWs having meanings in Japanese which differ from English, called false friends (Uchida, 2001; recruit, for example) would require re-learning and this challenge could be anticipated and prepared for. Not preparing for such false friends could lead to frustrating misunderstandings between students and instructor, especially if the latter is a native English speaker unfamiliar with the pervasiveness of English LWs in Japanese. Instead of complaining about the headache of English LWs in Japanese, teachers (and researchers) should improve their LW knowledge so that they can better assist their students acquire the English vocabulary they require.

Acknowledgements

I thank Hoke Shintarou sensei, as well as the Kyushu Sangyo University *Language Education and Research (LERC)* staff for their assistance with marking the passive recall test. I would also like to thanks Messrs. Chris O'Sullivan, Dave Latz and Eric Miller, and their students for participating in this study. Finally, I also thank the many reviewers of this paper for their constructive comments.

References

Anderson, R. Cr., & Freebody, P. (1983). Reading comprehension and the assessment and acquisition of word knowledge. In B. A. Hutson (Ed.), *Advances in reading/language research* (vol. 2; pp. 231–256). Greenwich, CT: JAI Press.

Ashmore, E., Carter, E., Duke, T., Hauck, M., Locke, M., & Shearin, R. (2007). TOEIC® Bridge 公式ガイ ド&問題集 (Official Guide and Exercise Books). Princeton, NJ: Educational Testing Service.

Daulton, F. E. (2011). On the origins of gairaigo bias: English learners' attitudes towards English-based loanwords in Japan. *The Language Teacher, 35*(6), 7–12.

Daulton, F. E. (2008). *Japan's built-in lexicon of English-based loanwords.* New York, NY: Multilingual Matters.

Daulton, F. E. (1998). Loanword cognates and the acquisition of English vocabulary. *The Language Teacher, 22*(1), 17–25.

Eyckmans, J. (2004). *Measuring receptive vocabulary size*. Utrecht: LOT (Landelijke Onderzoekschool Taalwetenschap).

Hirsh-Pasek, K., & Golinkoff, R. M. (Eds.). (2006). *Action meets words: How children learn verbs*. New York, NY: Oxford University Press.

JACET Basic Word Revision Committee. (2003). *JACET list of 8000 basic words*. Tokyo: Japan Association of Language Teachers.

Kay, G. (1995). English loanwords in Japanese. *World Englishes, 14*(1), 67–76. doi:10.1111/j.1467-971X.1995.tb00340.x

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength and computer adaptiveness. *Language Learning, 54*(3), 399–436. doi:10.1111/j.0023-8333.2004.00260.x

Linacre, J. M. (2012). *Facets computer program for many-facet Rasch measurement, version 3.70.0*. Beaverton, OR: Winsteps.com. Retrieved from http://www.winsteps.com//index.htm

Masson, M.-E. (2013). How L1 loanwords can create a false sense of familiarity with L2 vocabulary meaning and usage. *Vocabulary Learning and Instruction, 2*(1), 8–14. doi:10.7820/vli.v02.1.masson

Meara, P., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language Testing, 4*(2), 142–154. doi:10.1177/026553228700400202

Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters.

Mochida, A., & Harrington, M. (2006). Yes/No test as a measure of receptive vocabulary. *Language Testing, 23*, 73–98. doi:10.1191/0265532206lt321oa

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston, MA: Heinle and Heinle.

Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston, MA: Heinle.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing, 18*, 55–89.

Shepherd, J. (1996). Loanwords: A pitfall for all students. *The Internet TESL Journal, 2*(2). Retrieved from: http://iteslj.org/Articles/Shepherd-Loanwords.html

Simon-Maeda, A. (1995). Language awareness: Use/misuse of loan-words in the English language in Japan. *The Internet TESL Journal, 1*(2). Retrieved from: http://iteslj.org/Articles/Maeda-Loanwords.html

Stubbe, R. (2012). Do pseudoword false alarm rates and overestimation rates in yes/no vocabulary tests change with Japanese university students' English ability levels? *Language Testing, 29*(4), 471–488. doi:10.1177/0265532211433033

Stubbe, R., & Stewart, J. (2012). Optimizing scoring formulas for yes/no vocabulary checklists using linear models. *Shiken Research Bulletin, 16*(2), 27. Retrieved from http://teval.jalt.org/node/12

Stubbe, R., & Yokomitsu, H. (2012). English loanwords in Japanese and the JACET 8000. *Vocabulary Education and Research Bulletin, 1*, 10–11. Retrieved from http://jaltvocab.weebly.com/publications.html

Stubbe, R. (2010). Exploring the lexical challenge of the TOEIC® *Bridge*. In A. M. Stoke (Ed.), *JALT 2009 conference proceedings* (pp. 710–721). Tokyo: JALT. Retrieved from http://jalt-publications.org/proceedings/articles/84-jalt2009-proceedings-contents

Stubbe, R., Stewart, J., & Pritchard, T. (2010). Examining the effects of pseudowords in yes/no vocabulary tests for low level learners. *Kyushu Sangyo University Language Education and Research Center Journal, 5*, 5–23.

Uchida, E. (2007). Oral and written identification of L2 loanword cognates by Japanese learners of English. *The Language Teacher, 31*(9), 19–22.

Uchida, E. (2001). *The use of cognate inferencing strategies by Japanese learners of English* (Unpublished doctoral dissertation). Essex University, Colchester, UK.

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language, 15*, 130–163.

Appendix
Yes–No Test (YN) Scores, Passive Recall Test (PR) Scores and the Difference (YN–PR) by J8000 Level, in Percentages.

LWs					NLWs				
Item	J8000	YN (%)	PR (%)	YN–PR (%)	Item	J8000	YN (%)	PR (%)	YN–PR (%)
Corner	1	88.8	83.3	5.5	Difficulty	1	88.6	29.2	59.3
Cup	1	98.5	91.2	7.3	Effort	1	78.7	45.7	33.0
Nation	1	81.5	48.1	33.4	Instead	1	75.4	38.0	37.4
Park	1	99.3	91.4	7.9	Join	1	93.8	73.0	20.9
Scene	1	66.8	54.3	12.5	Recent	1	74.3	22.4	51.9
West	1	94.1	82.0	12.1	Suffer	1	83.7	34.9	48.8
Connection	2	73.2	58.9	14.3	Elect	2	73.4	32.5	40.9
Crash	2	93.8	50.8	43.1	Enable	2	71.2	14.3	56.9
Dish	2	87.3	73.4	13.8	Novel	2	91.6	70.3	21.3
Drama	**2**	**82.9**	**84.8**	**−2.0**	Plenty	2	51.9	18.2	33.6
Gap	2	77.1	51.0	26.2	Previous	2	62.2	21.8	40.4
Swing	2	85.5	61.3	24.2	Sand	2	79.3	49.9	29.5
Chart	3	66.6	32.7	33.8	Honey	3	80.2	51.0	29.2
Container	3	46.2	26.4	19.8	Interfere	3	45.7	8.4	37.4
Crawl	3	26.4	14.1	12.3	Psychological	3	65.5	13.2	52.3
Nut	3	36.7	22.6	14.1	Similarity	3	51.2	19.1	32.1
Reverse	3	75.8	26.6	49.2	Suck	3	27.9	3.7	24.2
Trunk	3	59.6	48.6	11.0	Superior	3	54.7	20.4	34.3
Bargain	4	52.3	2.4	49.9	Availability	4	48.8	4.4	44.4
Chapel	4	52.1	44.0	8.1	Casualty	4	19.1	0.2	18.9
Crystal	**4**	**60.2**	**62.4**	**−2.2**	Exhibit	4	60.0	25.7	34.3
Disk	4	73.6	64.6	9.0	Fraud	4	5.3	0.9	4.4
Helicopter	**4**	**59.3**	**81.1**	**−21.8**	Incentive	4	28.6	4.6	24.0
Mild	4	87.5	23.5	64.0	Tactic	4	17.4	4.4	13.0

Appendix (*Continued*)

LWs					NLWs				
Item	J8000	YN (%)	PR (%)	YN–PR (%)	Item	J8000	YN (%)	PR (%)	YN–PR (%)
Canvas	5	49.0	37.8	11.2	Accusation	5	10.1	0.7	9.5
Cassette	5	25.1	23.5	1.5	Esthetic	5	5.3	0.2	5.1
Fever	5	76.3	46.6	29.7	Ditch	5	6.8	0.2	6.6
Handling	5	56.0	5.7	50.3	Methodology	5	18.2	3.3	14.9
Layout	5	44.0	33.6	10.3	Multiply	5	38.7	8.1	30.5
Salmon	**5**	**42.4**	**43.1**	**−0.7**	Rib	5	28.1	2.4	25.7
Accessory	6	44.6	36.0	8.6	Elementary	6	85.1	30.8	54.3
Cholesterol	**6**	**40.9**	**41.1**	**−0.2**	Idiot	6	23.1	7.3	15.8
Demonstrator	6	66.4	8.6	57.8	Siege	6	2.2	0.2	2.0
Jealousy	6	38.5	30.3	8.1	Sober	6	9.0	1.8	7.3
Pearl	6	20.2	9.7	10.5	Sting	6	23.3	5.9	17.4
Sour	6	45.5	21.8	23.7	Urine	6	3.7	0.4	3.3
Cork	7	28.1	4.0	24.2	Captive	7	5.9	0.4	5.5
Loaf	7	17.4	0.4	16.9	Coronary	7	9.7	0.2	9.5
Merry	7	56.7	0.9	55.8	Cripple	7	4.2	0.0	4.2
Ranking	7	88.1	78.9	9.2	Exacerbate	7	2.9	0.0	2.9
Typewriter	7	53.0	45.3	7.7	Expressive	7	52.7	3.3	49.5
Wireless	7	50.3	17.4	33.0	Unaffected	7	33.0	9.0	24.0
Biotechnology	8	84.6	27.9	56.7	Entrepreneurial	8	2.9	0.2	2.6
Delicacy	8	74.1	53.2	20.9	Grocery	8	42.0	11.2	30.8
Downtown	8	83.5	39.8	43.7	Relentless	8	3.5	0.0	3.5
Fried	8	48.1	24.0	24.2	Squirrel	8	9.9	2.2	7.7
Overcoat	8	33.2	5.3	27.9	Unresolved	8	31.2	16.5	14.7
Pianist	**8**	**74.1**	**80.2**	**−6.2**	Whine	8	36.5	0.4	36.0

Note: Bold denotes passive recall test score exceeding yes–no score.

Vocabulary Learning and Instruction
Volume 3, Issue 1, December 2014
http://vli-journal.org

Article

Is the Vocabulary Level of the Reading Section of the TOEFL Internet-Based Test Beyond the Lexical Level of Japanese Senior High School Students?

Masaya Kaneko
Tokyo Denki University
doi: http://dx.doi.org/10.7820/vli.v03.1.kaneko

Abstract

The main goal of the present study is to answer the question of whether the lexical level of the reading section of the TOEFL Internet-based Test (TOEFL iBT) is beyond the vocabulary level of Japanese senior high school graduates. The lexical level was measured in terms of text coverage. The present study builds upon Chujo and Oghigian's study. The notable difference in methodology compared to earlier text coverage studies on the TOEFL is an examination of real past TOEFL iBTs. Two objectives are explored in the present study. First, this study aims to examine how well a vocabulary of 3,000 word families, which is the lexical size target for Japanese high school graduates set by the Ministry of Education, Culture, Sports, Science, and Technology, enables students to comprehend reading passages in the TOEFL iBT. Second, it estimates a vocabulary size required to reach 95% and 98% coverage of these passages. Results showed that the most frequent 3,000 word families plus proper nouns as well as words that are defined in context yielded an average text coverage of 88.5% and that 6,000 word families plus proper nouns and defined words accounted for 95% of the text, and around 10,000 word families 98%. The findings suggest that Japanese high school graduates with a vocabulary of 3,000 word families would be expected to comprehend nearly 50% of reading passages in the TOEFL iBT and that learning a vocabulary beyond the 10,000-word frequency level may not be necessary unless 98% or more text coverage is required.

1 Background

On April 8, 2013 the education reform panel of the ruling Liberal Democratic Party of Japan suggested that the TOEFL Internet-based Test (TOEFL iBT) replace the National Center Test, which has been used for admission to all national, public, and 520 private universities in Japan (as of 2013). The proposal has provoked widespread criticism. One of the main objections to the adoption of the TOEFL iBT stems from arguments that a vocabulary beyond the 10,000-word frequency level is required to comprehend passages in the TOEFL and that the lexical level of the TOEFL therefore surpasses the level expected of Japanese senior high school graduates. In a newspaper article appearing in the Asahi Shimbun (Anonymous, 2013), Erikawa claims that the TOEFL requires a vocabulary beyond

Please cite this article as: Kaneko, M. (2014). Is the vocabulary level of the reading section of the TOEFL Internet-based Test beyond the lexical level of Japanese senior high school students? *Vocabulary Learning and Instruction*, *3* (1), 44–50. doi: 10.7820/vli.v03.1.kaneko

the 10,000-word frequency level. Sato (2013) maintains that a 15,000–30,000-word-frequency-level vocabulary is needed for the TOEFL.

Past L2-reading research investigating the effect of the percentage of known words in discourse on comprehension suggests that achieving 95% (Laufer, 1989) or 98% (Hu & Nation, 2000; Laufer & Ravenhorst-Kalovski, 2010; Schmitt, Jiang, & Grabe, 2011) coverage allows readers to gain adequate comprehension. Findings from several text coverage studies on the TOEFL partially support claims made by Erikawa and Sato.

Hirai (2000) examined six reading passages from an unofficial TOEFL preparation book. The vocabulary level was measured using the General Service List (GSL) (West, 1953) and the University Word List (UWL) (Xue & Nation, 1984). She found that the 2,800 word families on the two lists provided an average coverage of 86.6% and that proper nouns made up around 2% of the texts. Hirai concluded that learning low-frequency vocabulary is necessary to prepare for the TOEFL, because it constituted more than 10% of the text.

Mizumoto (2006) compiled a one-million-word TOEFL corpus consisting of various sources including both official TOEFL practice tests and unofficial test preparation materials. The lexical level was calculated using the Japan Association of College English Teachers (JACET) 8000 word list (Ishikawa et al., 2003). The JACET list is a lemma-based frequency list that contains 8,000 words plus 250 additional basic words. It was compiled using the British National Corpus (BNC) and many other materials that are specifically aimed at Japanese English as a Foreign Language (EFL) learners such as Japan-developed Eiken tests, National Center Tests, and English textbooks for junior and senior high school students approved by the Ministry of Education, Culture, Sports, Science, and Technology (MEXT). Thus, the list is tailored for the Japanese EFL learning environment. Mizumoto found that the most frequent 3,000 lemmas from the list made up 82.4% of the tokens in the TOEFL corpus and that its lexical level did not show a striking difference to the level of vocabulary in other genres of texts.

Chujo and Oghigian (2009) examined one official TOEFL iBT practice test, the version of the TOEFL administered in Japan since 2006. They concluded that with a vocabulary of the most frequent 4,000 word families on Nation's BNC 14,000 word lists (2006), an examinee would gain 95% text coverage of the TOEFL iBT.

Although informative, these earlier studies have several limitations. First, except for Chujo and Oghigian's (2009) study, the samples analyzed were all based on the older version of the TOEFL. However, different versions of the TOEFL may employ different lexical levels, as Chujo and Oghigian (2009) point out. These researchers compared text coverage figures obtained from the older and the current versions of the TOEFL. They found that to reach 95% text coverage, there was a difference of around 1,500 word families. They concluded that the TOEFL iBT may employ more high-frequency vocabulary than the older version although they acknowledged that their sample was quite small. Second, the frequency lists used in some of the earlier studies are limited in that they are too small to provide 95% (Laufer, 1989) or 98% text coverage (Hu & Nation, 2000; Laufer & Ravenhorst-Kalovski, 2010; Schmitt et al., 2011). The GSL (West, 1953) and the UWL (Xue &

Nation, 1984) are not sufficient for this purpose. Similarly, Mizumoto (2006) found that the JACET 8,000 list accounted for just 89.4% of his one-million-word TOEFL corpus. Third, several past studies of lexical levels of the TOEFL contained mixed sources. Test items for both receptive and productive knowledge were incorporated into one corpus (Chujo, 2004; Chujo & Oghigian, 2009; Mizumoto, 2006). The past TOEFL except for the original paper-based test consisted of three sections: reading, listening, and writing. The TOEFL iBT adds a speaking section to these existing sections. Past studies employed a text coverage approach to examine the vocabulary level of all three (Chujo, 2004; Mizumoto, 2006) or four sections (Chujo & Oghigian, 2009), which is a methodological flaw. If we want to determine the likelihood of success in the reading section of the TOEFL, only reading passages should be analyzed.

With these limitations in earlier text coverage studies on the TOEFL, earlier findings and their subsequent suggestions should be considered tentative.

2 Aims

The present study examines the vocabulary level of reading passages in the TOEFL iBT from the perspective of Japanese senior high school graduates. The notable difference in methodology compared to earlier text coverage studies on the TOEFL is an examination of real past TOEFL iBTs, which had not been available to the public before 2013. Earlier studies thus had to examine materials from either *practice* tests developed by the Educational Testing Service (ETS), the developer of the TOEFL test, or unofficial TOEFL preparation books. Therefore, the present study should provide a more accurate measurement of the vocabulary level of reading passages in the TOEFL iBT.

3 Sample

The present study examined 15 reading passages from five past TOEFL iBTs (Educational Testing Service, 2013). Each test contains three reading passages. The total number of tokens in the passages was 10,624, with an average of 708 per passage.

No modifications were made to the original passages. Chujo and Oghigian (2009) deleted all proper nouns, but the present study retained proper nouns because past studies on text coverage suggested that the deletion of proper nouns affects the results of text coverage studies (Brown, 2010; Kaneko, 2013). Kaneko compared coverage figures obtained from the reading passages in the 2004 Tokyo University entrance examination, with proper nouns deleted, and those with proper nouns left in. It was found that the text with proper nouns removed yielded around 1% higher coverage than the text with proper nouns.

4 Methods

The reading passages were examined using RANGE (Heatley, Nation, & Coxhead, 2002). The text coverage obtained through the most frequent 3,000 word

families from Nation's (2006) BNC word family lists and the vocabulary size required to achieve 95% and 98% coverage were estimated.

5 Results

Table 1 depicts cumulative text coverage figures on each reading section of the five past TOEFL iBTs. The BNC 3,000 word families plus proper nouns yielded an average of 87.8% (standard deviation, $SD = 2.8$) text coverage. To reach 95% text coverage, 6,000–7,000 word families were required. With regard to 98% text coverage, 12,000–13,000 word families were necessary.

If we consider words whose meaning is provided in the test, more accurate estimates will be achieved. This modification should be taken into consideration because definitions or pictures are provided for some words appearing in passages of the TOEFL iBT. For instance, a picture of a *chipmunk* is provided in Test 5. The definition of *feudalism* is given in Test 4. Readers do not need to know these words beforehand. Thus, those defined words should be considered to be known words. Table 2 shows revised cumulative text coverage figures taking into account those defined words. The most frequent 3,000 word families plus proper nouns and defined words yielded an average of 88.5% text coverage ($SD = 2.4$). An average of 6,000 word families were necessary to reach 95% text coverage ($SD = 0.8$). This coverage may allow readers to gain a comprehension of 60.5% according to the estimates made by Schmitt, Jiang, and Grabe (2011). Converting this comprehension figure into a score on the TOEFL iBT, students with knowledge of 6,000 word families would be estimated to achieve a score of 18 out of 30 points in the reading section. To reach 98% text coverage, around 10,000 word families were required.

Table 1. Cumulative Text Coverage Figures on the Reading Section from five Past TOEFL iBTs by Nation's BNC Lists

Frequency band	Test 1	Test 2	Test 3	Test 4	Test 5	Mean
Proper nouns	2	2.2	1.6	1.1	1.2	1.6
1,000	72.9	67.5	78.2	72.9	71.9	72.7
2,000	83.8	80	87.6	84.5	81.3	83.4
3,000	88.2	83.9	91.2	89.4	86	87.8
4,000	91.9	87.9	92.5	92.1	89.6	90.8
5,000	94.9	91.3	94	93.9	91.3	93.1
6,000	95.6	92.6	94.8	95.3	93.6	94.4
7,000	96.2	94	95.7	96.1	94.6	95.3
8,000	97.1	95	96.4	97.3	95.2	96.2
9,000	97.5	95.9	96.7	97.7	95.4	96.6
10,000	97.9	96.6	97.4	98.1	95.6	97.1
11,000	98.2	96.9	97.8	98.2	95.6	97.3
12,000	98.4	97.2	98.2	98.5	95.7	97.6
13,000	98.7	98.2	98.4	98.8	96	98
14,000	98.8	98.7	98.9	99.1	96.7	98.4
Not in the lists	99.9	100	100	99.9	99.9	99.9

Note. Totals of percentages are not 100% because of rounding.

Table 2. Cumulative Text Coverage Figures on the Reading Section from five Past TOEFL iBTs by Nation's BNC Lists, Defined Words Included

Frequency band	Test 1	Test 2	Test 3	Test 4	Test 5	Mean
DWs + PNs	2.3	3.2	1.8	1.5	2.9	2.3
1,000	73.1	68.5	78.4	73.2	73.7	73.4
2,000	84.1	81	87.8	84.8	83.1	84.2
3,000	88.5	84.9	91.4	89.7	87.7	88.5
4,000	92.2	88.9	92.8	92.4	91.3	91.5
5,000	95.2	92.3	94.2	94.1	93.1	93.8
6,000	95.8	93.6	95	95.5	95.4	95
7,000	96.4	95	95.9	96.4	96.3	96
8,000	97.3	96	96.6	97.6	96.9	96.9
9,000	97.7	96.9	96.9	97.9	97.1	97.3
10,000	98	97.6	97.6	98.2	97.4	97.8
11,000	98.4	97.9	97.9	98.2	97.4	97.9
12,000	98.6	98.2	98.3	98.5	97.5	98.2
13,000	98.8	98.9	98.6	98.8	97.7	98.6
14,000	98.9	99	99	99.2	98.3	98.9
Not in the lists	99.9	100	100	100	99.9	99.9

Note. Totals of percentages are not 100 because of rounding.
DWs = defined words; PNs = proper nouns.

Schmitt et al. (2011) suggest that this coverage figure would allow readers to gain 68.3% comprehension, or 20 out of 30 possible points in the reading section of the TOEFL.

6 Conclusions

To summarize, if we add proper nouns and words that are defined in the passages to the counts of the most frequent 3,000 word families, readers would be able to gain 88.5% text coverage of reading passages on the TOEFL iBT. Assuming that senior high school graduates acquire a vocabulary of 3,000 word families, which is the lexical size target set by the current Course of Study guidelines (MEXT, 2009), then students would be estimated to achieve nearly 50% comprehension in the TOEFL iBT according to the estimates made in a recent large-scale lexical and reading comprehension study (Schmitt et al., 2011). If students acquire 3,000 additional word families, they would be able to reach 95% text coverage (SD = 0.8). If students wish to enroll in an English-speaking university, 95–96% text coverage will be the first lexical-coverage target because this coverage figure may enable test takers to gain a comprehension of 60–63% as suggested by Schmitt et al. (2011). According to the official TOEFL iBT guide (Educational Testing Service, 2009), the average score for test takers who were applying for admission to English-medium colleges or universities as undergraduate students was 76 points (63%). For instance, Temple University Japan, which is the only four-year English-speaking American[1] university accredited by the MEXT, requires applicants to submit a score of 71 points (59%) or more to be matriculated for the undergraduate course.

In conclusion, the answer to the question of whether the lexical level of reading passages in the TOEFL iBT is beyond the vocabulary level of Japanese senior high school graduates depends on the desired comprehension level. If we assume that students would aspire to 45 points out of 120 in the TOEFL iBT, or 37.5%, as proposed by the head of the ruling Liberal Democratic Party's education reform panel (Anonymous, 2013), then 3,000 word families would be sufficient. In contrast, if students aim to enroll in an English-speaking university, the ideal lexical size target would be beyond the vocabulary size goal proposed by the MEXT, because at least 3,000 additional word families would be required.

Lastly, the findings of the present study should be confirmed with a different methodology. Lexical level of texts can be measured in two ways: by examining the text coverage that different frequency bands yield, or by examining students on text comprehension and relating that to their vocabulary size. The present study took the first approach. Aizawa and Iso (2008) explored the vocabulary level of reading passages in the TOEFL employing the latter method. However, the limited number of subjects and short length of the materials they sampled do not allow us to reach any solid conclusion. Measuring students' comprehension on reading passages from real past TOEFL iBTs and relating these scores to their lexical knowledge needs to be done to confirm the results of the present study.

However, two findings in the present study seem to be clear. First, the TOEFL proposal made by the education reform panel might not be as senseless as has been claimed because the 3,000 word-family goal for high school students would allow test takers to gain nearly 50% comprehension of the reading section in the TOEFL iBT. Second, learning a vocabulary beyond the 10,000-word frequency level would not be necessary unless 98% or more text coverage is desired. Hopefully, the findings of the present study can provide new insights into the discussion on the use of TOEFL as part of university entrance examinations.

Note

1. The word "American" was unfortunately omitted from the original version of this paper (published Online First on March 15, 2014, page 5).

References

Aizawa, K., & Iso, T. (2008). Identifying the minimum vocabulary size for academic reading. *Annual Review of English Language Education in Japan, 19*, 121–130.

Anonymous (2013, May 1). 大学入試にＴＯＥＦＬ [Adoption of the TOEFL as part of university entrance examinations]. *Asahi Shimbun*, p. 15.

Brown, D. (2010). An improper assumption? The treatment of proper nouns in text coverage counts. *Reading in a Foreign Language, 22*(2), 355–361.

Chujo, K. (2004). Measuring vocabulary levels of English textbooks and tests using a BNC lemmatized high frequency word list. In J. Nakamura, N. Inoue, & T. Tabata (Eds.), *English corpora under Japanese eyes* (pp. 231–249). Amsterdam: Rodopi.

Chujo, K., & Oghigian, K. (2009). How many words do you need to know to understand TOEIC, TOEFL and EIKEN? An examination of text coverage and high frequency vocabulary. *The Journal of Asia TEFL, 6*(2), 121–148.

Educational Testing Service. (2009). *The official guide to the TOEFL test* (3rd ed.). New York: McGraw-Hill.

Educational Testing Service. (2013). *Official TOEFL iBT tests with audio* (Vol. 1). New York: McGraw-Hill.

Heatley, A., Nation, I.S.P., & Coxhead, A. (2002). RANGE and FREQUENCY programs. Retrieved from http://www.vuw.ac.nz/lals/staff/Paul_Nation

Hirai, A. (2000). An analysis of vocabulary used in STEP 1st and pre-1st grades, TOEFL, and TOEIC. *言語文化論集, 54*, 1–16.

Hu, M., & Nation, I.S.P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language, 13*(1), 403–430.

Ishikawa, S., Uemura, T., Kaneda, M., Shimizu, S., Sugimori, N., Tono, Y., … Mochizuki, M. (2003). *大学英語教育学会基本語リスト* [JACET list of 8000 basic words]. Tokyo: JACET.

Kaneko, M. (2013). Estimating the reading vocabulary-size goal required for the Tokyo University entrance examination. *The Language Teacher, 37*(4), 40–45.

Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines* (pp. 316–323). Clevedon: Multilingual Matters.

Laufer, B., & Ravenhorst-Kalovski, G.C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language, 22*(1), 15–30.

Ministry of Education, Culture, Sports, Science, and Technology (MEXT). (2009). *高等学校学習指導要領* [The course of study guidelines for upper secondary school]. Tokyo: Author.

Mizumoto, A. (2006). TOEFL 語彙の語彙レベル、真正性、特徴の検証 [Word levels, authenticity, and characteristics verification of the TOEFL vocabulary]. *言語文化学会論集, 27*, 3–15.

Nation, I.S.P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review, 63*(1), 59–82. doi:10.3138/cmlr.63.1.59

Sato, M. (2013, May 31). 大学入試ＴＯＥＦＬ導入への疑義 [Dispute over the adoption of the TOEFL as part of university entrance examinations]. *内外教育, 6250*, 1.

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal, 95*, 26–43. doi:10.1111/j.1540-4781.2011.01146.x

West, M. (1953). *A general service list of English words*. London: Longman, Green and Co.

Xue, G., & Nation, I.S.P. (1984). A university word list. *Language Learning and Communication, 3*(2), 215–229.

Vocabulary Learning and Instruction
Volume 3, Issue 1, December 2014
http://vli-journal.org

Article

A Methodology for Identification of the Formulaic Language Most Representative of High-frequency Collocations

James Rogers[a], Chris Brizzard[a], Frank Daulton[b], Cosmin Florescu[c], Ian MacLean[a], Kayo Mimura[a], John O'Donoghue[d], Masaya Okamoto[e], Gordon Reid[a] and Yoshiaki Shimada[f]
[a]*Kansai Gaidai University;* [b]*Ryukoku University;* [c]*University of New England;* [d]*Osaka Board of Education;* [e]*University of Manchester;* [f]*State University of New York at Albany*
doi: http://dx.doi.org/10.7820/vli.v03.1.rogers.et.al

Abstract

Researchers have stated that learning formulaic language is key to achieving fluency. It has also been stated that studying vocabulary in this way is more efficient than isolated vocabulary learning. However, there is a lack of research in regards to which formulaic language should be taught. There is a further lack of research about how such formulaic language can be identified. This study aimed to evaluate a methodology for identifying the most common formulaic language. It compared multi-word unit identification results from both 500 and 1,000 example sentences and quantified how often native speakers opt to extend multi-word units beyond their core pivot and collocate. This study also identified and quantified colligational issues affecting multi-word unit identification. The results showed no difference in multi-word unit identification between 500 and 1,000 example sentences, that native speakers opted to extend multi-word units more than half of the time, and that colligational issues only affected approximately 3% of the items examined. This study concluded that 500 example sentences are just as reliable as 1,000 when identifying multi-word units. It also found that extending multi-word units beyond their core pivot and collocate is an essential step researchers should take. This study also found that a colligational treatment is necessary if the aim is to achieve the most accurate data; however, the percentage of items that were affected were small and the methodology time-consuming. This finding indicates that there is a need for improved software to better automate the steps taken.

1 Introduction

Researchers agree that knowledge of formulaic language is essential if native-like fluency is to be achieved (Cowie, 1998; Wray, 2002). However, many researchers also state that there is a severe lack of emphasis on formulaic language (Gitsaki, 1996; Nesselhauf, 2005). Thus students fail to develop formulaic knowledge and struggle to obtain native-like fluency. So why is there a lack of emphasis on this important aspect of language fluency despite that researchers have

Please cite this article as: Rogers, J., Brizzard, C., Daulton, F., Florescu, C., MacLean, I., Mimura, K., ... Shimada, Y. (2014). A methodology for identification of the formulaic language most representative of high-frequency collocations. *Vocabulary Learning and Instruction*, *3* (1), 51–65. doi: 10.7820/vli.v03.1.rogers.et.al

agreed on its importance? The reason is connected to the complexity of identifying such items, and the resulting lack of resources to help develop such fluency. Identifying high-frequency collocations and the formulaic language most representative of those collocations is a complex process. In addition, there is a lack of consensus regarding what a collocation is. Furthermore, various different methodologies have been used in the past to identify high-frequency collocations/formulaic language, but all have flaws and/or lack comprehensiveness. Thus despite recent advancements concerning how words and cooccurrence of words can be counted, there is still a lack of research incorporating this knowledge.

In response, this study introduces a methodology that identifies the formulaic language most representative of high-frequency collocations. This study also provides examples of the types of data that can be identified when using such a methodology.

2 Literature Review

Researchers agree that knowledge of formulaic language is central to language fluency and that collocation is a major part of formulaic language. Lewis (2000) believes teaching collocations to be "a top priority in every language course" (p. 8). But what exactly is a collocation? In fact, many researchers still struggle to agree on a comprehensive definition. Traditionally most researchers have defined collocations as the tendency for words to frequently cooccur (Biber, Johansson, Leech, Conrad, & Finegan, 1999; Shin, 2006). Other criteria have been utilized to delimit what can be considered as collocations, such as utilizing mutual information data, but such criteria have been found unreliable (Shin, 2006; Stubbs, 1995).

Other researchers recommend that only semantically opaque words that frequently cooccur be considered collocations (Moon, 1994) and that only such items be taught directly because they have a higher learning burden. However, researchers such as Nesselhauf (2005) and Wray (2000) highlight flaws in this approach. Semantic transparency does not necessarily equate to a lower learning burden since other criteria often affect a collocation's learning burden, such as L1-L2 congruency. In fact, Feyez-Hussein (1990) found that 50% of collocation errors were due to L1 influence. Thus, the most reliable criterion to initially identify collocations worth teaching is still frequency of co-occurrence, and therefore this study will utilize this definition.

If collocations are defined as words that frequently cooccur, how cooccurrence is counted must also be addressed. What exactly is a word and how should we count words? In fact, there are many ways to approach this. For instance, the simplest way to count words would be as *word types*. A *word type* distinguishes all lexical items that have different spellings. For instance, the word *eat* is considered separately to *eats*. However, this method of counting is not ideal for a study of high-frequency collocations because of the sheer number of collocations that exist. Hill (2000) estimates there to be hundreds of thousands of collocations in English. Thus consolidating data is essential.

Data consolidation can be accomplished in a number of ways. For example, it would be preferable to count the formulaic sequences *eat dinner* and *eats dinner*

together because the learning burden of learning one after the other is very low. With general affix knowledge, a learner can handle such differences on their own without both items needing to be taught at separate times. This can be accomplished by counting words as *lemma.* A lemma is a "set of related words consisting of the stem and inflected forms that are all the same part of speech" (Nation & Meara, 2002, p. 36). For example, the verbs *run, runs, running* and *ran*, are all counted together when lemmas are used, while the noun *run* would be counted separately. Counting word as *word families* consolidates data even further. A *word family* is "a headword, its inflected forms, and its closely related derived forms" (Nation, 2001, p. 8). When counting words with word families, the verb and noun forms of *run* are counted together and listed as a singular entry under the headword *run.*

Counting words as word families certainly has advantages in specific types of research, however, such a methodology can be problematic as well. This is because the headword that represents a word family is not always the most frequent lexical item that the family includes. For example, in Table 1, it is clear how the headword *depress* can be misleading. The word family is represented by the verb *depress,* despite the fact that the noun *depression* has significantly higher frequency. Furthermore, it is erroneous to make the assumption that learners can simply extend their affixed knowledge, thus equating to knowledge of the verb *depress* extending to knowledge of the whole word family. Both Schmitt and Meara (1997) and Daulton (2008) found Japanese learners to struggle with this task. So if the goal of a study is to identify a specific example of formulaic language to teach directly to learners, then breaking down word families into smaller groups of words, such as *lemma*, would be preferable. Lemma grouping can be broken down even further into word *types*, but this does not take advantage of a learner's ability to extend simple affix knowledge, such as the difference between the nouns *dog* and *dogs*, which is more plausible than, for instance, extending knowledge of a noun to an adjective.

For the above reasons, lemmatized collocation pairs are preferable when identifying high-frequency collocation, and therefore, this method of counting

Table 1. Frequency Counts in the Corpus of Contemporary American English (COCA; Davies, 2008) for Word Types in the Word Family *depress*

Word type	Frequency in the corpus
depression	19,176
depressed	6,715
depressing	2,032
depressive	1,598
anti-depressants	758
anti-depressant	533
depress	411
depressingly	152
depresses	144
depressant	58
depressives	31
depressants	18

words will be utilized in this study. The way this is achieved is by counting cooccurrence of words as *concgrams*.

A *concgram*, as defined by Cheng, Greaves, and Warren (2006), "constitutes all the permutations of constituency and positional variation generated by the association of two or more words" (p. 411). *Constituency variation* (AB, ACB) involves a pair of words not only cooccurring adjacent to one another (*lose weight*) but also with a constituent (*lose some weight*). *Positional variation* (AB, BA) refers to counting total occurrences of two or more particular lexical items that includes occurrences on either side of each other. Thus *provide you support* and *support you provide* would both be included in the total counts for a formulaic language concordance search for the lemma *provide* and *support*. Table 2 shows the first five results of an actual concgram search for the lemma *provide* and *support*. These data are sourced from the Corpus of Contemporary American English (COCA)'s online interface, which allows for lemma concgram searches and provides snippets of the sentences these concgrams are occurring in.

Then, this concordance data could be processed to identify the formulaic language, or *multi-word unit*, most representative for the lemma *provide* and *support*. When 500 such snippets from the COCA are processed, it is revealed that *provide support* is the most common multi-word unit. Table 3 shows the top three multi-word units for this lemma pair.

Concgramming has significant advantages when the goal is to identify formulaic language most representative of high-frequency collocations. Cheng et al. (2006) state that "searches which focus on contiguous collocations present an incomplete picture of the word associations that exist" (p. 431). In other words, attempts to identify formulaic language that are not done as concgram searches are not reliable. However, much of the previous research that aimed to identify high-frequency formulaic language was actually conducted in this way (Biber, Conrad, & Cortes, 2004; Shin, 2006; Simpson & Mendis, 2003). Therefore, there is a severe gap in the research that this study aims to fill.

Table 2. A Sample of Data from the COCA for a Concgram Search for the Lemma *provide* and *support*

. . . low-cost measures, the United States can extend the same lifesaving **support** that it has **provided** to the little boy in a rural, dusty village to the working-age woman living . . .
. . . psychiatrists, nurses, addiction and employment counselors, and peer **support** specialists. PHF **provides** community-based services, and a service coordinator is always on call to help clients address . . .
. . . it, then provide technical support to assist them. This **support** can usually be **provided** through a single phone call or demonstration. If needed, seek assistance from school . . .
. . . losing those aid dollars that we need in order to get **support** when Pakistan does **provide** it, which is real and does help us in the case of drones to . . .
. . . for low-income adults in occupational programs as well as financial **support** to colleges to **provide** support services for such students. States and colleges interested in adopting a model similar . . .

Table 3. Top Three Multi-word Units for the Lemma *provide* and *support* Found after Examining 500 Concordance Strings in the COCA

Multi-word unit	Frequency
provide support	55
support provided	39
support provided by	32

However, simply identifying lemma pairs that co-occur frequently is insufficient to provide learners with specific items to study. For instance, *take/walk* collocate, but it is not enough to simply expose students to this lemma pair. Rather, a more specific example of how the two collocate as a multi-word unit needs to be identified. Is it *taking walks, took walks, take a walk,* etc.? Thus steps are required to identify the multi-word unit most representative of that lemmatized concgram. This is accomplished via concordance software. However, working with concgrams is not simple, and thus this paper provides guidance on how this can be accomplished. Furthermore, another pertinent question is whether a multi-word unit identified as most representative of a lemmatized concgram should go beyond the pivot and collocate. For instance, should an identification method stop at *take a walk* or should it extend beyond this to identify *take a walk to*?

Colligation, or the counting various lexical items that can easily substitute for one another as grammatical categories (Gitsaki, 1996; Renouf & Sinclair, 1991), is another important criterion for formulaic sequence identification about which there is a lack of research. An example of colligation is counting the collocates *early* and *century* as *early* [*year*] *century* when they occur with years, which would account for instances, such as *early twentieth century*, *early nineteenth century*, etc., together. Table 4 shows the advantage of processing corpus data with consideration for colligation. One thousand example sentences were collected from the COCA (Davies, 2008), and a concordance search identified the multi-word unit most representative of how *century* and *earlier* occur together. One search was done with consideration for colligation, replacing every instance of a year with the marker *[year]*. By considering colligation, the top multi-word unit identified was shown to have nearly double the frequency in comparison with the top multi-word unit identified without consideration for colligation.

However, depending on the goal of the research, colligation also has the potential to create more problems than it solves. For instance, when major content word categories, such as nouns or verbs, are replaced with colligational markers, the limitations of how a multi-word unit can be formulated may not be conveyed to the learner. Take the colligational framework *[adjective] tea,* for instance. Typical examples such as *hot tea, brown tea,* or *strong tea* are perfectly logical, but it becomes very difficult to explain why *powerful tea* is not an option. Due to this idiosyncratic way collocations occur, grammar alone is not sufficient to determine which lexical items cooccur (Lewis, 2000). Regardless, colligation may be an important criterion to consider when identifying formulaic language. Yet how this criterion can be implemented and the extent of its value remains to be seen. Thus

Table 4. A Comparison between Two Multi-word Unit Searches, One with and One without Consideration for a Specific Type of Colligation

% of occurrences in 1,000 example sentences	Multi-word unit with cooccurrence of *century* and *early*	% of occurrences in 1,000 example sentences	Multi-word unit with cooccurrence of *century* and *early*
Without consideration for colligation		With consideration for colligation	
10.7	*century earlier*	19.2	*early in the [year] century*
9.5	*a century earlier*	10.7	*century earlier*
8.5	*early in this century*	9.7	*early [year] century*
7.3	*early in the century*	9.5	*a century earlier*
6.4	*centuries earlier*	8.5	*early in this century*
5.0	*early in the 20th century*	8.3	*early as the [year] century*
		8.3	*as early as the [year] century*
		7.3	*early in the century*
		6.4	*centuries earlier*

this paper aims to clarify the value of specific types of colligational searches and provide examples of the types of data that result from such consideration.

3 Research Questions

1. What percentage of the most common multi-word units is affected when specific types of colligations are considered?
2. Compared with the results of multi-word unit searches without consideration for colligation, what percentage of items identifies a different multi-word unit as being most representative of the lemmatized concgram?
3. Should multi-word units be extended beyond the pivot and collocates, at the beginning and end of a multi-word unit, to provide learners with more information about how the target items commonly occur formulaically?

4 Materials

This study will begin by utilizing Rogers et al.'s (in press) list of 12,604 high-frequency lemmatized concgrams. This list was originally derived from Davies' (2010) *Word List Plus Collocates,* a list of collocations that occur with the most frequent 5,000 lemmas of the COCA. To distinguish only items from this list that are useful for learners of general English, Rogers et al. (in press) delimited the list by frequency (approximately one occurrence per million tokens), and only included items with balanced range and chronological data.

Concordance data for each of the 12,604 concgrams was collected from the COCA. This study's approach necessitated the writing of custom concordance software to identify the most common multi-word units. Using normal concordance software, such as Anthony's (2011) *AntConc,* was not an option because

this study aimed to identify only multi-word units in which both lemma occurred, a function not possible with *AntConc* or other concordance software. Furthermore, the large amount of data (over 12,000 pairs) required a batch processing option, another feature not possible with current concordance software. Thus this study used the custom concordance software *AntWordPairs* (Anthony, 2013), a program written specifically for this study. It utilizes Someya's (1998) *E-lemma list*. For coding purposes, Someya's lemma list could not contain duplicate entries, and thus was modified to remove homonyms. For part of speech tagging, the software *GoTagger Version 0.7* (Goto, 2005) was utilized, and for colligational marker substitution, the software *Textcrawler* (Digital Volcano, 2011) was utilized.

5 Procedure

The first step was to collect concordance data (example sentences) for each of the 12,604 lemma pairs. Lemmatized concordance searches were conducted, using the COCA's online concordance interface, to identify instances when the collocate occurred either three words to the left or right of the node word. The rationale for this length (seven words) was influenced from findings on typical human memory limitations (Miller, 1956). The COCA's interface provides options for 100, 200, 500, or 1,000 example sentences to be extracted. Since more data provide the most reliable results, this study began by collecting 1,000 example sentences for each pair. However, because of COCA download limits, and the time required for sentences to load, 1,000 sentences was deemed impractical. However, to ensure that 500 example sentences provided as reliable data as 1,000 sentences would, results from 10 random lemma pairs were compared using both 500 and 1,000 example sentences. Starting with pairs which had frequency counts of 1,000 or more, every 500th pair was selected from the list which was sorted by frequency. Extracting 500 example sentences per lemma pair essentially created a mini corpus for each pair consisting of approximately 13,000 words per pair.

The next step was to identify specific categories of lexical items that occur in high frequency that could be substituted with colligational markers. Essentially, the goal was to experiment with a number of items that could be substituted with a marker that does not impede the meaning of the multi-word unit as a whole, while providing more accurate frequency counts. Table 2 is a perfect example of such an item. However, since no previous research existed, a number of items needed to be chosen and experimented with. A multi-word unit search was conducted on all 12,604 lemma pairs without consideration for collocation. A scan of the full data by a native English speaker revealed that particular categories of words (pronouns) occurred quite often in the multi-word units identified and could easily be substituted without disruption of the meaning of the multi-word units as a whole. In addition, a number of other word categories were used in the colligation treatment: *months, days of the week, oridinal numbers,* and *cardinal numbers.*

To use the colligational categories, adjustments for homonyms in the corpus data were necessary. This was done by part of speech tagging using *GoTagger* and making replacements using *Textcrawler.* First, all instances of the pre-nominal possessive pronoun *her* were changed to *his* as to not interfere with the object pronoun *her.* Then, instances of the ordinal number *second* were changed to 2nd as

to not interfere with the noun *second.* Next, instances of the nominal possessive personal pronoun *his* were changed to *hers* to not interfere with the pre-nominal possessive pronoun *his.* Then, the nominal possessive personal pronoun *mine* was replaced with *yours* to not interfere with the noun *mine.* Furthermore, instances of the month *May* and *March* were replaced with *January* to not interfere with the auxiliary verb *may* and the verb *march,* respectively. In addition, the day of the week abbreviations *Sun, Wed,* and *Sat* were replaced with *Mon* to not interfere with the noun *sun* and the verbs *wed* and *sat*, respectively.

Then, *Textcrawler* was used to replace all the pronouns, months, days of the week, ordinal and cardinal numbers with distinct colligational markers in each mini-corpus. The data were then processed with *AntWordPairs* to identify the most common multi-word units each lemma pair occur in. Because the amount of resulting data was excessive, only multi-word units occurring in 5% or more of the corpora were collected. Furthermore, a limit of seven words was set for the length of the multi-word units.

Next, five native English speakers examined the data to not only extract the most frequent multi-word unit but to also extend the multi-word unit beyond the most frequent item to its left or right when the native speaker judged any additions to be part of the natural unit.

The next step was a random sample of the multi-word units that were affected by the colligational treatment, and a concordance search with the original data not treated for colligational to judge whether a different multi-word unit was identified.

The final step taken in this study was to examine a random sample of 100 multi-word units identified and determine which percentage native speakers extended beyond the pivot and collocate.

6 Results

Data from 10 random concordance searches were examined for differences between using 500 and 1,000 example sentences.

Between the two amounts, the same top multi-word unit was identified for every pair examined, regardless of whether 500 or 1,000 example sentences were used. The data also show that the frequency counts varied very little when comparisons were made. Table 5 shows the top multi-word unit identified for each of the 10 pairs examined.

After the initial concordance search, distinct categories of words were found to occur frequently in the multi-word units identified. The vast majority of these was pronouns. Thus colligational markers were created for the following types of pronouns:

1. Pre-nominal possessive pronouns (your, his, her, their, my, our, its)
2. Subject pronouns (I, you, he, she, they, we, it)
3. Object pronouns (me, us, him, her, them)
4. Nominal possessive personal pronouns (theirs, his, hers, yours, mine)
5. Singular reflexive personal pronouns (myself, yourself, himself, herself, itself, yourselves, themselves, ourselves)

Table 5. The Top Multi-word Unit Identified When 500 and 1,000 Example Sentences Were Utilized

Lemma	POS	Lemma	POS	Multi-word unit identified	% out of 500 sentences	% out of 1,000 sentences
announce	verb	week	noun	announced last week	21.6	20.0
trade	noun	deficit	noun	trade deficit	85.6	84.7
body	adj.	upper	adj.	upper body	87.2	86.2
up	adv.	high	adv.	high up	70.0	66.5
little	adv.	better	adv.	little better	100	97.5
court	noun	hold	verb	court held	40.2	42.5
take	verb	charge	noun	take charge	46.4	38.7
care	verb	people	noun	people who care	15.4	10.8
get	verb	look	noun	get a look	23.2	15.7
too	adv.	often	adv.	too often	57.4	33.4

It was also determined that four other additional colligational categories should be replaced with colligational markers since they were seen occurring in the original concordance search, did not disrupt the meaning of the multi-word unit as a whole, and could potentially provide more accurate frequency counts. There were:

1. Months (January, Jan, February, Feb, Mar, April, Apr, May, June, Jun, July, July, August, Aug, September, Sept, October, Oct, November, Nov, December, Dec)
2. Days of the week (Sunday, Sun, Monday, Mon, Tuesday, Tue, Wednesday, Wed, Thursday, Thurs, Friday, Fri, Saturday, Sat)
3. Ordinal numbers (1^{st}, 2^{nd}, 3^{rd}, 4^{th}, 5^{th}, 6^{th}, 7^{th}, 8^{th}, 9^{th}, 10^{th}, 11^{th}, 12^{th}, 13^{th},14^{th}, 15^{th}, 16^{th}, 17^{th}, 18^{th}, 19^{th}, 20^{th}, 21^{st}, 30^{th}, 40^{th}, 50^{th}, 60^{th}, 70^{th}, 80^{th}, 90^{th}, 100^{th}, first, second, third, fourth, fifth, sixth, seventh, eighth, ninth, tenth, eleventh, twelfth, thirteenth, fourteenth, fifteenth, sixteenth, seventeenth, eighteenth, nineteenth, twentieth, twenty-first, thirtieth, fortieth, fiftieth, sixtieth, seventieth, eightieth, ninetieth, one-hundredth)
4. Cardinal numbers (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38,39,40,41,42,43,44,45,46,47,48,49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1,000, 2,000, 3,000, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 20,000, 30,000, 40,000, 50,000, 60,000, 70,000, 80,000, 90,000, 100,000, 200,000, 300,000, 400,000, 500,000, 600,000, 700,000, 800,000, 900,000, 1,000,000, one, two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, thirteen, fourteen, fifteen, sixteen, seventeen, eighteen, nineteen, twenty, thirty, forty, fifty, sixty, seventy, eighty, ninety, one-hundred, one-thousand, ten-thousand, one-hundred thousand, one-million)

It should be noted that these selections are not all-encompassing and other potentially useful colligational patterns may certainly be present in the data.

However, due to practical time and computing limitations this paper could only deal with the above colligational categories and the items listed within them.

After all the mini-corpora were adjusted for homonyms and processed with *AntWordPairs* to identify the multi-word units, and native speakers extracted the multi-word units most representative of how each lemma pair cooccurs, and the amount of multi-word units identified that were affected by the colligational treatment were counted. The results are shown in Table 6.

The colligational treatment for prenominal possessive pronouns was shown to be the most common. As much as 2.1% of the lemma pairs' top multi-word units were affected by this colligational treatment. Treatments for subject pronouns and cardinal numbers also resulted in a significant amount of items being affected. In total, 5.8% of all of the top multi-word units (727 items) identified were affected by all the colligational treatments conducted.

Ten random samples were then taken from the top three types of colligation treatment found to affect the top multi-word unit identification. These were then compared to a top multi-word unit search with untreated data. Out of the 30 items selected, only 13 (43.3%) resulted in different multi-word units being identified. For items affected by the prenominal possessive pronoun treatment, only 4 out of 10 top multi-word units differed. With the subject pronoun treatment, only 3 out of 10 top multi-word units differed. With the cardinal number treatment, 6 out of 10 of the top multi-word units differed. These results are summarized in Tables 7, 8, and 9.

Native speakers opted to extend multi-word units beyond the core pivot and collocate in 53 percent of the 100 random multi-word units sampled. For instance, the most frequent multi-word unit for the lemma pair *come* and *term* was found to be *come to terms*, at 243 occurrences (see Table 10). However, the next most common string in the data beyond *come to terms* was *come to terms with* (229 occurrences), and beyond that, *to come to terms with* (129 occurrences). Thus a native speaker judged *to* **come to terms** *with* as being the multi-word unit most representative of the lemma pair *come* and *term.* Core multi-word units were

Table 6. Amount of Top Multi-word Units That Were Affected by Each of the Colligational Treatments

Colligational treatment	Number of top multi-word units affected	Percentage of total lemma pairs (%)
Pre-nominal possessive pronouns	259	2.1
Subject pronouns	208	1.7
Cardinal numbers	171	1.4
Object pronouns	74	0.6
Ordinal numbers	14	0.1
Singular reflexive personal pronouns	1	0.007
Nominal possessive personal pronouns	0	0
Months	0	0
Days of the week	0	0
Grand totals	727	5.8

Table 7. Comparison between 10 Random Samples of Top Multi-Word Units Affected by the Colligational Treatment for Prenominal Possessive Pronouns and the Results That Would Have Occurred without the Treatment

Lemmatized concgram pair	Multi-word unit identified w/ colligational treatment	Multi-word unit identified w/o colligational treatment
***hand* (noun) *wave* (verb)**	***waved * hand***	***waved a hand***
***live* (verb) *life* (noun)**	***live * life***	***live life***
***base* (verb) *experience* (noun)**	***based on * experience***	***based on experience***
***attention* (noun) *focus* (verb)**	***focus * attention***	***focus attention***
head (noun) *gun* (noun)	*gun to * head*	*gun to his head*
hand (noun) *extend* (verb)	*extended * hand*	*extended his hand*
eye (noun) *wipe* (verb)	*wiped * eye*	*wiped her eye*
life (noun) *ruin* (verb)	*ruin * life*	*ruin your life*
put (verb) *hand* (noun)	*put * hand*	*put her hand*
sit (verb) *desk* (noun)	*sitting at * desk*	*sitting at his desk*

Note: Items in bold indicate those that showed differences in the top multi-word unit identified, and instances of a slot in which a pre-nominal possessive pronoun exists are represented with "*".

identified in bold and any strings present in the data and also judged to be typically cooccurring with the multi-word unit were added in italics. To accomplish this, native speakers relied on their intuition to not only add strings that truly represented common usage, but that also provided learners with useful information.

7 Discussion

Regarding the amount of data collected to create each mini-corpus used in this study, 500 example sentences were deemed as reliable as 1,000 example sentences when concordance data were compared. The example shown in Table 3

Table 8. Comparison between 10 Samples of Top Multi-Word Units Affected by the Colligational Treatment for Subject Pronouns and the Results That Would Have Occurred without the Treatment

Lemmatized concgram pair	Multi-word unit identified w/ colligational treatment	Multi-word unit identified w/o colligational treatment
***see* (verb) *mirror* (noun)**	***mirror * saw***	***mirror and saw***
***wear* (verb) *dress* (noun)**	***dress * wore***	***wearing a dress***
***take* (verb) *back* (adverb)**	***take it back***	***take back***
how (adverb) *interact* (verb)	*how * interact*	*how they interact*
get (verb) *when* (adverb)	*when * got*	*when I got*
make (verb) *hard* (adverb)	*makes * hard*	*makes it hard*
could (verb) *suppose* (verb)	*suppose * could*	*suppose you could*
belong (verb) *where* (adverb)	*where * belong*	*where I belong*
think (verb) *pretty* (adverb)	*think * is pretty*	*think she is pretty*
want (verb) *whenever* (adverb)	*whenever * want*	*whenever you want*

Note: Items in bold indicate those that showed differences in the top multi-word unit identified, and instances of a slot in which a subject pronoun exists are represented with "*".

Table 9. Comparison between 10 Random Samples of Top Multi-word Units Affected by the Colligational Treatment for Cardinal Numbers and the Results That Would Have Occurred without the Treatment.

Lemmatized concgram pair	Multi-word unit identified w/ colligational treatment	Multi-word unit identified w/o colligational treatment
***get* (verb) *second* (noun)**	***got * seconds***	***seconds to get***
***nearly* (adverb) *decade* (noun)**	***nearly * decades***	***nearly a decade***
***just* (adverb) *year* (noun)**	***just * years***	***just a few years***
***live* (verb) *mile* (noun)**	***live * miles***	***live within 50 miles***
***nearly* (adverb) *mile* (noun)**	***nearly * miles***	***nearly a mile***
***minute* (noun) *second* (noun)**	***minutes * seconds***	***seconds to one minute***
estimate (verb) *percent* (noun)	*estimates that * percent*	*estimates that 80 percent*
divide (verb) *group* (noun)	*divided into * groups*	*divided into two groups*
over (adverb) *month* (noun)	*over * months*	*over six months*
roughly (adverb) *percent* (noun)	*roughly * percent*	*roughly 10 percent*

Note: Items in bold indicate those that showed differences in the top multi-word unit identified, and instances of a slot in which a cardinal number exists are represented with "*".

demonstrates that collection of 500 versus 1,000 example sentences for each lemma pair made no difference in identifying the most common multi-word unit. However, collecting the data was a manual process of copy and pasting from the COCA's interface, something it was not designed for. Thus through the process unnecessary data were also copied, and therefore, a multi-step process of pasting into an Excel file, then copying only the sentences and pasting again into a Word file, and then saving the file, was necessary to remove this data. Being a cumbersome, time-consuming process, corpus computer interface designers may want to consider this for future design.

When the initial concordance data were examined after processing the compiled mini-corpora, various types of pronouns occurred quite often within the multi-word units identified. Other categories of words, such as cardinal numbers, also frequently occurred. Thus such word categories became the focus of this study's colligation experiment. However, because of a lack of previous research, other categories were experimented with as well. Not all of these proved

Table 10. Multi-word Units Identified from 500 Example Sentences in Which the Lemma Pair *come* and *term* Both Occur

Multi-word unit	Occurrences in 500 sentences
come to terms	243
come to terms *with*	229
to **come to terms**	133
to **come to terms** *with*	129
coming to terms	96
coming to terms with the	86
to **come to terms** *with the*	44
come to terms *with* [pre-nominal possessive pronoun]	28
coming to terms with the	26

fruitful, however, the resulting data did provide an insight as to specific types of colligation that, when addressed, can improve upon the reliability of multi-word unit identification.

The colligational treatment for prenominal possessive pronouns was shown to be the most useful. Treatments for singular reflexive personal pronouns, nominal possessive personal pronouns, months, and days of the week did not prove useful; only one item was affected in the entire list by all of these treatments. At first glance, the colligational treatment was shown to be an important step in the identification of the most frequent multi-word units, most representative of lemmatized concgrams, in that 727 (5.8%) of the total concgrams examined had their most common multi-word unit change. However, when a sample of the multi-word units was compared to the multi-word units that would have been identified without a treatment for colligation, only 43.3% of the items actually had differing results. Therefore, while frequent counts were always improved upon, the treatments did not always end with improved results.

Yet before the colligational treatment could be conducted, homonym interference in the data had to be dealt with. The process was complex, cumbersome, and very time-consuming due to the lack of dedicated software to conduct such a task. It would be useful if software developers considered such functionality and ways to improve the efficacy of conducting such data modification.

In regards to the value of extending multi-word units beyond the core pivot and collocate, the data suggest that this is an important criterion to consider when attempting to identify multi-word units most representative of lemmatized concgrams. Native speakers opted to extend multi-word units in more than half of the items examined. Corpus data and software alone cannot accurately identify such extensions, and thus this aspect of the study highlighted the importance of native speaker intuition and intervention in multi-word unit identification.

8 Conclusion

This study discussed a methodology to identify multi-word units most representative of lemmatized concgrams. It highlighted the value of counting words as lemma to identify formulaic language most representative of high-frequency collocations, compared results from different sized pivot word/collocate corpora, provided hard data as to the type of results one can expect when conducting specific colligational treatments on data, and highlighted the value of extending multi-word units beyond the core pivot and collocate.

This study showed how 500 example sentences that contain a target pivot word and collocate are just as reliable as 1,000 example sentences. This study also showed that multi-word unit searches for 5.8% of the lemma pairs examined were affected by the steps taken in this study. However, when a sample of these items was examined more deeply, it was found that nearly half showed no difference in the top multi-word unit identified. While results did improve for approximately 3% of the items examined, the steps needed to achieve these improvements were time-consuming and complex. Therefore, this study indicated the need for a more efficient methodology for such colligation treatments. Software designers should

thus consider ways to automate some of the steps taken in this study. This study also highlighted the importance of extending multi-word units beyond the core pivots and collocates, as over half of the items examined benefited from this procedure.

This study had its limitations. Due to the lack of previous research and no standard on how to conduct such a data analysis, choices for the types of colligation examined were subjective. Quite possibly other types of colligation exist in the data that could also prove fruitful if treated. Thus more research is needed in regards to other types of colligation that may improve results if treated. Despite these limitations, this study did provide new insights into a previously unexplored area of linguistic analysis that certainly has the potential for creating improved resources that help learners achieve fluency in a second language.

References

Anthony, L. (2011). *AntConc* (Version 3.2.2) [Computer Software]. Tokyo, Japan: Waseda University. Retrieved from http://www.antlab.sci.waseda.ac.jp/

Anthony, L. (2013). *AntWordPairs* (Version 1.0.2) [Computer Software]. Tokyo, Japan: Waseda University.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at. . .: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, *25*, 371–405. doi:10.1093/applin/25.3.371

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London, UK: Pearson Education.

Cheng, W., Greaves, C., & Warren, M. (2006). From n-gram to skipgram to concgram. *International Journal of Corpus Linguistics*, *11*, 411–433. doi:10.1075/ijcl.11.4.04che

Cowie, A. (Ed.). (1998). *Phraseology: Theory, analysis, and applications*. Oxford, UK: Oxford University Press.

Daulton, F. (2008). *Japan's built in lexicon of English-based loanwords*. Clevedon, Oh: Multilingual Matters.

Davies, M. (2008). *The corpus of contemporary American English: 450 million words,* 1990–present. Retrieved from http://corpus.byu.edu/coca/

Davies, M. (2010). *Word list plus collocates*. Retrieved from http://www.wordfrequency.info/purchase1.asp?i=c5a

Digital Volcano. (2011). *Textcrawler* (Version 2.5) [Computer Software]. Retrieved from http://www.digitalvolcano.co.uk/content/textcrawler

Feyez-Hussein, R. (1990). Collocations: The missing link in vocabulary acquisition amongst EFL learners. In J. Fisiak (Ed.), *Papers and studies in contrastive linguistics: The Polish English contrastive project* (vol. 26, pp. 123–136). Poznan: Adam Mickiewicz University.

Gitsaki, C. (1996). The development of ESL collocational knowledge (Unpublished doctoral dissertation). University of Queensland, Brisbane.

Goto, K. (2005). *GoTagger* (Version 0.7) [Computer Software]. Retrieved from http://web4u.setsunan.ac.jp/Website/GoTagger.htm

Hill, J. (2000). Revising priorities: From grammatical failure to collocational success. In M. Lewis (Ed.), *Teaching collocation: Further developments in the lexical approach* (pp. 47–67). Hove, UK: Thomson Heinle Language Teaching.

Lewis, M. (2000). Language in the lexical approach. In M. Lewis (Ed.), *Teaching collocation: Further developments in the lexical approach* (pp. 8–10). Hove, UK: Thomson Heinle Language Teaching.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97. doi:10.1037/h0043158

Moon, R. (1994). The analysis of fixed expressions in text. In M. Coulthard (Ed.), *Advances in written text analysis* (pp. 117–135). London, UK: Routledge.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139524759

Nation, P., & Meara, P. (2002). Vocabulary. In N. Schmitt (Ed.), *An introduction to applied linguistics* (pp. 35–54). London, UK: Edward Arnold.

Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam: John Benjamins.

Renouf, A., & Sinclair, J. (1991). Collocational frameworks in English. In K. Aijmer & B. Altenberg (Eds.), *English corpus linguistics* (pp. 128–143). Harlow, UK: Longman.

Rogers, J., Daulton, F., Brizzard, C., Florescu, C., MacLean, I., Mimura, K., … Shimada, Y. (in press). On using corpus frequency, dispersion and chronological data to help identify useful collocations. *Research in Corpus Linguistics*.

Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework. *Studies in Second Language Acquisition*, *19* (1), 17–36. doi:10.1017/S0272263197001022

Shin, D. (2006). *A collocation inventory for beginners* (Unpublished doctoral dissertation). Wellington, New Zealand: Victoria University of Wellington.

Simpson, R., & Mendis, D. (2003). A corpus-based study of idioms in academic speech. *TESOL Quarterly*, *37*, 419–441. doi:10.2307/3588398

Someya, Y. (1998). *E-lemma list*. Retrieved from http://www.antlab.sci.waseda.ac.jp/software/resources/e_lemma.zip

Stubbs, M. (1995). Collocations and semantic profiles: On the cause of the trouble with quantitative methods. *Functions of Language*, *2* (1), 23–55. doi:10.1075/fol.2.1.03stu

Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics*, *21*, 463–489. doi:10.1093/applin/21.4.463

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge, UK: Cambridge University Press.

Vocabulary Learning and Instruction
Volume 3, Issue 1, December 2014
http://vli-journal.org

Commentary

Reaction Time Methodologies and Lexical Access in Applied Linguistics

John P. Racine
Dokkyo University
doi: http://dx.doi.org/10.7820/vli.v03.1.racine

1 Introduction

The first issue of this journal featured a paper by Iso (2012) in which the author described research conducted to validate his Lexical Access Time Test (LEXATT2). While the details of the test procedure are scant in the write-up, it appears that there are a number of methodological issues that require thorough examination before this test can be considered a valid measure of lexical access. Notable among these issues are the accuracy of the reaction time (RT) measurements and the manner in which the RTss are interpreted. Other aspects of the study–including its relation to prior research and theory–also deserve scrutiny.

As pointed out by Mochizuki (2012) in a discussion of four vocabulary test studies which included Iso's, very few researchers in applied linguistics are conducting research on lexical access. Indeed, lexical research involving RT measurement of any kind is particularly rare in our field, despite the abundance of such studies in cognitive psychology and psycholinguistics circles. The comments below are thus not intended as criticism of Iso's paper in particular. Rather, this commentary is intended to highlight some of the general principles central to cognitive approaches to the mental lexicon, particularly where RT is to be measured. It is hoped that the issues raised below will serve as an introduction to some key elements of this kind of research and encourage lexical researchers who wish to undertake further studies in this area.

2 The Original Study

Iso's LEXATT2 program attempts to test lexical access by examining the difference in RTs between nonverbal and verbal recognition tasks. Nonverbal task trials began with subjects clicking and holding down a mouse button. A stimulus (in this case, a circular shape) then appeared on screen and subjects were required to release the button when they saw it. When they released the button, the stimulus disappeared. The RTs (i.e., the time between presentation of the stimulus and when the button was released) from five such trials were recorded and the average time was calculated. Iso calls this the "mean reaction time for nonverbal stimulus" (MRT (NV); p. 78). In the verbal task trials, subjects once again began by holding down a mouse button. In this case a word appeared (either four, six, or eight letters in length). After reading the word, subjects again released the button and the RT

Please cite this article as: Racine, J.P. (2014). Reaction time methodologies and lexical access in applied linguistics. *Vocabulary Learning and Instruction*, *3* (1), 66–70. doi: 10.7820/vli.v03.1.racine

was recorded (RT for verbal stimulus, RT (VL)). Lexical access time was thus calculated for each word as:

$$RT\,(VL) - MRT\,(NV)$$

Verbal task trials included a second stage in which subjects were then given a test of receptive knowledge in which they had to choose the meaning or a synonym of the word from a number of options. For trials in which subjects failed to recognize the meaning of the word in this stage, RTs were discarded as subjects had either not known the word or had presumably failed to read it correctly.

3 Technical Considerations in RT Research

One of the fundamental issues that researchers in applied linguistics should realize about RT measurements, such as the one employed here, is that they are extremely delicate. In any experiment involving RT, many trials fail to record what they are intended to. Accidental keystrokes, lapses in subject attention, and distractions in the laboratory environment are common factors. In research where more than one key are used to trigger the timer (e.g., yes/no responses), slower left-hand reactions (for right-hand dominant participants) need to be accounted for and counterbalanced (Shen & Franz, 2005). Some researchers have attempted to circumvent these problems with voice-activated triggers, only to find that hems and haws, coughing, and extraneous noises in the lab can lead to just as many lost trials. There are more technical issues to consider as well. Screen and keyboard refresh rates can vary widely and can contribute dozens of milliseconds to RTs. Insensitive or oversensitive microphones regularly lead to lost data in voice-triggered RT studies as well. While researchers hope that these issues occur randomly across conditions, this may not always be the case and thus certain measures should be taken to limit the influence these factors have on resulting data.

Common methods of preventing undue influence on RTs from the factors listed above include allowing subjects ample practice on task and the inclusion of an abundance of experimental trials. Iso makes no mention of practice at all and bases the MRT (NV) measure on the mean from only five nonverbal trials. Compare this to the eight practice trials followed by 160 experimental trials in a study of attribution (Bassili & Racine, 1990; Expt. 1). Lexical decision studies similar to Iso's invariably include such practice sessions followed by an abundance of experimental trials. Fitzpatrick and Izura's (2011) study, for example, included 20 practice trials followed by 72 experimental ones. Balota and Chumbley (1984) utilized as many as 60 practice trials before subjects received experimental trials. This study also employed five buffer trials at the beginning of each experimental block. A buffer trial is one that appears to subjects as any other experimental trial, but the results from which are not included in the data. The decision to discard this data is made by the researcher a priori, and is an excellent way to allow participants to become accustomed to experimental procedures without adversely affecting the final results.

Another common procedure for avoiding undue influence on RT scores is to normalize data via the elimination of outliers. In experiments where average

response latencies may only be 200 or 300 ms in duration, a distraction of 5 or 6 seconds can have a huge impact on experimental cells. Iso does not report utilizing any measures to adjust or omit such erroneous data. As can be seen in the sampling of lexical decision task (LDT) studies listed in Table 1, however, it is very commonplace for researchers to employ such adjustments. In these studies, outliers were either adjusted (e.g., a score exceeding 2.5 standard deviations (*SD*s) above a subject's mean RT was replaced with a score of precisely 2.5 *SD*s above the mean) or they were eliminated from the data entirely.

4 Conceptual issues

A more serious concern for the validity of this study can be seen in the way the RT scores are interpreted. Iso's verbal task, as explained above, consists of two stages: the RT measure and a meaning confirmation stage. Subjects were asked to read the word and release the button when they had done so. The word recognition task is a separate stage, Iso explains, "so that LEXATT2 is not measuring word recognition time" (p. 79). Then the question is: What is it measuring? What exactly are subjects doing before they release the button? Have they said the words to themselves? Have they thought about the meaning? What are the L2-learning subjects doing during this time if they do not know the word? More essentially, can a subject "read" a word without considering its meaning? And if that was possible, would subjects allow themselves to do so, knowing that they are about to be given a multiple choice recognition task? Iso would have us believe that the RT measure here is equal to fixation time on text as measured in experiments involving passage reading. Thus "lexical access" is something akin to perception of the orthography of a word before semantic processing has occurred. In fact, these two processes are not necessarily serial tasks and are often treated as inseparable. This is precisely how they are treated in the RT studies cited above. All of these LDTs measure RTs when subjects have completed judgments as to whether a string of letters constitutes a word. Researchers typically employ words and nonwords as stimuli in these studies. Semantic processing is inherent to the subject's task and that is

Table 1. Outliers in Lexical Decision Task Studies

LDT study	Outliers
Altarriba and Knickerbocker (2011)	2.5 *SD*s above or below a participant's mean RT
Fitzpatrick and Izura (2011)	3 *SD*s above or below a participant's mean RT
Hu and Jiang (2011)	2.5 *SD*s above or below a participant's mean RT Responses exceeding a 2500 ms time limit
Segalowitz and Segalowitz (1993)	2 *SD*s above or below a participant's mean RT Responses exceeding a 3000 ms time limit

what the RT scores are intended to measure. Iso should have designed his methodology similarly.

It is also interesting to note that Segalowitz and Segalowitz's (1993) research (listed in Table 1) is the only RT study cited in the Iso (2012) paper. This gives the impression that Iso's research is based upon, or is perhaps even a replication attempt, the former study. In fact, there are fundamental differences between the two studies. Some of these differences have already been explained above (i.e., Iso did not make adjustments for outliers and employed a lexical task for which the RT scores may not be interpretable). Another difference is that the Segalowitz and Segalowitz (1993) study includes a review of relevant prior research. While it may have been due to editorial considerations beyond Iso's control, his is the only study cited here that does not include a thorough literature review as background to the study. Another approach would be to cast the study as an attempt to explore the validity of a model or models of second-language lexical access (e.g., de Bot, 2010; de Bot, Paribakht, & Bingham Wesche, 1997; Kroll & Dijkstra, 2010). Without a broader literature review and a more detailed write-up, it is difficult to position Iso's work within the field.

Finally, as I stated in the introduction to this commentary, there are as yet very few applications of RT methodologies in applied linguistics. For that reason, Iso should be commended for his exploratory work. It is hoped that these comments will help to inspire further studies in the field and I welcome a LEXATT3 program that addresses the issues raised here.

References

Altarriba, J., & Knickerbocker, H. (2011). Acquiring second language vocabulary through the use of images and words. In P. Trofimovich & K. McDonough (Eds.), *Applying priming methods to L2 learning, teaching and research* (pp. 21–48). Amsterdam: Benjamins.

Balota, D.A., & Chumbley, J.I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(3), 340–357. doi:10.1037/0096-1523.10.3.340

Bassili, J.N., & Racine, J.P. (1990). On the process relationship between person and situation judgments in attribution. *Journal of Personality and Social Psychology, 59*(5), 881–890. doi:10.1037/0022-3514.59.5.881

de Bot, K. (2010). Cognitive processing in bilinguals: From static to dynamic models. In R.B. Kaplan (Ed.), *The Oxford handbook of applied linguistics* (2nd ed., pp. 335–348). Oxford, UK: Oxford University Press.

de Bot, K., Paribakht, T.S., & Bingham Wesche, M. (1997). Toward a lexical processing model for the study of second language vocabulary acquisition. *Studies in Second Language Acquisition, 19*, 309–329.

Fitzpatrick, T., & Izura, C. (2011). Word association in L1 and L2: An exploratory study of response types, response times, and interlingual mediation. *Studies in*

Second Language Acquisition, *33*(03), 373–398. doi:10.1017/S0272263111000027

Hu, G.G., & Jiang, N. (2011). Semantic integration in listening comprehension in a second language. In P. Trofimovich & K. McDonough (Eds.), *Applying priming methods to L2 learning, teaching and research* (pp. 199–218). Amsterdam: Benjamins.

Iso, T. (2012). Examining the validity of the Lexical Access Time Test (LEXATT2). *Vocabulary Learning and Instruction, 01*(1), 78–82. doi:10.7820/vli.v01.1.iso

Kroll, J.F., & Dijkstra, T. (2010). The bilingual lexicon. In R.B. Kaplan (Ed.), *The Oxford handbook of applied linguistics* (2nd ed., pp. 349–370). Oxford, UK: Oxford University Press.

Mochizuki, M. (2012). Four empirical vocabulary test studies in the three dimensional framework. *Vocabulary Learning and Instruction*, *01*(1), 44–52. doi:10.7820/vli.v01.1.mochizuki

Segalowitz, N.S., & Segalowitz, S.J. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics*, *14*(03), 369–385. doi:10.1017/S0142716400010845

Shen, Y.-C., & Franz, E.A. (2005). Hemispheric competition in left-handers on bimanual reaction time tasks. *Journal of Motor Behavior*, *37*(1), 3–9. doi:10.3200/JMBR.37.1.3-9