Welcome from the Organizers

The Conference Committee comprising Rob Waring, Laurence Anthony, Charles Browne, and Tomoko Ishii welcome attendees and presenters to Vocab@Tokyo.

We would like to extend our thanks to Meiji Gakuin University, its staff and faculty for supporting the conference. We would also like to thank the Japan Association of Language Teachers (JALT) Vocabulary SIG for their support, for sponsoring the conference Plenary speaker Dr. Tess Fitzpatrick, and for organizing the Vocabulary Round Table on Day 3.

Our gratitude also extends to the Conference Exhibitors: Cambridge University Press, English Central, National Geographic Cengage, Nellie’s English Books, Oxford University Press, Pearson Education, Seed Learning, and xReading.

Lastly we would like to thank the many other people who have helped us, including the proposal reviewers, general helpers, and the dozens of Meiji Gakuin University student volunteers, all of whom are 3rd and 4th year seminar students of Charles Browne, training to become English teachers.

We hope you all have a great conference.

The Conference Committee:
Dr. Rob Waring
Dr. Laurence Anthony
Dr. Charles Browne
Dr. Tomoko Ishii
A Message from the University President

It is my honor to welcome you to the Vocab@Tokyo Conference at Meiji Gakuin University.

In 1863, Dr. James Curtis Hepburn, a missionary from the United States, founded the Hepburn School which eventually developed into Meiji Gakuin University, and we therefore have a history of over 150 years of contributing to English education in Japan.

Our university was founded based on Christian principles, and “Do for Others” has always been one of the key philosophies of our education. This remains true today as we try to help our students to develop an international perspective, understanding and sympathy towards others, and a desire to make a contribution to society.

We feel honored to be able to host such a gathering of great vocabulary researchers from all over the world, and hope that the history and tradition of our school will provide a good atmosphere for the exchange of ideas and information.

We sincerely wish that your conference will be a great success.

Yasuo Matsubara
President, Meiji Gakuin University
About Vocab@Tokyo

The Vocab@Tokyo Vocabulary Conference is the second in the series of “Vocab@” conferences, where researchers from around the world can gather and discuss the latest trends in vocabulary learning and teaching. The inaugural “Vocab@Vic” conference was held at Victoria University, Wellington, New Zealand in December 2013. We hope that our second conference in the series can maintain the outstanding quality and friendly atmosphere of the first.

Nominations for the next Vocab@ conference are open! On the final day, after the final session, we will discuss where we can hold it next.

Conference Organizing Committee

Conference Chair: Rob Waring (Notre Dame Seishin University, Japan)
Program Chair: Laurence Anthony (Waseda University, Japan)
Co-Site Chair: Charles Browne (Meiji Gakuin University, Japan)
Co-Site Chair: Tomoko Ishii (Meiji Gakuin University, Japan)

Vocab@Tokyo Handbook
Brandon Kramer (Editor), Imogen Custance, Clint Denison, Steve Porritt

Scientific Committee
Kiyomi Chujo, Tom Cobb, David Coulson, Averil Coxhead, Tad Kamimoto, Batia Laufer, Atsushi Mizumoto, Tatsuya Nakata, Geoffrey Pinchbeck, John Racine, Diane Schmitt, Anna Siyanova, Jeff Stewart, Yukio Tono, Stuart Webb

And a big thank you to all of the student volunteers from Charles Browne’s seminar!

Conference Website

The Vocab@Tokyo Vocabulary website is sites.google.com/site/vocabattokyo/home
Conference Overview

Monday, September 12th

from 8:30
9:00-9:30 Registration
9:30-9:55 Conference Opening
10:00-12:25 Featured Speaker: Batia Laufer
13:30-14:40 Paper Presentations
14:50-15:45 Plenary: Tess Fitzpatrick
16:20-16:45 Paper Presentations
16:50-18:00 Featured Speaker: Tom Cobb
18:00-18:30 Colloquiums
18:30-21:00 Group Photo
18:30-21:00 Conference Banquet

Tuesday, September 13th

8:00-8:45 Registration and Refreshments
9:00-9:25 Featured Speaker: Yukio Tono
9:30-12:25 Paper Presentations
13:30-14:40 Poster Session
14:50-15:45 Paper Presentations
16:20-16:45 Featured Speaker: Diane Schmitt
16:50-18:00 Colloquiums
18:00-21:00 “Dinner on Tokyo Bay”

Wednesday, September 14th

8:00-8:45 Registration and Refreshments
8:30-9:15 Vocab SIG Annual Meeting
9:30-12:25 Paper Presentations
13:30-14:40 Poster Session
14:50-15:45 Paper Presentations
16:20-17:30 Vocabulary Round Table
17:30-17:45 Closing Ceremony
17:45-18:00 Closing Photo
General Conference Information

Venue Description
Meiji Gakuin University, Shirokanedai Campus
Address: 1-2-37 Shirokanedai, Minato, Tokyo 108-8636

Registration & Help Desks
The registration and help desks will be located in the Palette Zone in Building 11 and will remain open throughout the conference. Feel free to ask us any questions. Student volunteers will be wearing blue conference t-shirts and will also be glad to help answer your questions or point you in the right direction.

Storage
Bags can be stored in the cloak room in the Main Building. Please note that neither the conference nor the university will take responsibility for any bags left but will do their best to protect your luggage.

Catering
There will be free coffee, tea, and water offered throughout the day in the Palette Zone (Building 11) starting from 8:30 AM. Free Japanese and Western snacks will also be offered during the mid-morning and mid-afternoon conference breaks. No lunches are included in the conference fee but lunch is available in the university canteen (Bldg. 11, 1st Floor). The banquet on the first evening is free for conference participants, while the ‘Dinner on Tokyo Bay’ event costs 7,000 yen and pre-registration is required.

Parking
The university will not offer parking, but there are some pay parking lots nearby.

Train Lines
The nearest train stations are Shirokanedai, Shirokanetakanawa, and Takanawadai. It should take about 7 minutes to walk from any of these stations. If you are arriving at Shinagawa Station the Toei bus or a taxi are recommended. Check the website for more information:
<www.meijigakuin.ac.jp/en/access/>

Computer & Internet Connections
Each presentation lecture room will be equipped with a LAN Internet-connected PC (Windows), projector, and screen. We recommend any presenters wanting to use the Internet during their presentation to use this equipment. If you choose to use your own computer, you will not be able to use the LAN Internet. However, it will be possible to connect your computer to the projector via a standard RGB cable. Please bring your own computer dongle as necessary. The university is part of the Eduroam Wifi network, so if your own institution belongs to Eduroam, it should be possible to easily connect via Wifi. Note, however, that the connections in the campus buildings are a little weak and cannot be guaranteed to be sufficient for presentation demonstrations.

Poster Sessions
Posters will be on display on Day 2 and Day 3 of the conference from 13:30 to 14:45 in the Palette Zone in Building 11. The poster board locations will be set up during the morning break and should not be moved. Presenters should put up their posters during the lunch break and may leave the posters up until after the mini-break at 17:15.
Conference Events

Opening Session
The Opening Session is on Day 1 from 9:00 to 9:30. There will be an opening address by the organizing committee, Rob Waring, Laurence Anthony, Charles Browne, and Tomoko Ishii. A short welcoming message will be given by the Meiji Gakuin University President Yasuo Matsubara.

Photo Sessions - Day 1, Day 3
There will be two group photo opportunities. The first will be before the conference Banquet on Day 1 at about 18:00 to 18:15 in the Palette Zone (registration/resting area). The second will be after the closing session on Day 3 in the same place. The photos will be put on the conference website for you to download.

Conference Banquet – Day 1
A free banquet dinner for all attendees will be held in the Palette Zone (registration/resting area) from 18:15 on Day 1 after the group photo. First there will be a live performance of traditional Japanese music for about 20 minutes during which the food will be put out and attendees will be able to listen to the performance while enjoying an aperitif. After the toast, attendees may enjoy the banquet and background music from the performers. The banquet finishes at around 21:00. Please leave the conference site quietly and respectfully. The conference organizers have made efforts to accommodate as many dietary requirements as is possible. The food will be labelled as such.

Evening Out – Day 2
The ‘Dinner on Tokyo Bay’ event has been organized at TY Harbor in Shinagawa. We will be organizing buses to take people from the conference site to the beautiful Shinagawa Bay area of Tokyo, where you can enjoy dinner overlooking the water at the TY Harbor restaurant. The cost of the event will be ¥7,000 and covers a gourmet buffet meal, a range of beers, wines, and non-alcoholic drinks, and transportation to the site. The site is a short 5 minute walk to Tennozu Isle Station or a 15 minute walk to a Shinagawa station so returning from the site should be trouble free. Pre-registration is required and must have been made by the end of August. There is a strict maximum of 100 places and no more may be added after that.

Extensive Reading Foundation Language Learner Literature Awards – Day 2
The Extensive Reading Foundation Language Learner Literature Award winners will be announced at Vocab@Tokyo around lunchtime on Day 2 in the posters and presentations area (Palette Zone, Building 11).
Conference Partners and Exhibitors

Conference Partners

Conference Exhibitors

Please visit our exhibitors’ promotional materials in the Palette Zone.
Venue & Getting Around

Campus Map

Ceremonies, Paper Presentations, Featured Speakers, and Plenaries are in the Main Building.

The Palette Zone, which includes the Registration, Poster Presentations, Publisher Booths, and Refreshments, is in Building 11.

Main Building

2nd Floor (in name only, more like the 1st floor)

Rooms 1254, 1255, & 1201 will hold the paper presentations.

1st Floor (in name only, more like the basement)

Opening, featured speakers, plenary, & Round Table in Room 1101.
## Abstract Contents at a Glance

<table>
<thead>
<tr>
<th>Section</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plenary</td>
<td>pp. 10-11</td>
</tr>
<tr>
<td>Featured Speakers</td>
<td>pp. 12-18</td>
</tr>
<tr>
<td>Vocabulary Round Table</td>
<td>pp. 19-20</td>
</tr>
<tr>
<td>Colloquiums: Day 1</td>
<td>pp. 21-25</td>
</tr>
<tr>
<td>Colloquiums: Day 2</td>
<td>pp. 26-31</td>
</tr>
<tr>
<td>Paper Presentations: Day 1</td>
<td>pp. 32-36</td>
</tr>
<tr>
<td>Paper Presentations: Day 2</td>
<td>pp. 37-42</td>
</tr>
<tr>
<td></td>
<td>pp. 43-48</td>
</tr>
<tr>
<td></td>
<td>pp. 49-54</td>
</tr>
<tr>
<td></td>
<td>pp. 55-60</td>
</tr>
<tr>
<td></td>
<td>pp. 61-66</td>
</tr>
<tr>
<td>Paper Presentations: Day 3</td>
<td>pp. 67-71</td>
</tr>
<tr>
<td></td>
<td>pp. 72-77</td>
</tr>
<tr>
<td></td>
<td>pp. 78-83</td>
</tr>
<tr>
<td></td>
<td>pp. 84-89</td>
</tr>
<tr>
<td></td>
<td>pp. 90-93</td>
</tr>
<tr>
<td></td>
<td>pp. 94-98</td>
</tr>
<tr>
<td></td>
<td>pp. 99-104</td>
</tr>
<tr>
<td>Paper Presentations: Day 3</td>
<td>pp. 105-110</td>
</tr>
<tr>
<td></td>
<td>pp. 111-116</td>
</tr>
<tr>
<td></td>
<td>pp. 117-122</td>
</tr>
<tr>
<td></td>
<td>pp. 123-128</td>
</tr>
<tr>
<td></td>
<td>pp. 129-134</td>
</tr>
<tr>
<td></td>
<td>pp. 135-140</td>
</tr>
<tr>
<td></td>
<td>pp. 141-145</td>
</tr>
<tr>
<td>Poster Presentations: Day 2</td>
<td>pp. 146-174</td>
</tr>
<tr>
<td>Poster Presentations: Day 3</td>
<td>pp. 178-210</td>
</tr>
</tbody>
</table>
1 Introduction

Understanding the process of word retrieval is crucial to research on vocabulary acquisition and word knowledge. It underpins models of lexical production, fluency, and communicative competence, and is a key consideration in measures of these constructs.

2 Aims of talk

This talk considers current conceptualisations of word retrieval in second language vocabulary research, focusing on models of word knowledge that accommodate notions of activation, automaticity, and the organisation of lexical networks, such as Crossley & Salsbury (2010), Henriksen (1999), Meara (1996), Qian (2002), Van Hell & de Groot (1998). These are then examined in the light of theoretical models and empirical findings from L1 studies and from research into communication disorders (including Caramazza 1997, Levelt 1989, Ramscar et al. 2014, Ulman 2004). This enables us to tease apart behaviour that is caused by incomplete word knowledge from other kinds of retrieval impairment. In practical terms, occasional and temporary impairment to word retrieval is an experience common to us all in both L1 and L2 use (Brown 1991), and in particular conditions impairment can be frequent and permanent (Foygel & Dell 2000, Knibb & Hodges 2005). Quality of word knowledge undoubtedly contributes to the efficiency of word retrieval, but it is not the whole story; a successful lexical production event can also be dependent on the ‘route’ used for retrieval.

The talk then presents a principled profiling framework that captures the retrieval routes used in word production (Fitzpatrick et al. 2015). This sheds light on the developing lexicon in two ways. Firstly, patterns common to all language users emerge, providing a baseline of generic word retrieval behaviour. But important individual differences emerge too, and these are stable across test times and across languages used (Fitzpatrick 2009). Secondly, the quantitative nature of the profile facilitates comparison with lexical size and with speed of word retrieval (Fitzpatrick & Izura 2011), and this enables us to track interactions between multiple dimensions of lexical storage and access.

By scrutinising word retrieval patterns in different kinds of language user, hypotheses can be generated about causes of individual difference. To demonstrate this, three profile comparisons are explored: between L1 and L2 language users, between older and younger adults, and, for a set of medical vocabulary, between professional and lay users. The participant groups represent different acquisition stages, resource size, and applications of lexical knowledge, and enable us to relate word retrieval patterns to these dynamic qualities of the lexicon.

The talk concludes by considering the extent to which changes in word retrieval behaviour can be deliberately induced through effective deployment of retrieval strategies, and by re-evaluating the contribution of word retrieval to models of vocabulary knowledge.

References
Caramazza, A. (1997). How many levels of processing are there in lexical access?. Cognitive neuropsychology, 14(1), 177-208.


Learning Collocations in L3: Effects of L1, L2 and Collocation Frequency

Batia Laufer
University of Haifa
batialau -at- research.haifa.ac.il

1 Background
Research on collocations has attempted to identify factors that make collocation learning easy or difficult for foreign language learners, such as collocation frequency, part of speech, restrictedness, and congruency with L1 (Paquot & Granger, 2012; Peters, 2016; Webb & Kagimoto, 2011; Wolter & Gyllstad, 2013; Yamashita & Jiang, 2010). To my knowledge, no studies have examined the relative and combined effects of L1, L2, and collocation frequency on the knowledge of collocations in L3.

2 Aims
Two studies investigate EFL learners' knowledge of ninety Verb + Noun and Adjective + Noun collocations as a function of 'congruency status' and frequency of collocations. Each collocation belonged to one of three types of 'congruency status': congruent with L1 (Arabic), congruent with L2 (Hebrew), but non-congruent with L1, and non-congruent with either L1 or L2. In each 'congruency status' group half of the collocations were frequent, i.e. appeared more than 300 times in COCA and half were infrequent, i.e. appeared less than 300 times. The aims of the studies were to find out which of the two factors, congruency status or frequency is more influential in collocation knowledge and to look at the collocation test scores as a function of possible combinations of these two factors.

3 Design and Procedure
The studies used a within-subject design. Sixty high school learners took two tests, recall and recognition. In the first test, they had to supply the missing collocate, verb or adjective, by translating L1 collocations into English. The nouns of the collocations and the first letters of the collocates were provided. A correct answer was considered a correctly spelled collocate. In the second test, participants chose the English equivalent of L1 translation from four options. Answers were scored dichotomously. Data were analyzed by 3 x 2 Repeated Measures and Bonferroni pairwise comparisons.

4 Results
Main effects of 'congruency status' and frequency were found for Adj + N and V + N collocations on recall and recognition tests. Pairwise comparisons showed that effect sizes were larger for differences between collocations of different congruency status than between frequent and non-frequent collocations. This suggests that congruency is a more influential factor than frequency in collocation learning. As expected, collocations congruent with L1 received the highest scores and non-congruent collocations—the lowest. Collocations congruent with L2 were known better than non-congruent collocations, suggesting that learners rely on any language they are familiar with. Infrequent collocations congruent with L1 or L2 received higher scores than frequent non-congruent collocations, supporting the contention that congruency between familiar languages and a new language is the most determining factor in collocation learning. The teaching implication of these results is that the burden of collocation learning can be reduced by excluding most of the congruent collocations from syllabi and focusing on incongruent collocations only.

References
Broad and Narrow Band Frequency Profiling

Tom Cobb
Université du Québec à Montréal
cobb.tom -at- sympatico.ca

1  What is LFP?

Lexical frequency profiling (LFP) is a type of computational text analysis (Nation & Heatley, 1994; Heatley, Nation, & Coxhead, 2002; Anthony, 2015) that draws on corpus-based frequency lists to assign each word of an entered text to one of several 1,000 word-family frequency bands (also known as k-lists). The profile takes the form of a percentage of words at each level, and allows texts to be compared to each other, or to learners’ vocabulary test scores, in several useful ways.

2  LFP research and practice

A promising early research use of LFP was to automatically assess second language (L2) learners’ written productions. The deployment of more words from less frequent frequency bands was shown to successfully predict expert human ratings with a high degree of consistency (Laufer & Nation, 1995). Since then, researchers and practitioners at the reading-vocabulary interface have used LFP for many purposes in instructed language learning (diagnostic, testing, text selection, awareness raising, and others), with practical uses probably receiving more attention than further research.

3  Varying the grain size

Attempting to carry the research side forward on frequency profiling, Crossley, Cobb and McNamara (2013) experimented with improving its predictive ability still further by assigning words not to 1,000-family bands but to averaged individual word-form frequencies. For example, cat and cats are both first 1,000 items in most LFP schemes but in fact have different individual corpus frequencies which are not employed in a band approach. This research found that individual word frequencies can predict human judgments of learner writing even more successfully than band-based LFP. The problem however is that greater precision is achieved at the expense of most or all of the practical value of LFP. It is useful for a learner to know that their receptive or productive vocabulary knowledge seems to fall at a certain 1,000 level; there are many other words at this level, and lists or filtering software for deploying these lists in natural texts are available. It is less useful to know that words appearing fewer than x,000 times in a corpus seem to be missing from their compositions.

4  A way down the middle?

The goal of the present research is to explore ways of combining the greater precision of the finer grain-size with some of the functionality that so many practitioners have found for the original LFP. A first guess at a grain size that might achieve a balance is 100 word families. The five highest frequency 1000-family LFP bands, from the most recent LFP lists Nation’s BNC-Coca, were assigned CELEX frequency values and carved into fifty 100-family bands that might arguably combine the advantages of both finer and broader measures.

5  More practical

The 100-family unit may have even more practical utility that the 1,000 in some typical cases. For example it is extremely difficult to make flashcards for 1,000 words; 100 word sets are doable and usable. In testing, the sampling ratios from a 1,000-family set of words has always been a problem (18 out of 1,000 families = 1.8% of items sampled in Nation and Beglar’s Vocabulary Size Test, but 18 from 100 is 18%). Some “C-level” deployments for smartphone from Lextutor will be briefly exhibited, time permitting.

6  More accurate?

But can the deployment of 100-family bands approach the superior prediction of the point-frequency measure? A set of 400 spelling normalized learner productions at four original levels, as established by expert raters, were re-categorized in two ways: first by proportion of post-1,000 family items, and second by the average individual word frequencies of all the words in each text. Which method put the texts back where
they came from?

References


Toward the Integration of a Wordlist with a Common Framework of English: The Case of the CEFR-J

Yukio Tono
Tokyo University of Foreign Studies
y.tono -at- tufs.ac.jp

1 Introduction

Wordlists are indispensable resources not only for foreign language teachers and learners but also syllabus designers and test and teaching materials developers. Several different approaches have been used to develop wordlists for educational purposes. According to Chujo (2015), wordlists are classified based on the following design criteria: (a) word frequency and dispersion, (b) composites of multiple wordlists, (c) other pedagogical factors such as cost, necessity, coverage, stylistic level, function, and intuitions of EFL teachers (See Chujo (2015) for a list of wordlists for each classification). Despite the availability of corpus-based wordlists, it was not until the release of the British National Corpus (BNC) (1995) that pedagogical wordlists became truly corpus-based or corpus-informed (e.g. SVL12000; JACET8000; New GSL, etc.).

As there is no perfect wordlist to meet everyone’s needs, each wordlist has its own shortcomings. Nation (2004), for example, argued that the BNC 3000 word family list and the GSL-AWL combined list show the characteristics of “adults, British” and “young learners, US” respectively, which is a good example showing that the source data could influence the nature of the list. Tono (2013a) also shows that if the list is based solely on a single corpus such as the BNC, the frequency list tends to be influenced by sampling biases and that in the frequency band lower than 3,000 lemmas are not very reliable and their ranking will vary within the span of 1,000 to 1,500 words if a word list is created based on the repeated sampling of the same number of texts as the BNC. Besides, the wordlists are often based on different assumptions. For example, frequency-based wordlists tend to show how likely L2 learners are to encounter words in the list in real life situations, but this does not always correspond to those words which are likely to be introduced in the classroom (Chujo & Oghigian, 2009). If we pay too much attention to the coverage of native speaker texts, the number of words in the wordlist naturally increases, and thus we lose sight of the potential benefits of limiting vocabulary to a manageable size. Also there is a growing interest in going beyond a mere wordlist.

A list of formulas or phrasal expressions (Martinez & Schmitt 2012; Simpson-Vlach & Ellis 2010) is a good example showing that in the frequency band lower than 3,000 lemmas are not very reliable and their ranking will vary within the span of 1,000 to 1,500 words if a word list is created based on the repeated sampling of the same number of texts as the BNC. Besides, the wordlists are often based on different assumptions. For example, frequency-based wordlists tend to show how likely L2 learners are to encounter words in the list in real life situations, but this does not always correspond to those words which are likely to be introduced in the classroom (Chujo & Oghigian, 2009). If we pay too much attention to the coverage of native speaker texts, the number of words in the wordlist naturally increases, and thus we lose sight of the potential benefits of limiting vocabulary to a manageable size. Also there is a growing interest in going beyond a mere wordlist.

It is about time for us to consider how a wordlist can be integrated into a core language teaching syllabus in a meaningful way. In this paper, I will report on the project of the CEFR-J, in which vocabulary is integrated into a common framework of English in terms of CEFR levels, corresponding illustrative descriptors, and language functions in such a way that both teachers and learners can appreciate how productive and receptive vocabulary will develop through different stages of learning.

2 What is the CEFR-J?

The CEFR-J is an adapted version of the CEFR, designed especially for Japanese learners of English. Based upon a large-scale survey of the learning objectives of English learners in different educational sectors in Japan and other countries, we decided to launch a project to create a framework of English based on the CEFR and more finely tuned to learners of English in our country. The CEFR-J consists of a series of can-do descriptors according to a scale with further branching and additions to the original CEFR (pre-A1, A1.1-A1.3, A2.1-2.2, B1.1-B1.2, B2.1-B2.2, C1, C2). All the descriptors were scaled following a similar calibration procedure to the original CEFR. The first version was released in 2012 (Tono, 2013) and has been widely used for placement purposes, materials development, syllabus construction at prefectural or school levels, and test development.

3 The CEFR-J Wordlist

Various accompanying resources have been developed to support the users of the CEFR-J. One of such efforts is the CEFR-J Wordlist. It is based on the analysis of CEFR-level classified texts originally taken from textbooks used in China, Korea and Taiwan. The original aim of the text analysis was to analyze how vocabulary was introduced at the primary school level in EFL environments similar to Japan. The results show that the size of the vocabulary commonly taught at primary school level across the three regions above
is approximately 1,000 words, which is almost equivalent to the vocabulary of A1-level words defined by English Vocabulary Profile (EVP). We compiled a wordlist by merging the results of our corpus analysis with the EVP. This wordlist is now freely available for research as well as commercial purposes.

4 Corpus-based grammar and text profiling for the CEFR levels

It is also important to identify criterial grammar and text characteristics for CEFR levels in order to introduce vocabulary in relation to grammar and text at different CEFR levels. To this end, we compiled a corpus of CEFR textbooks published in the UK (96 titles, c. 1.64 million running words) tagged for POS, lemma and sub-sections of 4 skills, and 493 grammatical items were extracted from the corpus. A Support Vector Machine (SVM) was used over the matrix of [levels] x [grammar] to determine the attribute weights of grammatical items that are crucial in classifying texts into a particular CEFR level. In this way, grammatical items which are criterial for a particular CEFR level are selected.

In the same way, text characteristics (fluency and complexity measures) were extracted from the corpus and SVM or random forests were used to determine important text characteristics for classification purposes. These two results were summarized into the CEFR-J Grammar Profile and Text Profile respectively.

5 Linking vocabulary to the profile information

Currently we are working on the linking between the CEFR-J Wordlist, the Grammar Profile and our can-do descriptors. Since a can-do descriptor is a mere statement of what you can do with language, in order to perform the given action specified in a certain can-do descriptor, we need to present learners a set of vocabulary and grammar to achieve the goal. For productive can-do’s, the conditions under which an action specified in a particular descriptor is taken are further defined in terms of learner’s CEFR levels, relevant situations and topics. By properly controlling these factors influencing the performance, we can define a set of words and grammar to go with a particular can-do descriptor, which is quite useful when designing language tasks or textbooks.

For receptive can-do’s, the Text Profile information will be linked to the Wordlist. There are many words above 2,000 lemmas, which may not be used as active vocabulary for many learners. In this case, all the words are first introduced in reading materials with controlled vocabulary and recycled in subsequent lessons.

6 Conclusion

The CEFR-J Grammar/Text Profiles can successfully integrate vocabulary into a set of grammar or expressions linked to can-do descriptors at a particular CEFR level. This comprehensive framework linking can-do descriptors with vocabulary, grammar, and text will serve as a useful resource for designing curriculums, textbooks, classroom tasks and task-based tests. Here, vocabulary is not a mere static list, but is dynamic, integrated into a set of coherent frameworks, which gives learners a clear image of what words should be learned for what purposes and in what order. By classifying vocabulary according to can-do descriptors, we can distinguish a group of words which is more closely linked to productive knowledge than receptive knowledge. Our goal is to complete this set of resources as far as the B1 levels and provide it for the next revision of the Course of Study by the MEXT.

References


Beyond Caveat Emptor: Applying Validity Criteria to Word Lists

Diane Schmitt
Nottingham Trent University
diane.schmitt -at- ntu.ac.uk

1 Introduction

Both West’s General Service List (1953) and Coxhead’s (2000) Academic Word List are much loved work horses for the fields of Applied Linguistics and TESOL as both have informed curricula, teaching and learning materials, tests, dictionary development, text profiling software and other aspects of language description. Yet, in recent years, the field has been bombarded with a proliferation of new general and specific purposes word lists (e.g. Brezina and Gablasova, 2013; Gardner and Davies, 2013; Mungra and Canziani, 2013; Hsu, 2014; Lei and Liu, 2016) seeking to take their place, to the extent that word list construction appears to have become something of a cottage industry. Counter to definitions of such industries in economics which focus on their labour intensiveness (e.g. Investopedia, n.d.), the growth in the number of word lists and word list developers has been supported by advances in technology which both streamline the construction of word lists and facilitate their dissemination. While it seems obvious that we should hail these developments as great leaps forward for our field, with the growing number of word lists now on the market, it may be time to pause and consider what it is we want from lists and what the field needs from their compilers.

2 Do we really need more lists?

A number of researchers have offered critiques of existing lists in terms of age, size of and/or methods of compiling the corpora, and the claims made about text coverage, learnability, and usefulness (e.g. Neufeld, Hanciog’lu, and Eldridge, 2011; Hyland and Tse, 2007; Martinez, 2009). Newer list compilers have attempted to address some of these critiques by taking advantage of improvements in corpus methodology to identify lists of words that better reflect contemporary English and changes in thinking about the unit of counting (e.g. Brezina and Gablasova, 2013; Gardner and Davies, 2013). Others have taken advantage of the greater ease of compiling and analysing corpora to create more targeted specialist lists (Hsu, 2015). Yet others seem to rather uncritically compile new lists simply because the opportunity to do so is available (e.g. Lei and Liu, 2016, Yang, 2015).

3 Why should we think more carefully about word list development and dissemination?

For word lists to be relevant to users, word list compilers need to make a number of decisions about, for example, what types of texts to include in the corpora from which the list will be drawn, the unit of counting, and the relative value of word frequency, range, and dispersion. Although most new lists are purportedly developed to assist curriculum developers, materials writers, teachers and students, it is surprising how little reference is made to the needs of these groups when word list compilers report on the decisions taken when developing their lists. It is common to see uncritical replications of previous lists without clear justification provided for why the decisions taken by earlier lists developers are appropriate for the development of a new list for the stated user groups. The result can be overstated claims about usefulness, and confusion for potential users and fellow researchers. Usefulness is not an inherent trait transferable from context to context. When disseminating new lists, list makers need to provide stronger arguments for the relevance of their lists for particular contexts and provide more explicit guidance on list use for potential users. In sum, just as the field of language testing has begun to require that test developers provide validity arguments for test use, vocabulary studies would benefit from similar validity arguments for word lists.

References


Hsu, W. (2014). Measuring the vocabulary load of engineering textbooks for EFL undergraduates. English for Specific...


1 Unanswered questions in corpus analysis (Laurence Anthony)

Vocabulary researchers seem to have an almost love-hate relationship with corpora and corpus methods. Clearly, vocabulary researchers to date have shown a ‘love’ of corpora, using them to calculate the frequencies of general, academic, and technical vocabulary for use in the construction and testing of vocabulary size and levels lists. Some vocabulary researchers have even gone so far as to construct vocabulary lists and tests based solely on the frequencies of items in large corpora, regardless of whether or not these lists and tests make pedagogic sense. On the other hand, there is clearly less interest among vocabulary researchers in other aspects of corpus analysis. For example, few vocabulary researchers seem interested in constructing new, interesting, and useful corpora, or probing corpora beyond surface-level forms in order to gain insights on the fundamental concept of "words", their relationships with other "words", or investigating the lexico-grammatical relationships that exist in language. Do these trends reflect a 'distrust' of corpora or perhaps a 'hate' of the statistical methods on which much corpus analysis is based?

In recent years, an increasing number of vocabulary researchers have been investigating word combinations in the form of n-grams, collocations, and all-encompassing multi-word units (MWU). We also see more rigorous statistical methods used to validate the results of vocabulary work. Over the next 10 years, we can predict that these trends will continue and we will begin to use corpora to answer deeper questions on the nature of vocabulary, including:

- How should we define and identify vocabulary items and combinations?
- How stable are general, academic, and technical vocabulary items and combinations in and across specialist disciplines?
- How can we measure and visualize the development of a learner's vocabulary depth and breadth over time?
- How can we measure and visualize the relationships between vocabulary items and combinations in ways that reveal their interdependence?

2 Unanswered questions in L2 vocabulary assessment (Norbert Schmitt)

The area of vocabulary assessment had remained relatively stagnant since Paul Nation first published the original *Vocabulary Levels Test* in 1990 and Paul Meara developed various versions of Yes-No tests around the same period. True, a number of new tests have appeared, including new versions of the *VLT* (e.g. Schmitt, Schmitt, & Clapham, 2001; McLean & Kramer, 2015), and versions of the Yes-No test format (e.g. the *DIALANG* battery). There have also been a number of new frequency-based multiple-choice tests (e.g. the *Vocabulary Size Test*, Nation & Beglar, 2007). Likewise, tests of vocabulary depth have mainly been measured by various versions of the *Word Associates Format* (Read, 1998). Most of these new tests have been developed based on the templates of older tests, without much consideration of whether the assumptions underlying those templates are still tenable, and whether those templates are still workable.

Very recently, there has been a discernible move towards challenging the existing assumptions, and thinking about what kind of shape the next-generation vocabulary tests should take. Notable pieces of this new trend include:

- Taking advantage of computer technology in terms of delivery (tests on *Lextutor*) and computer adaptive testing (e.g. the *CATSS*, Laufer)
- Questioning how frequency should be used in developing size tests (Kremmel, in press)
- Testing both form recall and meaning recall to get a better indication of the quality of lexical knowledge
- Questioning what multiple-choice items can tell us about learners’ ability to use vocabulary

Many new vocabulary tests are being disseminated without proper (if any) validation being done, or any indication of how the scores should be interpreted. But once they are in the public domain, most end users
(e.g. teachers) will use them assuming they must be valid as they are publicly-available. Perhaps the greatest need in vocabulary assessment is create a mindset in the field that tests not be released to the public until a reasonable range of validation evidence has been collected and the test developers are confident they can describe what the resulting test scores mean. The days of a one-size-fits-all vocabulary test are probably over, as tests need to be developed with a particular purpose in mind (which then allows validation research to carried out).

3 Unanswered questions in L2 vocabulary teaching and learning (Paul Nation)

I take a very broad view of vocabulary learning which includes opportunities for incidental learning through reading and listening, and through speaking and writing, and deliberate learning. An early piece of research in this area was a qualitative study by Hosenfeld (1976) which involved the examination of sentence-based grammar activities. Hosenfeld found that when learners did the activities, they were actually defeating the learning purpose behind the activities.

There is some research on a few activities. Frank Boers’ (2013) research on 4/3/2 for example while not focusing particularly on vocabulary provides a very useful examination of this important fluency development technique. Wilkinson’s study of making word cards is another example. Newton’s (2014) study of vocabulary learning through speaking tasks nicely combined receptive and productive learning. My own research (Nation, 2014) on the opportunities for repetition through extensive reading is a corpus-based example.

The most thoroughly researched area is that of vocabulary learning using flash cards which also includes learning using the keyword mnemonic. The fruitfulness of such research can be seen in the detailed guidelines it is now possible to provide for such learning (see Nakata, 2011), and in the important questions that still remain unanswered about this technique (Ishii, 2013; Elgort, 2011). The goal of such research would be to allow informed choice between various vocabulary learning activities and to show how such activities can be adapted and used to produce the best learning.

<table>
<thead>
<tr>
<th>Partly Researched Activities</th>
<th>Largely Un-researched Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keyword technique</td>
<td>Extensive listening</td>
</tr>
<tr>
<td>Word card learning (paired associates)</td>
<td>Speed reading</td>
</tr>
<tr>
<td>Extensive reading</td>
<td>Linked skills</td>
</tr>
<tr>
<td>Glossing and dictionary look-up</td>
<td>Information transfer</td>
</tr>
<tr>
<td>Group speaking activities</td>
<td>Prepared talks</td>
</tr>
<tr>
<td></td>
<td>Listening to stories</td>
</tr>
<tr>
<td></td>
<td>10-minute writing</td>
</tr>
<tr>
<td></td>
<td>4/3/2</td>
</tr>
<tr>
<td></td>
<td>Substitution tables</td>
</tr>
<tr>
<td></td>
<td>Writing with feedback</td>
</tr>
<tr>
<td></td>
<td>Dictation</td>
</tr>
<tr>
<td></td>
<td>Intensive reading</td>
</tr>
</tbody>
</table>

Table 1: Partly Researched and Unresearched Vocabulary Learning Activities

The items in Table 1 have been ranked, in column 1 according to the amount of vocabulary-focused research done (most researched at the top), and in column 2 according to the need for vocabulary-focused research (the most needed at the top). Most of the activities in Table 1 are described in detail in Webb & Nation (2016 in press). In that book we looked at the kinds of mental conditions that could occur in particular vocabulary learning activities. This kind of examination, as is carried out in involvement load analysis and technique feature analysis however, is one step away from a more empirical study. It is important that this next step is taken because it is only through such examination that we find out what teaching techniques are truly doing.
Charting the Vocabulary of a Minoritised language: Challenges and Opportunities in the Creation and Application of the National Corpus of Contemporary Welsh

<table>
<thead>
<tr>
<th>Dawn Knight</th>
<th>Steve Morris</th>
<th>Tess Fitzpatrick</th>
<th>Laurence Anthony</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiff University</td>
<td>Swansea University</td>
<td>Cardiff University</td>
<td>Waseda University</td>
</tr>
<tr>
<td>KnightD5</td>
<td>s.morris</td>
<td>FitzpatrickT</td>
<td>anthony -at- waseda.jp</td>
</tr>
<tr>
<td>-at- cardiff.ac.uk</td>
<td>-at- swansea.ac.uk</td>
<td>-at- cardiff.ac.uk</td>
<td></td>
</tr>
</tbody>
</table>

1 Introduction to the CorCenCC corpus

The National Corpus of Contemporary Welsh (Corpws Cenedlaethol Cymraeg Cyfoes –CorCenCC) is a major new language resource being created to represent spoken, written, and electronically-mediated Welsh. The corpus has a unique user-centered design, with language data collected and validated through crowdsourcing, and an in-built pedagogic toolkit developed in consultation with representatives of all anticipated academic and community user groups.

CorCenCC will break new ground both as a language resource and as a model of corpus construction. It will be the first large-scale, representative corpus of Welsh language use across communication types (circa 4m spoken words, 4m written, 2m e-language), genres, language varieties (regional and social) and contexts. CorCenCC will forge transformative methods for corpus creation, impact and sustainability. It is community-driven, harnessing opportunities afforded by mobile technologies, specifically crowdsourcing and community collaboration, allowing users to collect and upload their language data, complete with facilities for accurately and appropriately tagging/annotating data in terms of its source, type, topic, date of collection, and other relevant metadata categories. CorCenCC will be open-source and freely available for use by professional communities and anyone with an interest in language. The corpus will enable, for example, community users to investigate dialect variation or idiosyncrasies of their own language use, professional users to profile texts for readability or develop digital language tools, researchers to investigate patterns of language use and change, and learners to study from real life models of Welsh.

2 Social context and preliminary findings from e-language corpus data.

In minority language settings there is often extensive community interest and activity surrounding and ‘growing’ the language, and this presents specific opportunities (and challenges) for applied and corpus linguists. The CorCenCC project includes novel design features that will capitalize on this community buy-in and engagement in the Welsh language context throughout the project: community users will be involved in the content, tagging, utilities and application of the corpus.

In Wales, where 20% of the population are users of Welsh, references to language use, skills, and awareness permeate the discourse of policy, education, and society explicitly and deliberately. This is a rich environment in which to introduce a resource that focuses on language description rather than prescription. It also means that talking about language, as well as using language to talk, is a feature of Welsh speakers’ repertoire. A Welsh Government call for increased digital media content in Welsh exemplifies this dual focus on language promotion and language use: “the existence of Welsh-language digital media content and applications not only allows the Welsh language to flourish, but it also enables Welsh speakers to participate fully as digital citizens and demonstrates to all that the Welsh language is a creative, powerful, adaptive and modern language” (A Living Language, 2012:5)

This perception of digital media in the Welsh context prompted us to create a 500,000-word pilot corpus of electronically-mediated Welsh, as a proof of concept for the CorCenCC project. Permissions were obtained from all contributors, and their prompt and positive responses supported our vision of corpus creation as a community enterprise in the Welsh context. We analysed the corpus data in a number of ways, including by word frequency and by word cluster. Online communication is often cited as a medium in which instances of code-switching and lexical borrowing are more frequent than in other genres (Nguyen and Dogruoz 2013), and therefore can be viewed warily by linguistic purists; a corpus of e-language therefore represented an opportunity to demonstrate ways in which corpus data can inform prescriptive/descriptive debates. The pilot did indeed reveal frequent use of lexical borrowings, but also highlighted that word use in a ‘modern’ context, via a ‘modern’ medium, can sometimes mask the fact that a lexical item has been embedded in the language for centuries. Examples in our corpus included ‘stwff’ (=stuff), which first appeared in Welsh in the 15th century, and ‘cyri twrci’ (=turkey curry) – ‘twrci’ first appeared in Welsh around 1600. The pilot corpus, then, immediately makes an evidence-based contribution to debates on language change and use, and enables close scrutiny of the “creativity, power and flexibility” of the language, referred to in the Living Language document.
CorCenCC is designed to serve a useful resource in the teaching of the Welsh language in schools, colleges, and adult learning centers. To facilitate this learning, the corpus will be released with a set of online pedagogic tools and learning activities. As the corpus contains spoken and spoken data from various real-life contexts, these tools and resources provide many opportunities for teachers to introduce learners to practical, authentic, modern Welsh language. For vocabulary teaching and learning, the pedagogic toolkit will first feature standard vocabulary profiling functions similar to those in Compleat Lexical Tutor (Cobb, 2016) and AntWordProfiler (Anthony, 2014). Using these functions, learners will be able to quickly generate lists of the most frequent Welsh vocabulary in particular settings, such as daily conversation or television discourse. They will also be able to compare the frequencies of vocabulary items in one setting against the frequencies in the corpus as a whole to reveal contextual ‘keywords’.

One of the main pedagogic features of CorCenCC is a set of ‘language awareness raising’ tools that allow learners to interact with CorCenCC and see how vocabulary varies with context and speaker. Using the concept of Key-Word-In-Context (KWIC) searches, teachers can guide learners to navigate through the corpus to find common vocabulary patterns, collocations, and multi-word phrases in different settings, and then help them apply what they have learned through various learning activities, also included in the pedagogic toolkit. In addition, the toolkit offers learners a way to measure they vocabulary development through a vocabulary level test. The results of this test can be immediately applied in the selection of suitable materials within the CorCenCC online system, and also beyond, for example, in the choice of materials and textbooks for standard teacher-led classrooms.

Results and applications of CorCenCC in the Welsh local and national context

The Welsh language holds a unique position within the education system in Wales (which is a devolved responsibility within the UK) from nursery level up to Higher Education. This includes post-compulsory education language acquisition provision in the community. Within the Welsh education system, the language is taught both as a first language as well as a second language to students. In the Welsh context (in a similar way to other European minoritised language contexts, for example Basque or Catalan), the pedagogical landscape overlaps with other policy areas as well, in particular language planning (both status and corpus planning). CorCenCC will inform development in these policy areas and has the potential to bring about change here too. CorCenCC - as a major new resource in the context of a minoritised language like Welsh – will impact on the larger linguistic landscape in particular in such areas as:

- Lexicography – CorCenCC will not only give evidence of contemporary usage, neologisms, frequency and anacronisms, but also will allow lexicographers in Welsh to compile dictionaries which give real life examples of how contemporary Welsh is used.
- The translation industry – detailed information on frequency of words in Welsh will allow translators to target their work at appropriate registers.
- Mass media (in particular the BBC and S4C) – these areas also need to be sure that they are engaging with their viewers and listeners at a linguistically appropriate level and CorCenCC will facilitate this both in broadcast materials as well as on-line resources.
- The publishing industry – Welsh language journals and publishing companies will be able to ensure that their materials are written for an appropriate reading age based on real- life language evidence rather than supposition and assumption.

References

Vocabulary Pedagogy and Teaching Materials: Applying the Research in Japanese Higher Education

Paul Wadden  
International Christian Univ.  
paulwadden -at- gmail.com

Paul Nation  
Victoria Univ. of Wellington  
paul.nation -at- vuw.ac.nz

Phil Bennett  
Miyazaki International Col.  
philbennett59 -at- gmail.com

Tim Stoekel  
Univ. of Niigata Prefecture  
timstoeckel -at- gmail.com

Tamio Okumura  
Goken Publishing Inc.  
okmr_t -at- mac.com

Guy Cihi  
Lexxica  
ecihi -at- lexxica.co.jp

1 Overview—Wadden

Since Xue and Nation published the UWL (University Word List) in 1984, research in vocabulary has grown exponentially as linguists have examined vocabulary acquisition, categorization, assessment, and corpus construction. Among the corpus-based lists that have further enriched the resources for vocabulary learning are the EAP (English for Academic Purposes) Word List (Masuko et al., 1997), AWL (Academic Word List) (Coxhead, 2000), GAV (Global Academic Vocabulary) lexicon (Wadden, 2013), NGSL (New General Service List) and NAWL (New Academic Word List) (Browne et al., 2013). Unfortunately, vocabulary pedagogy and learning materials have scarcely kept pace with research and resources. To partly address this deficit within the Japanese university setting, this colloquium considers corpus-based vocabulary in the Japanese context (Okumura, Wadden, Cihi); outlines a comprehensive “four strands” approach (Nation); describes assessment tools to evaluate lexical growth, program efficacy, and curricular construction (Bennett & Stoekel); identifies common vocabulary learning styles of higher level Japanese learners (Okumura); and presents two corpus-based resources (Wadden, Cihi) which can help Japanese university students and advanced learners build a strong foundation in academic vocabulary.

2 Using the Four Strands for Curriculum and Pedagogy—Nation

A well-balanced language course should consist of four equal strands—meaning-focused input, meaning-focused output, language-focused learning, and fluency development—with each strand receiving a roughly equal amount of time across skill areas in a course or curriculum.

For teachers, the four strands can organize the roles, techniques, relative time, and range of activities for each of the four skills in a course. Critical to applying the four strands principle is the criteria an activity must meet to be included in a particular strand, such as meaning-focused input providing comprehensible material, a low density of unfamiliar language features, and focus on the message; meaning-focused output providing material largely but not completely familiar and focus on message; language-focused learning providing conscious focus on language features; and fluency development providing easy material, pressure to go faster, and focus on message. The four strands also provide a basis for autonomous vocabulary-learning activities such as use of notecards, dictionaries, graded readers, podcasts, and email exchanges.

3 Using Assessment to Inform Curriculum —Bennett & Stoekel

A principled use of vocabulary assessment is vital to monitor lexical growth, assess program efficacy, and inform curricular revision (Nation, 2001). Two institutional settings in Japan offer demonstrative examples of such use: Miyazaki International College (MIC) and the University of Niigata Prefecture (UNP). In each setting, knowledge of high frequency and academic vocabulary was considered important for success in EMI (English Medium Instruction) classes in the third and fourth years. To assess knowledge of these words, a Test of General and Academic Vocabulary (Bennett & Stoekel, 2013) was developed and later converted into two separate instruments—the New General Service List Test (NGSLT) and New Academic Word List Test (NAWL Test) (Stoekel & Bennett, 2015) aligned with the NGSL for high frequency and NAWL for academic vocabulary.

At MIC, testing showed that many students entered college with sizeable gaps in high frequency and academic vocabulary, and had relatively small gains, averaging 120 words per semester, during the first three semesters. This information supported a proposal to add a dedicated reading and vocabulary course in each of the first three semesters, and it guided the development of lexically appropriate learning materials for
these courses. Continued assessment indicates a 33% increase in the uptake of high frequency and academic vocabulary under the new program. Moreover, first-year TOEIC gains have increased approximately 70%.

At UNP, initial assessment shows that learners in a third- and a fourth-year EMI course began with good knowledge of NGSL words but considerable deficiencies in NAWL terms. Unfortunately, no measurable NAWL gains were made over the semester, which suggests a need for continued intentional study of targeted vocabulary in EMI programs, where the focus of learning tends to be on content rather than language.

The NGSLT and NAWLT are available in monolingual and Japanese-English bilingual formats from http://www.newgeneralservicelist.org/ngsl-levels-test/.

### 4 Lexical Background and Learning Styles—Okumura & Wadden

From the late 1950s to the early 21st century, the Ministry of Education suggested that middle schools teach 900-1200 words (including 100-520 words the Ministry specifically selected) and high schools teach an additional 1200-1800 words. College-prep institutes and textbook publishers typically added further vocabulary for entrance exam study, such as the additional 1,850 common vocabulary Kenkyusha combined with the two lists in its dictionary of the “4,800 vocabulary that high school students should know” (Kosakai, 1967, 1979, 1980). With the adoption of *yutori kyoiku* in secondary schools (“less stressful education” or “education that gives students room to grow”), class days, content, and coursework have been reduced and there has been no nation-wide vocabulary benchmark at any level. Instead, individual publishers, schools, and teachers compile and teach English vocabulary according to their best judgment. This era may now be ending. The Ministry of Education announced on June 21 that it plans to set a target of 4,000-5,000 words for secondary schools. It is unclear, as yet, whether the Ministry will create the lists itself. In the past, the Ministry published lists of 100-520 words as the minimum vocabulary for junior high school students and left it to textbook publishers and teachers to decide what additional vocabulary to teach and in what order.

The Ministry has no recommended word list for Japanese higher education, although some Japanese college teachers draw upon the corpus-based JACET 8000. To fill this gap, over the past ten years Goken has published TOEFL vocabulary books for university students, particularly those with the added motivation of test preparation or plans for overseas study. The first, TOEFL TEST 究極単語5000 [TOEFL 5,000 Essential Vocabulary], utilizes bands of the Goken Word List (GWL) which has 32,000 word entries and 24,000 word families drawn from corpora (3.1 million tokens) of TOEFL and TOEIC tests, TIME magazine articles, VOA and FEN radio news, junior high and high school textbooks, and other sources. The second book, TOEFL TEST 究極単語 ADVANCED 3000 [TOEFL Advanced 3,000 Specialized Vocabulary] offers students disciplinary-specific, high frequency words in 44 academic fields.

Many Japanese students have the ability—and the preference in exam preparation—for rapid learning and rote memorization of vocabulary. Yet retention of this vocabulary is short-lived unless reinforced by active use, challenging courses, or overseas study. In fact, if such students learn 3,000-5,000 words principally for university exam prep, they tend to suffer language attrition and vocabulary loss during college. At present, some university English teachers report spending a whole year reviewing vocabulary and grammar that students were already expected to have learned in junior high and high school.

### 5 Tapping the Global Academic Vocabulary (GAV) Lexicon—Wadden

Nation and Waring observe that a vocabulary size of around 2,000 words enables a learner to know only about 80% of the vocabulary in an academic text, a ratio insufficient for successful comprehension. They note, “At least 95% coverage is needed for that” (1997). Each of the major corpus-based word lists to date (UWL with 836 headwords, EAP with 874, AWL 570, NAWL 963) can crucially boost learners’ knowledge of academic vocabulary, but with each having fewer than 1,000 headwords, none by itself is likely to result in learners reaching the critical 95% threshold. To increase the chance of achieving this comprehension level, the GAV combines the headwords from all four lists in 23 lessons comprising 1,850 words and generally progressing from most to least frequent. The pedagogical premise is efficacy: to lower as much as possible the “learning burden” (Nation, 2006, p. 70). The GAV Lessons, Quizlet cards, and Learner’s Dictionary (1) identify the headword (e.g., the verb “access”) as well as important related word forms (e.g., the noun “access,” the adjective “accessible”), (2) provide the primary meaning of headwords and related words in simple English and bilingually in Japanese, (3) draw attention to common phrases the words appear in (such as “access the internet”), and (4) demonstrate their “use in context” in sample sentences. The quizzes for each lesson can be used autonomously or in the classroom to motivate students to study the words, assess learning, and give students feedback on their progress. The GAV lessons, tests, Quizlet cards, and Learner’s Dictionary have been used on Moodle, Schoology, Edmodo, and Quizlet platforms.
6 WordEngine: Principles and Premises—Cihi

Lexxica’s vocabulary testing and learning applications are informed by the best available research with the specific goal of providing learners the highest possible outcome relative to time-on-task. Of great importance are the collective experiences and outcomes of students and teachers using the applications. While there are many examples where the research saved Lexxica and its developers time and money, perhaps more instructive in the context of Vocab -at- Tokyo are these four examples where user feedback and data analysis overrode common approaches noted in the research. 1) To be pedagogically effective, digital flashcards should display high quality distractors, not only words and definitions. 2) Item difficulty and item frequency quanta are not meaningfully correlated (Japan Patent No. 4908495, 2012), as Cihi, Culligan, and Browne have found. For example, Figure 1 Vocabulary Gaps depicts a Japanese learner who knows 3,908 total words yet is still missing 541 words from the first three 1,000-word frequency bands. Regardless of the frequencies, those 541 items are above the learner’s ability. 3) With regard to counting words and calculating domain coverage, Rasch Analysis, with millions of item responses, reveals inflected word forms have the same statistical difficulties as their base words and therefore can be counted together as one item; however, derived word forms, phrasal verbs, and idiomatic polywords have different statistical difficulties and logically should be counted as separate base words. 4) Distractors are as impactful on test scores and learning outcomes as the items, translations, and definitions themselves. High quality distractors are of the same POS and from a higher difficulty level, with meanings and spellings that are plausible but wrong, and regularly rotated to limit fossilization. There are typically four to five times as many distractors as items, therefore, the onus on developers is much greater than generally imagined.

Figure 1. Vocabulary Gaps.

References


Revisiting the Word Family: What is an Appropriate Lexical Unit for Japanese EFL Learners?

Stuart McLean
Osaka Jogakuin University
Kansai University
stuart93 -at- me.com

Paul Nation
Victoria University of Wellington
Paul.nation -at- vuw.ac.nz

Geoffrey G. Pinchbeck
Werklund School of Education, University of Calgary
geoffpin -at- me.com

1 Introduction

Word lists based on language corpora have been widely used for multiple purposes in applied linguistics research and language pedagogy, and they have played an essential role in language testing, language curricula development, estimations of vocabulary size, and estimations of text difficulty. The choice of the lexical unit, where inflected and derived forms are included with a word’s base form, has ramifications for the validity of corpora-derived word lists, word-list-based vocabulary tests, and research based on these instruments (Gardner, 2007). This colloquium includes two studies which answer calls for research “on the actual performance of second-language learners when they encounter members of a word family” (Macalister & Webb, 2013, p. 853), and research which introduces a counting unit for Asian EFL learners.

2 Study one (McLean, S.)

Japanese university students (N = 279) were separated into three groups, 1) 1K mastery 2) 1K-3K mastery and 3) 1k-4/5K mastery based on their New Vocabulary Levels Test scores (McLean & Kramer, 2015). All students completed a form-recall and then meaning-recall instrument testing knowledge of 12 base forms and various inflected and derived forms (23 items consisting of multiple affixes). One Japanese native speaker marked all of the meaning-recall tests. Two additional native speakers of Japanese marked 20% of the meaning-recall tests. If the L1 Japanese meaning produced by participants demonstrated accurate knowledge of the target word including the meaning of the target form’s affix(es), markers marked the answer as correct. The study only analyzed data from inflected and derived items where participants first correctly recalled the meaning of the associated base form. Thus, if a participant failed to recall the meaning of publish, that participant’s responses to the base form, and inflected and derived members of the publish word-family, were excluded from the analysis. This step was taken as Bauer and Nation’s (1993) assumption is that once the base word or even a derived word is known, the recognition of other members of the family requires little or no extra effort. The interrater reliability values established among the three markers were all ≥ .91. The KR20 reliability yielded from the entire meaning-recall data set and that marked by the second and third markers was ≥ .93. This study provides evidence that the lemma is a more appropriate counting unit than the word family at level 6 (WF6) for all but the most able students. This brings into question the validity of using WF6 word lists and WF6 wordlist based tests with Japanese EFL students.

3 Study two (Pinchbeck, G.G., McLean, S., Brown, D., & Kramer, B.)

This study is part of a larger project examining how word lists for learners of English might be optimized for research and pedagogical purposes. Here, we investigated two aspects of word list development for use in Japanese post-secondary contexts: (1) the choice of reference corpus (or corpora) and (2) the definition of word. Word lists used for applied linguistics and language pedagogy have been based primarily on word frequency/distribution of large corpora that have been compiled purposefully to ‘balance’ various registers, genres and modalities. Biber (1993) introduced the question of how representative such corpora are of real world language, and this issue continues to be germane to word list development because it is not clear, a priori, how and in what proportions frequency/dispersion data from different sub-corpora should contribute to word ranks in any given word list. Word ranks will vary considerably depending on how data from corpora that were derived from different genres/registers, modalities (e.g. spoken vs. written), and/or geographic origins (e.g. UK vs. US) are balanced.
The second issue of how word should be best defined has been previously investigated empirically and theoretically, but the word family (i.e., inclusive of levels 1-6 in Bauer and Nation, 1993) continues to be widely used in word-list-based applications for a variety of purposes and in diverse contexts. This topic will be covered in more detail in the third presentation in this colloquium, but it is important to note that ranks for words in word lists also vary significantly depending on how word is defined. We therefore explored these two issues as they pertain to word list design specifically for English language learning in the Japanese university context.

In a modification of the methodology of Brysbaert and New (2009) who used lexical decision times to evaluate word lists, the present study employed the New Vocabulary Levels Test (McLean & Kramer, 2015) word item difficulty - the proportion of examinees that answered the test item correctly - as an external reference criterion upon which currently available word lists could be validated. Corpora-derived ranked word lists based on either lemma (roughly inclusive up to Bauer and Nation’s levels 1-2) or word family (levels 1-6) were first generated from frequency and dispersion (e.g., range, arf, Gries’ DP) data from: (1) major sub-corpora of large general corpora (e.g., COCA, BNC), (2) the corpora upon which both New General Service Lists were based (Brezina & Gablasova, 2013; Browne, Culligan, & Phillips, 2013), and (3) corpora based on TV and movie subtitles (Brysbaert & New, 2009; Van Heuven, et al., 2014). UK and US spelling forms were merged into the same word groupings in both the lemma and family based lists. We then compared corpora frequency/dispersion rank data (20,000+ items) with vocabulary test word item difficulty rank data (149 target words) from 625 students attending four Japanese universities. Using Spearman rank correlations and a Hotelling-Williams significance test of correlation differences, word lists using lemma word groupings resulted in significantly higher correlations with item difficulty rankings compared with those using word family groupings across all word lists tested. Furthermore, lists derived from corpora of spontaneous speech, TV/movies for younger viewers, and narrative written texts consistently showed higher correlations with test item difficulty ranks than those derived from scripted speech or any non-fiction written text genres. The size of the originating corpus and its geographic origin were less important than was the register, genre, and modality.

These results suggest that the prevailing trend to compile mega-corpora from easily available electronic written texts may not be ideal for the creation of either general service lists or for word lists that represent developmental scales of academic English lexis as they might be used in Japanese English language learning contexts. Furthermore, the results also support examining the derivational morphology knowledge of the learners prior to using word lists in English language research and pedagogy in Japan.

4 Study three (Nation, P.)

When deciding on a unit of counting (word type, lemma, word family, and level of inclusiveness of the word family), we need to look at the reasons why the counting is being done and who the lists will be used with. There is considerable debate about the choice between lemmas (or flemmas) and word families. The discussion regarding the unit of counting is often posed as lemmas versus word families (Gardner, 2007; Brown, 2010; Nation, 2012). Bauer and Nation’s (1993) levels however included types and lemmas as being part of the word family levels – Level 1 for word types, Level 2 for lemmas. The discussion could thus be more usefully focussed on what level of word family is suitable for particular learners and for particular purposes. This is not just playing with terminology. Choosing to count lemmas involves accepting the idea that lies behind word families, namely that words with the same stem may be seen as related to each other and thus the recognition and learning of a word which is morphologically related to a known word is likely to be substantially easier than dealing with a completely unrelated word. The current problem is that the existing lists of word families are at either end of the Bauer and Nation scale with nothing in between. That is, there are lists of types (Level 1), lemmas and flemmas (Level 2) which include only headwords and inflected forms as family members, and families (Level 6) which include headwords, inflected forms and a wide range of derived forms. With such a wide contrast (stem and eight inflections at Level 2, and 91 affixes at Level 6), it is tempting to see Level 2 and Level 6 as a binary opposition rather than points on the same shaky scale. Some researchers with an interest in native-speaker knowledge see Level 6 word families as not being inclusive enough (Brysbaert, Stevens, & Mandera, 2015 forthcoming). The major issue is what level of word family is suited to a particular group of learners.

Gardner and Davies (2014) criticize the use of word families in the making of the Academic Word List (Coxhead, 2000) for the following reasons: (1) All the members of a word family are not closely related in meaning, (2) Word families do not make part of speech distinctions, (3) Learners, including young native speakers, may not have knowledge of the word-building devices of English. Gardner and Davies conclude that using lemmas will solve many of these problems. However, these criticisms are not correct. (1) Bauer
The debate on the unit of counting needs to turn into an effort to make word families and various levels of inclusiveness and then find which ones are most appropriate for certain groups of learners.

References


Processing Specificity in Vocabulary Learning

Joe Barcroft  
Washington University in St. Louis  
barcroft -at- wustl.edu

Shusaku Kida  
Hiroshima University  
skida -at- hiroshima-u.ac.jp

Henrik Gyllstad  
Lund University  
henrik.gyllstad -at- englund.lu.se

1 Introduction

Central to the study of vocabulary learning are the following questions: To what extent is vocabulary knowledge multi-componential in nature? What are the key subcomponents of vocabulary knowledge? To what extent are these subcomponents dissociable from one another? How do specific types of tasks, such as semantic and structurally oriented tasks, affect learning of different subcomponents of vocabulary knowledge? The present set of research reviews and analyses focuses on empirical findings and theoretical developments related to these questions, all of which concern processing specificity, or the extent to which multiple types of processing are dissociable or independent from one another (cf. domain specificity). The reviews and analyses attend to three key areas: (a) intentional vocabulary learning (Barcroft), (b) incidental vocabulary learning (Kida), and (c) developing knowledge of collocation (Gyllstad).

2 Processing specificity in intentional vocabulary learning

Research on intentional second language (L2) vocabulary learning suggests that (a) the formal, meaning-oriented (semantic/conceptual), and mapping components of vocabulary learning are, at least, largely dissociable from one another and that (b) increases in processing for one of these three can necessitate decreases in processing for one or both of the other two provided that overall processing demands are sufficiently high. These findings are consistent with predictions of the type of processing – resource allocation model (TOPRA) (Barcroft, 2002), which posits that increases in one or more dissociable types of processing, such as increased processing for the semantic/conceptual component of word learning, can result in decreases others, such as decreases in processing for the formal and mapping components of word learning. Different measures of vocabulary learning gain reflect necessary trade-offs of this nature.

Many studies have provided evidence for TOPRA predictions, but the combined findings of four studies are particularly telling. First, Wong and Pyun (2012) found that requiring L2 learners to write target L2 words in original sentences led to markedly less productive L2 vocabulary learning (see picture, produce word) as compared to no sentence writing, a decrease that was even more pronounced when formal features of the L2 words were more novel, that is, when first-language (L1) English speakers learned L2 Korean as opposed to L2 French words. Second, Barcroft (2002) revealed a double dissociation between task type (semantic vs. structural) and language of free recall (new vs. known) for target L2 words that had been studied while performing semantic versus structural tasks. Third, Sommers and Barcroft (2013) found that while semantically oriented referent token variability (e.g., seeing six different pictures of the Spanish target word oso ‘bear’ instead of one picture each of six times) yielded better memory for items in L1, it also decreased productive L2 word learning (in this case in the spoken mode) in an incremental manner. Finally, Kida and Barcroft (submitted) demonstrated that increases in either semantic or structurally oriented processing can decrease processing of the mapping component of vocabulary learn when L2 learners asked to map known L2 word forms onto known L1-based meanings, such as when mapping the known L2 word foot as 足 (ashi) as in hand and foot onto the meaning conveyed by 支払う (shiharau) as in foot the bill.

The combined findings of these studies are consistent with the dissociability of the formal, semantic, and mapping components of vocabulary learning and with predictions of the TOPRA model regarding the interactive relationship between these dissociable components. They are inconsistent, however, with claims of more generalized benefits of increased semantically oriented “deeper” processing in the realm of vocabulary learning. Claims of evidence for more generalized benefits of this nature are often complicated by how one given task is “deemed” to involve more (or less) semantic processing than another.

3 Processing specificity in incidental vocabulary learning

Research on incidental L2 vocabulary learning has revealed a number of factors that affect word learning. Early experimental research indicated that pedagogical interventions, such as dictionary use, single/multiple-choice gloss, or exposure frequency to the same word, promote word learning whereas other research has demonstrated how incidental L2 vocabulary learning takes place at multiple levels, such as the levels of
orthography, form-meaning mapping, and so forth. Several limitations on this research exist, however. Research in this area has been rather exploratory, for example, as evidenced by the many types of post-reading vocabulary tests that have been employed, ranging from simple word recognition tests without recall of meaning (Hulstijn, Hollander, & Greidanus, 1996) to more productively oriented tests that involve word use in a sentential context, as in the case of one level of the Vocabulary Knowledge Scale (Paribakht & Wesche, 1993). More recent research has also involved more psycholinguistic techniques, such as eye-tracking (Pellicer-Sánchez, 2016) or priming (Elgort & Warren, 2014), to investigate incidental acquisition of rapid and automatic word recognition skills. Although the use of variety of assessment measures has helped us to deepen our understanding of the effects of different variables on incidental L2 vocabulary learning, the variety also suggests that there seems to be no agreement among researchers in terms of how to measure L2 learners’ incidental vocabulary gain.

One reason for this lack of agreement has been an overall limited amount of theoretically motivated hypothesis testing in this field. This is particularly prominent when it comes to processing specificity. One exception is the application of the TOPRA model (Barcroft, 2002) into incidental learning research. For example, Barcroft (2009) demonstrated that performing synonym-generation, a semantically oriented task, negatively affected L2 word form learning even in the incidental learning context. Another example comes from Kida (2010), who demonstrated that writing L2 word’s pronunciation in L1 script, a structurally oriented task, had a positive impact of on L2 incidental word form learning. These studies suggest that the TOPRA model is applicable in incidental learning research as well. However, since the model is not designed to address the role of processing specificity specifically in the incidental learning context, there is still room for further investigation. For example, what other dissociable types of processing might be specific to incidental contexts of vocabulary learning? Processing for inferencing word meanings? Other dissociable sub-processes within processing for inferencing word meaning? How can we investigate other aspects of processing specificity that have not been previously examined, such as L2 words’ syntactic behaviour or the development of automatization? Future research is needed in order to construct a model that deals with role of processing specificity as it relates to incidental L2 vocabulary learning in particular.

4 Processing specificity in learning collocations

It is commonly claimed in the literature that L2 collocational knowledge is acquired late and not mastered even by relatively proficient learners. One fundamental aspect to initially consider when discussing ‘collocation’, however, is how its varied definition in the literature may encompass anything from semantically transparent word combinations in which the word components refer to literal senses (e.g., kick a ball), to word combinations where the word components make up a non-compositional, unitary meaning as a whole, with a high degree of figurativeness (e.g., pull strings) (Barfield & Gyllstad, 2009; Gyllstad & Wolter, 2016). Importantly, the level of figurativeness and holistic meaning is likely to affect how collocations are processed and acquired in a second language (L2), and we need to acknowledge that collocations may reside at different points of a continuum of figurativeness and semantic transparency.

Some authors have argued that the reason why L2 learners seemingly have problems with collocations is that they focus too much on individual words rather than collocations in their input, possibly because these constructions rarely cause comprehension problems (Henriksen & Stenius Staehr, 2009; Wray, 2002). Firstly, however, this explanation presupposes that the collocations dealt with come from the more semantically transparent end of the assumed continuum. Secondly, there are studies that have shown that L2 learners are capable of noticing and retaining novel collocational patterns in their input (Durrant & Schmitt, 2010) and that any deficit in collocational knowledge would stem from insufficient exposure. Wolter and Gyllstad (2013) provided evidence of the fact that at least advanced learners are highly sensitive to input frequency effects for L2 collocations. Thus, it could be that type of input and task play a significant role. As a case in point, Boers, Demecheleer, Coxhead, and Webb (2014) report on a small-scale study in which input type was manipulated, comparing learning gains from various matching exercises, where learners were expected to assemble whole verb + noun collocations from provided word component parts, with input in the form of intactly presented collocations. Interestingly, low, non-significant learning gains were obtained overall. The authors still cautiously advocate formal input of whole collocations but call for future studies where comparisons can be made to meaning-focused input with ample repetition of the same collocations.

Finally, it could be that many of the problems observed for L2 collocational behaviour are related to L1 influence. In production (Laufer & Waldman, 2011) as well as receptive processing (Wolter & Gyllstad, 2013; Yamashita & Jiang, 2010), incongruent L2 collocations, i.e. collocations for which there is not a corresponding L1 collocation available that can be felicitously translated in terms of word components’ core meanings, seem to be the biggest stumbling block.
5 Summary and conclusion

These research reviews and analyses underline the critical role of processing specificity in vocabulary learning as a whole and in particular with regard to the type of input that learners are provided and the types of tasks that they may be asked to perform when exposed to novel words and lexical phrases in the input. Research findings in this area draw into question the viability of discussing how a given variable may affect “vocabulary learning” or “word gain” in general. Instead, they invite consideration of how a given variable may affect multiple dissociable components of vocabulary learning in different ways.

References


The More Emotionally Intelligent, the More Likely to Remember Words without Elicited Emotional Elaboration: A Possible Individual Difference in L2 Lexical Retention

Yu Kanazawa
Graduate School, Kwansei Gakuin University
yu-kanazawa-at-outlook.com; yu-kanazawa-at-kwansei.ac.jp; @knzw783

1 Introduction

Emotion is the pervasive phenomenon whose pivotal impacts on cognitive processes, including learning and memory, have become widely acknowledged (e.g., Immordino-Yang, 2016; Talmi, et al., 2007). The movement to focus on affective dimensions in SLA has also become salient (e.g., Pavlenko, 2013). Although the majority of previous SLA studies with interests on emotion focused on conscious and socio-psychological aspects of emotion (i.e., macro-level emotion), such as the perception and construction of one’s selves and motivation, neuroscientific findings suggest that subconscious cognitive psychological emotion (i.e., micro-level emotion) plays an even more significant role in cognition including language processing and memory.

Kanazawa (2016a) investigated whether emotional attributes which dwelled as a microscopic object in words (i.e., lexical emotional valence) had any impact on L2 memory and discovered that micro-level positive valence facilitated L2 lexical memory without the need for deliberate effort to remember. Kanazawa (2016b) extended the target of emotion-involvement from individual L2 words to the processing mode and revealed that the lexical items which were processed under the condition of Emotion-Involved Processing (E-IP), in which intra-lexical emotional elaboration was elicited, had a significant advantage of incidental recall and recognition over the lexical items which were processed semantically but non-emotionally (non E-IP). The finding corresponds to and extends Nation’s (2015) remark on the significant effectiveness of elaboration on vocabulary learning.

2 Trait emotional intelligence

On the other hand, it has also been reported that the effectiveness of emotional elaboration on memory differs depending on psychological personality traits, viz., one’s Emotional Intelligence (Toyota & Sato, 2009). Emotional Intelligence (EmInt) refers to “the subset of social intelligence that involves the ability to monitor one’s own and others’ feelings and emotions, to discriminate among them and to use this information to guide one’s thinking and actions” (Salovey & Mayer, 1990, p. 189). EmInt has also reported to affect L2 communication (Dewaele, Petrides, & Furnham, 2008). Trait EmInt can be measured via self-report questionnaires. One of such questionnaires is the J-WLEIS (Toyota & Yamamoto, 2011). J-WLEIS is the Japanese translated version of WLEIS (Wong & Law, 2002), which consists of 16 items to estimate the trait EmInt. J-WLEIS was selected for the present study in view of its practicality and empirically proven validity.

The present study incorporated EmInt as the between-participants factor and investigated whether and how the difference of EmInt affects incidental lexical retention of either emotional or non-emotional processing.

3 Method

The participants were Japanese users of English, among whom a high-EmInt group (N = 11; EmInt M = 48.45, EmInt SD = 3.65) and low-EmInt group (N = 11; EmInt M = 37.64, EmInt SD = 4.07) were subject to the analysis. Their EmInt were estimated using J-WLEIS questionnaire. Since a 4-point Likert scale was adopted, the theoretically possible score ranged from 16 to 64. In the study session, English words were presented on a computer screen one by one and s/he was asked to (a) judge the emotional valence (E-IP condition) or (b) judge lexicality of the words (non E-IP condition). In the following test session, s/he was instructed to recall as many words which had been presented in the study session as possible.

4 Results

Due to the limited number of data, Wilcoxon signed-rank test was implemented for the statistical analysis. It was found that although E-IP resulted in better memory performance regardless of EmInt, the statistical significance and effect size for low-EmInt group (Z = -2.85, p < .01, r = -.86) were bigger than those for high-EmInt group (Z = -2.26, p < .05, r = -.68). The correlational analysis revealed a significant positive correlation between EmInt and recall performance of words in non E-IP condition (rho = .49 [df = 21, p < .05]; Figure 1).
Emotion is the pervasive phenomenon which intra-personally affects our psychological reactions to others and to our own and others’ behaviors, and in turn, affects our performance in social interaction (Pavlenko, 2008). Emotion-Involved Processing (EmInt) is an individual difference in emotional processing (e.g., emotional elaboration) that is well-validated in second language acquisition research (Salovey & Mayer, 1990; Immordino-Yang, 2016; Toyota & Sato, 2009). EmInt learners who need to be explicitly elicited to activate and utilize their emotional perception for cognition.

While it has been reported that lexical emotional elaboration plays an important role in L2 lexical memory (Salovey & Mayer, 1990), micro and macro aspects of emotion (i.e., emotionality, emotion intensity, and mood) have also been reported to affect language learning (Dell & Amsel, 2000; Pavlenko, 2008). For example, Pavlenko (2008) showed that emotionality and intensity were positively correlated with vocabulary learning, while mood was negatively correlated with vocabulary learning.

Table 1. Descriptive Statistics of the Results of Free Recall Performance

<table>
<thead>
<tr>
<th>Participant Group</th>
<th>High-EmInt</th>
<th>Low-EmInt</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>E-IP</td>
<td>Non E-IP</td>
</tr>
<tr>
<td>M</td>
<td>9.18</td>
<td>5.64</td>
</tr>
<tr>
<td>SD</td>
<td>3.45</td>
<td>4.20</td>
</tr>
</tbody>
</table>

Figure 1. Positive Correlation between Individual EmInt Score and Recall Performance of Non E-IP Items.

5 Discussion

The present results indicate that low-EmInt people benefit even more from E-IP concerning L2 lexical retrieval than high-EmInt people do. Considering together the relatively higher recall performance of non E-IP items for high-EmInt participants (Table 1), the likely rationale is that since high-EmInt people can regulate and make use of their emotion better, they may have consciously or unconsciously “woven” their micro-level emotional disposition into lexical processing. In pedagogical terms, the present study implies that it is low trait EmInt learners who need to be explicitly elicited to activate and utilize their emotional perception for cognition.

References


Making Word Lists

Paul Nation
Victoria University of Wellington and Temple University Japan
Paul.Nation -at- vuw.ac.nz

1 Introduction

Word lists lie at the heart of good vocabulary course design, the development of graded materials for extensive listening and extensive reading, research on vocabulary load, and vocabulary test development (Nation, 2016). It is important that word lists are well made, but most lists have not been made using an explicit and well supported set of guidelines like the following.

2 Guidelines

1. The purpose and audience for a word list need to be clearly described. These descriptions provide a basis for evaluating the list.

2. Very careful thought needs to be given to the unit of counting words, namely what level of word family should be chosen? This decision needs to be explicitly justified not by accepting what others have done, but by explaining the purpose of the word lists and showing how the chosen unit best reflects that purpose.

3. List makers need to deal with homonymy but not polysemy.

4. Within the category of proper names, a list of proper names requiring previous knowledge should be developed. This is because these words require learning in much the same way other words are learned. This list will also require continual updating as particular individuals, places, enterprises and events become well known. Clear criteria for the inclusion of words in this list should be developed.

5. A well-conducted word frequency count should distinguish the members of homonyms where one member is a proper name with a meaning that is clearly different from its common noun use.

6. Within the category of proper names, a list of proper names requiring previous knowledge should be developed.

7. Transparent compounds and hyphenated words need to be split into their parts.

8. Multiword units should be in separate lists.

9. There need to be explicit criteria for deciding what are multiword units and these criteria need to be followed.

10. The content of a corpus should represent the actual or potential language uses of the target audience for the resulting word lists.

11. The corpus size should match the frequency level of the words that are the focus of the count, and should be enough to get a reliable list. Around 20 million tokens is recommended for high-frequency and mid-frequency words.

12. To ensure that the most useful words occur early in the lists, both range and frequency of occurrence need to be considered when gathering data on the words.

13. Because any corpus is likely to be only an approximate representation of what learners need to know, and because of the limitations of word counting programs, the use of the criteria of frequency and range may need to be accompanied by more subjective criteria.

14. The procedures for making a word list need not be replicable but should be transparent.

It is not likely that a general service list of high frequency words can be developed using the current methodologies. Such a list may need to be modular with different words added from different homogeneous corpora.

Reference

Grammatical Class and Transitivity Influence in Word Association

Peter Thwaites
Yeungnam University and Cardiff University
petertheteacher -at- gmail.com

1 Introduction

Word association responses are subject to the influence of numerous variables pertaining to individual differences amongst respondents, environmental factors, and the features of cue words. Not all of these influences are well understood. Research into lexical factors in word association, in particular, is currently not well developed, in spite of increasing interest in the influence of such variables on other psycholinguistic tasks (Pexman, 2012). To date, what little research has been done in this area has suggested that grammatical class (Deese, 1962; Nissen & Henriksen, 2006) and concreteness (Van Hell & De Groot, 1998) are both capable of influencing word association responses to some extent. The current paper is part of a larger study aiming to build on these findings by systematically investigating the influence of a range of variables on word association responses.

2 Aims

The current paper has two aims. The first is to investigate the influence of verb transitivity on word association responses. While previous research has suggested that nouns and verbs yield different patterns of response behaviour, the influence of transitivity has not yet been investigated. Secondly, the paper presents an attempt towards a solution to a long-standing problem in word association research, namely, the subjectivity of response categorization. Most word association research uses one or more human coders (usually the researcher) to classify responses into a single response-type category. This method suffers not only from being subjective (even in the event that dual coders are used), but also from the fact that the categorization of a word into a single class obscures other forms of relationship between the cue and its responses. While a total solution to this problem is some way off, the present study employed a corpus research method in order to validate the researcher’s intuitions about the coding of collocational associations.

3 Findings of subjective codings

Responses were collected to 100 cues, equally divided between transitive and intransitive verbs. Cues were matched on a number of variables. 54 adult participants provided responses. Subjective coding of these responses revealed a significant effect of transitivity: transitive verbs were significantly more likely than intransitive verbs to yield position-based responses categorized as “cue-response” (i.e., the response was a word which typically follows the cue, e.g., cue = vanquish, response = foe), while intransitive verbs were more likely to elicit responses in the opposite direction (cue = erupt, response = volcano). Both effects were significant at the p < .01 level.

Figure 1. Mean Number of Position-Based Responses to Each Cue, by Transitivity.
4 Findings of corpus analysis

In order to verify the above findings, Mutual Information (MI) data was collected on each cue-response pair. Data was collected from the TenTen corpus (Jakubíček et al., 2013). For each cue, three measures were collected, representing a window size of 2 words before, 2 words after, and two words on either side of the cue. In view of the results of subjective coding, it was expected that responses to intransitive verbs would yield higher MI scores than transitive verbs in the MI -2 range, while transitive verbs would have higher scores in the MI +2 range. This prediction held in the MI -2 range, but not in the MI +2 range (see Figure 2, below). As Figure 2 shows, intransitive verbs yielded responses with significantly higher MI scores than did transitive verbs, irrespective of directionality.

![Figure 2. Mean MI Score per Cue-Response Pair. Data from enTenTen Corpus (Jakubicek et. al., 2013). * = p < .01.](image)

5 Conclusion

While the subjectively coded response types suggest a clear effect of transitivity on response behaviour, this picture is complicated by the corpus analysis of directional MI scores. The failure of this analysis to support the dominance of transitive verbs in the cue-response (i.e., MI +2) range requires further investigation. The results for the MI +/-2 range suggest that intransitive verbs simply yield more position-based responses than do transitive verbs. However, it is also possible that the correspondence between MI score and subjective codings of response type are less straightforward that previously assumed. For example, sometimes words which are categorized as synonyms (e.g., cue = irrigate, response = flush) also contribute directional MI effects (MI -2 = 0, MI +2 = 7.22). This is an issue requiring further research.

References


Finding Meaning in the Form of L2 Words: What Takes Place in the Learner’s Mind During the Meaning-making Process?

1 Introduction
Creating a link between the form and meaning of unknown words is the initial step in vocabulary learning, enabling learners to build up receptive and productive knowledge of these new words. Research has shown that prompting learners to elaborate on the appropriateness of form-meaning links can be an efficient vocabulary learning exercise (Deconinck, Boers & Eyckmans, 2015). In this paper, we wish to shed more light on the mental processes that occur during this specific elaborative task by investigating which types of elaborations learners make. In addition, we explore the influence of three individual learner factors on the type and number of elaborations that learners make: vocabulary size, number of known languages and vocabulary learning strategy.

2 Method
The participants were fifty Dutch-speaking EFL learners, of which 13 were male and 37 were female. They completed the LexTale test (Lemhöfer & Broersma, 2012) which measures receptive English vocabulary size and the Productive Vocabulary Levels Test (Lafer & Nation, 1999) which gauges productive English vocabulary size. In a language background questionnaire, the participants indicated which languages they speak.

Next, the learners were instructed to rate the congruency they perceived between the form and meaning of 24 English words on a 6-point Likert scale. After completing this task, we elicited their motivation for these ratings by means of a think-aloud protocol. We then transcribed the think-alouds and analyzed them with regard to the type of elaborations made.

Following the form-meaning motivation task, we inquired after the learners’ use of three vocabulary learning strategies which can be employed during the form-meaning motivation task: performing a word analysis (i.e. breaking the word down into smaller parts), comparing the L2 word with another word which has a similar form, and comparing the form of the L2 word with the form of another word which has the same meaning.

3 Results and discussion
We identified five types of elaborations: cross-lexical associations, sound-symbolic associations, word form comparisons, morphological associations and idiosyncratic associations. Cross-lexical associations (CLA) imply that the L2 word was associated with another known L1, L2 or L3 word that has a similar form. Sound-symbolic associations (SSA) indicate that the learner attributed sound-symbolic features to the word. Word form comparisons (WFC) signify that the learner compared the form of the L2 word with the form of another L1 or L2 word which has the same meaning. Morphological associations (MA) are based on the learner’s morphological knowledge of English. In this case, the learner performed a word analysis or associated its form with a specific word class. Finally, idiosyncratic associations (IA) incorporate elaborations that do not fit any of the previous categories, but this only amounts to a small percentage. The distribution of these elaborations is demonstrated in Table 1.

<table>
<thead>
<tr>
<th>Type of association</th>
<th>Number of occurrences</th>
<th>Percentage (n=870)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLA</td>
<td>399</td>
<td>45.8%</td>
</tr>
<tr>
<td>SSA</td>
<td>126</td>
<td>14.5%</td>
</tr>
<tr>
<td>WFC</td>
<td>168</td>
<td>19.3%</td>
</tr>
<tr>
<td>MA</td>
<td>148</td>
<td>17%</td>
</tr>
<tr>
<td>IA</td>
<td>29</td>
<td>3.3%</td>
</tr>
</tbody>
</table>

Table 1. Distribution of Elaborations Made by the Learners
A multiple linear regression analysis demonstrated that the larger the participants’ receptive L2 vocabulary size, the lower the form-meaning rating \( (t = -4.329, p < 0.0001, \text{Adjusted } R^2 = 0.243) \). When learners with a large receptive L2 vocabulary size first see an unknown L2 word, this will trigger more words than in the case of learners with a smaller receptive L2 vocabulary size. Consequently, the word will be associated with a number of different meanings and its form-meaning fit might be perceived as less motivated. We also found that increased receptive L2 vocabulary size resulted in more associations \( (t = 3.716, p = 0.0002, \text{Adjusted } R^2 = 0.045) \). The same effect was observed when we zeroed in on cross-lexical associations with the L2 \( (t = 3.265, p = 0.001, \text{Adjusted } R^2 = 0.135) \) and on word form comparisons with the L2 \( (t = 2.174, p = 0.03, \text{Adjusted } R^2 = 0.037) \). Productive L2 vocabulary size also positively influenced the number of cross-lexical associations that were made with the L2 \( (t = 2.615, p = 0.009, \text{Adjusted } R^2 = 0.135) \). Naturally, it will be easier for the learners to associate an L2 word with other L2 vocabulary if they have more linguistic knowledge to exploit.

Surprisingly, learners who knew more languages made significantly fewer elaborations \( (t = -2.141, p = 0.03, \text{Adjusted } R^2=0.045) \). Possibly, the more multilingual learners had already elaborated on the vocabulary implicitly when first seeing the vocabulary or when rating the form-meaning fit of the words, and so consequently they did not make these elaborations explicit during the think-aloud. The less multilingual language learners may not have made these implicit elaborations during the previous tasks, and therefore would have benefited more from the think-aloud procedure to produce them explicitly.

Reported use of the word analysis strategy appeared to positively impact the total number of elaborations \( (t = 2.596, p = 0.01, \text{Adjusted } R^2=0.045) \). The other vocabulary learning strategies did not have an influence on the total number of elaborations. The elaboration types do seem to reflect the learners’ vocabulary learning strategies. Making cross-lexical associations was the most popular vocabulary learning strategy among participants, with 36 participants indicating that they employ this strategy, and it is also the most frequently occurring type of elaboration in the think-alouds. Making word form comparisons and word analyses are on par, with half of the participants admitting to employing these strategies. When it comes to occurrence in the think-aloud procedure, the two categories are also level: word form comparisons make up 19.3% of the total number of elaborations, whereas morphological associations comprise 17% of the elaborations.

4 Conclusion

Deconinck et al. (2015) found that elaborating on the form-meaning connection of an unknown L2 word is a technique which can be exploited in the vocabulary learning classroom, and established four types of elaborations. Our study demonstrated that when a different set of target items is used, the same categories are discerned. As such, the study reaffirms that L2 learners possess an implicit ability to elaborate on new L2 vocabulary. Language teachers can tap into this ability and employ the form-meaning motivation method to promote acquisition of new L2 words. To that end, it is important that they are aware of the four main types of elaborations which can occur during such an activity: cross-lexical associations, sound-symbolic associations, word form comparisons, and morphological associations. When prompted, learners appear to spontaneously employ these types of elaborations, so further awareness-raising through the form-meaning technique could result in the increased use of these elaborations as a vocabulary learning strategy. Future research should therefore investigate whether making learners aware of these elaborations before they carry out the form-meaning motivation exercise leads to higher learning gains.

References


Investigating the Nature of Academic Spoken Vocabulary

Yen Dang  
Victoria University of Wellington  
New Zealand  
ngocyen1011 -at- gmail.com

Averil Coxhead  
Victoria University of Wellington  
New Zealand  
averil.coxhead -at- vuw.ac.nz

Stuart Webb  
University of Western Ontario  
Canada  
swebb27 -at- uwo.ca

1 Extended abstract

Comprehending academic speech (e.g., lectures, seminars, labs, and tutorials) is essential for second language (L2) learners to achieve success in their academic studies in English-medium programs (Biber, 2006). Given the close relationship between vocabulary knowledge and listening comprehension (Van Zeeland & Schmitt, 2013), it is beneficial for these learners to study the lexical items that occur frequently in a wide range of academic speech events. Previous research has suggested that an academic written word list may not fully cover the most frequent, wide ranging words in academic speech. Corpus-driven studies revealed a huge difference in the coverage of Coxhead’s (2000) written Academic Word List (AWL) in academic speech (around 4%) (Thompson, 2006; Dang & Webb, 2014) and academic writing (around 10%) (see Coxhead, 2011 for a review). Lexico-grammar research (Biber, Conrad, Reppen, Byrd, & Helt, 2002; Biber, 2006) reported a clear-cut distinction between the linguistic features of academic writing and academic speech. Importantly, Csomay (2006) found that university classroom teaching reflects the situation where speakers need to convey information focused message under real time producing circumstances. She called this distinctive feature of university classroom teaching on-line informational elaboration. Together these findings indicate the value of exploring the nature of academic spoken vocabulary.

Despite these pieces of research, little is known about what vocabulary L2 learners need to prepare for academic speaking and listening at university. To the best of our knowledge, only two studies have developed spoken academic word lists (Nesi, 2002; Simpson-Vlach & Ellis, 2010). The former focused on single word units while the latter investigated multi-word units. Unfortunately, to date, there is no precise description of Nesi’s (2002) list. To better understand the nature of academic spoken vocabulary, further research on developing an academic spoken word list of single word units is needed.

Our study was conducted to address the following research questions:

1. Which lexical items occur frequently and are evenly distributed in a wide range of academic speech?
2. What is the coverage of these items in independent collections of academic speech, academic writing, and non-academic speech?
3. What is the overlap between these items with lists of general high-frequency words and academic written words?

To identify the most frequent, wide ranging lexical items of academic speech, we created a 13 million running word academic spoken corpus. This corpus was made up of materials from 24 subject areas across four equally-sized disciplinary sub-corpora: hard-pure (e.g., maths, physics), hard-applied (e.g., engineering, medicine), soft-pure (e.g., history, art), and soft-applied (e.g., business, law). The division of the sub-corpora followed Becher’s (1989) classification of academic disciplines at higher education. The hard/soft dimension refers to the degree to which a paradigm exists. The pure/applied dimension is related to the application to practical problems. Four speech events were represented in the corpus, including those which are common to all subject areas (lectures and seminars) as well as those that are distinctive to hard subjects (labs) and soft subjects (tutorials). Based on the corpus, we used selection criteria to identify the 1,741 most frequent, wide ranging, and evenly distributed word families in the corpus and included them in an Academic Spoken Word List (ASWL). Given the large, balanced, and representative corpus from which the ASWL was developed, it is expected that the ASWL captures the most frequent, and wide ranging words in academic speech.

The 1,741 ASWL words were graded into four levels according to Nation’s (2012) BNC/COCA word lists. Levels 1 and 2 are general high-frequency words from Nation’s (2012) first and second 1,000 BNC/COCA frequency levels which are also high frequency, wide ranging, and evenly distributed items in academic speech. Some examples of ASWL words at these levels are alright, agree, though, maybe, stuff (Level 1), example, identify, determine, similar, and therefore (Level 2). Levels 3 and 4 include ASWL words that have high frequency in academic speech and are outside general high-frequency words. For instance, define, potential, focus, versus, and achieve are ASWL at Level 3, and diagram, arrow, domain, maximise, tutor are AWL words at Level 4. These levels serve as a guide for learners and teachers to determine which vocabulary they should
focus on to match learners’ existing vocabulary level.

To explore the nature of academic spoken vocabulary, we examined the coverage of the ASWL in three independent corpora: a second academic spoken corpus (with the same structure as the original spoken academic corpus in this study), an academic written corpus, and a non-academic spoken corpus. These corpora were approximately the same size as the corpus from which the list was developed. Also, we determined the overlap between the ASWL with Coxhead’s (2000) AWL, Nation’s (2012) BNC/COCA2000, and Gardner and Davies’s (2014) Academic Vocabulary List (AVL). Nation’s (2012) BNC/COCA2000 represents general high-frequency vocabulary while Coxhead’s (2000) AWL and Gardner and Davies’s (2014) AVL represent academic written vocabulary. Together, the results indicated that academic spoken vocabulary has features associated with vocabulary in general conversation and academic writing, but also reflects the distinctive feature of academic speech, which highlights the value of the ASWL.

2 Acknowledgements

We would like to thank the following publishers and researchers for their generosity in letting us use their materials to build our corpora: Cambridge University Press, Pearson, Dr. Lynn Grant, the lecturers at Victoria University of Wellington, the researchers in the International Corpus of English project, the British Academic Spoken English corpus project, the British Academic Written English corpus project, the Yale University Open courseware project, the Massachusetts Institute of Technology Open courseware project, the University of California, Berkeley Open courseware project, and the Stanford University Open courseware project.

We would like to thank the Postgraduate Students’ Association and the Faculty of Humanities and Social Sciences at Victoria University of Wellington, New Zealand.

References


Developing a Multi-Functional Online Word Tier Analyzer: Investigating Vocabulary Load, Genre, and Modification of Texts

Tatsuhiko Matsushita
University of Tokyo
matsushita -at- global.c.u-tokyo.ac.jp

1 Introduction

As Webb & Nation (2008) noted, adjusting the vocabulary load of a text material at an appropriate level is highly important for teaching and learning. This study aims at developing a multi-functional online word tier analyzer for education and research.

For lexical analysis of a text, there have been some useful tools such as Range (Heatley, Nation, & Coxhead, 2002) and AntWordProfiler (Anthony, 2014). Complete Lexical Tutor v.8 (Cobb, 2016) has collected many useful functions to analyze texts from a lexical viewpoint.

For Japanese texts, Reading Tutor (Kawamura, Kitamura, & Hobara, 1997), jReadability (Lee, Hasebe, Kubo, & Asao, 2013), Yasanichi Checker (Iwata, Mori, Matsushita, & Nakajima, 2015) and J-LEX (Suganaga & Matsushita, 2013) are often used. However, these tools do not always meet with the purpose as there are some insufficiencies in interface, functions and/or the quality of baseword lists.

Reading Tutor is widely used by teachers and researchers of Japanese as it provides an online multilingual dictionary function with a lexical frequency profile. However, it has two major problems since it incorporates the vocabulary lists of the former Japanese Language Proficiency Test (JLPT) as the baseword set for the profile. One problem is that the baseword set seems outdated; the other problem is that Reading Tutor only returns the five different lexical levels of the former JLPT. jReadability’s also only returns six different levels for the profile. Yasanichi Checker only returns four different levels.

J-LEX was designed to overcome the weaknesses mentioned above. The baseword set was created from a newer corpus and proved to provide better text coverage than the JLPT lists incorporated in Reading Tutor (Matsushita, 2012; 130–141). It returns the lexical frequency profile by each one thousand word level up to the 20K word level with the estimated text coverage (including the text coverage of the Assumed Known Words i.e. proper names and interjections). It also provides the text coverage using the range between any two frequency rankings and highlights the words which are not covered by the designated frequency range. This function is very useful for modifying the text to adjust the lexical level to the target learner group level.

This study aims at adding some functions to J-LEX to develop a more multi-functional and user-friendly word tier analyzer for teachers and researchers of Japanese as a second language.

2 The functions of the new J-LEX

The new J-LEX is mainly designed for the following four purposes:

1) to check the vocabulary load (lexical level) of the text;
2) to modify the text multiple times in the window to adjust to a learners’ lexical level by checking the vocabulary load multiple times;
3) to check the genre features of the text, i.e. how academic or literary the text is, and how specific the text is to which academic genre (i.e. humanities, social sciences, technology, or biological sciences); and
4) to create the target word list of the text.
Specifically, the new J-LEX will have the following functions:

a) Return different types of results by pasting a text in a window and clicking buttons to designate a target frequency level and a baseword set. For example, it highlights the words above the set word level in another window as well as the word frequency profile of the text. (This function is kept from the current J-LEX. See Figure 1.)

b) Users can also choose a baseword set of a textbook from major textbooks and designate the learned lessons. This will be useful for controlling vocabulary load of readers for elementary learners. The current J-LEX incorporate one textbook baseword set, but more baseword sets from other textbooks will be incorporated in the new J-LEX.

c) For analyzing the genre specificity (i.e. literature, humanities, social sciences, technology, biological sciences), it will return the Text Covering Efficiency (Matsushita, 2012; 296–297) by genre and frequency level. This will also return the ‘genre coefficient’ calculated from the TCE values to show the degree of genre specificity.

d) After specifying plural cells in the profile and right-clicking there, it will highlight the words correspondent to the cells, and enable the user to download the list of the highlighted words. This will be convenient for creating a word list for new reading material.

3 The ideas for checking the vocabulary load and the genre specificity of a text

One unique function with the new J-LEX is to return the Text Covering Efficiency (TCE) (Matsushita, 2012; 296–297) figures of the text. TCE is ‘the expected number of tokens of a lexeme in the tested group in a one-million-token text in the target domain’, or ‘the mean text coverage per unit number of words of the tested group of words’. ‘The tested group’ here means a group of words such as the Japanese Common Academic Words (Matsushita, 2012). TCE is calculated by the formula shown below.

\[ E = \frac{N_t}{L_s} \times \frac{1,000,000}{N_t} = \frac{N_t \times 1,000,000}{L_s \times N_t} \]

\[ E: \text{Text covering efficiency} = \text{Expected number of tokens of a lexeme in the tested group in a one-million-token text in the target domain} \]

\[ F_t: \text{Number of tokens of the tested group of words in the target text} \]

\[ L_s: \text{The total number of lexemes of the tested group of words} \]

\[ N_t: \text{Number of tokens in the target text} \]

The TCE figures of different types of words at different frequency levels will demonstrate the vocabulary load and the genre specificity of a text in a visual and intuitive way. TCE is a robust index by which different lexical features in different genres can be clarified. For example, such an analysis allows you to say things like, “Learning the intermediate Japanese Common Academic Words is 6.2 times more efficient in covering Japanese social science texts than learning other words at the same level, and 8.3 times more efficient than learning the advanced Common Academic Words”. In addition, it should be possible to calculate some kind of genre coefficient from those TCE figures.

References


Changes in University Students’ Receptive English Vocabulary Knowledge During EAP Instruction

Rachael Ruegg
Akita International University
rruegg -at- aiu.ac.jp

Naeko Naganuma
Akita International University
nana9801 -at- aiu.ac.jp

1 Introduction

Whether reading in the L1 or an L2, learners who know more words read better than those with a smaller vocabulary. In fact, “vocabulary knowledge correlates more highly with reading ability than any other variable” (Koda, 2005, p. 186). Previous studies have found correlations of between 0.5 and 0.78 (Laufer, 1996; Qian, 2002) between vocabulary knowledge and reading comprehension ability. Moreover, more recently some experts have stated that the two are reciprocal (Koda, 2005; Nation, 2001).

It has been stated that knowledge of around 5,000 words is necessary to read and understand unsimplified texts (Schmidt, Schmidt & Clapham, 2001). Previous research in Japan (Barrow, Nakanishi & Ishino, 1999) found that first year university students knew 2,304 of the 4,000 high frequency words. Research conducted in China (Zhang & Lu, 2013) found that after six years of formal English education, first year university students knew around 5,500 words. On the other hand, Hazenberg and Hulstijn (1996) found that at least 10,000 words would be necessary to cope with university-level study in English. There has been a large increase in the number of English-Medium Instruction (EMI) programmes within Japanese universities. However, referring to the previous research, it would appear that students who hope to study at the university level in English would first need to almost double their vocabulary knowledge. Meanwhile, in EFL contexts, most studies have found that students’ vocabulary increases by 500-600 words per year, suggesting that a further six years of English instruction might be necessary before students would achieve an appropriate vocabulary size for EMI.

It seems reasonable to assume that students who apply for and are accepted into an EMI programme would have a higher level of English proficiency than average first-year Japanese university students. Therefore, this study was conducted in order to understand how much vocabulary first-year Japanese university students in an EMI programme had receptive knowledge of and how much they learnt over their time in a preparatory EAP programme. The receptive vocabulary knowledge of 144 incoming first-year students in an EMI programme was measured using Version A of the Vocabulary Size Test (VST) (Nation & Beglar, 2007). Their vocabulary knowledge was then measured again at the time of completing the EAP reading programme and moving into higher level EMI studies.

2 Measuring vocabulary knowledge

There is a formula by which the knowledge at each vocabulary level can be multiplied in order to extrapolate the knowledge of vocabulary at one level to other levels. However, this practice is based on the frequency model of vocabulary learning, which posits that learners should know more vocabulary at higher frequency levels and progressively less at each level of lower frequency. A demonstrated by Milton (2007), a large number of learners do not fit this model for a number of reasons. Most relevant to the current study is that, learners of English in an EFL environment have had exposure to English almost exclusively through textbooks and the textbooks are not necessarily created on the basis of frequency level. In particular, Milton and Vassiliu (2000), as well as Ruegg and Brown (2014), found a large amount of infrequent vocabulary in English language textbooks. Learners who have learnt the language in this context, using these kinds of textbooks, will be disadvantaged by the idealistic model of vocabulary size measured using the Vocabulary Levels Test as the vocabulary size estimates will not give credit for the potentially large number of low-frequency words known by them.

Another widely used test to measure vocabulary size is the VST (Nation & Beglar, 2007). However, the VST has also been criticized due to the fact that the sample size is only 5 words per 1,000 word frequency level. Thus, although it measures vocabulary at every level up to the 20,000 word level, rather than extrapolating information from one level to another, it is a relatively superficial measurement. On the other hand, as mentioned by Read (2000, p. 115), it “can give a more representative picture of the overall state of a learner’s vocabulary level than an in-depth probe of a limited number of words” at a limited number of levels. Another criticism of the VST has been the multiple-choice format of the test. It has been stated that learners may use test-taking strategies, such as guessing. Due to the format of the test, it is clear that random guessing
alone could measure a learner’s vocabulary size at 5,000 words, a quarter of the 20,000 measured, even if they do not know a single word of English. However, the instructions of the test direct test-takers to skip items they do not know and only answer those they know, or think they know. If the test is being used only for research purposes and the results have no stakes for the learners, it is difficult to imagine why they would try to guess items they did not know when skipping them would be quicker and easier. Furthermore, Shillaw (1999) found that Japanese learners in particular guessed less than learners from other cultural contexts. In the present study, in order to ascertain whether learners had guessed, test papers were checked and it was found that students had skipped questions without exception.

3 Results

As assumed, it was found that the incoming first-year students entering an EMI programme on average knew more than those in previous studies (2,930 of the 4,000 high frequency words and 8,193 words in total). Similar to the results of previous studies, it was found that students on average learnt 541 words over a period of one year in the EAP reading programme. Moreover, they increased significantly in receptive knowledge of high-frequency and mid-frequency words, while the increase in low-frequency words was not significant. By the end of the EAP reading programme, the students knew on average 3,159 of the 4,000 high frequency words and 8,734 words in total. Although the total vocabulary size is still less than the 10,000 suggested by Hazenberg and Hulstijn (1996), it is relatively close, and given the non-anglophone context of the EMI programme, as well as the abundant input still to come during several years of undergraduate EMI study, there is reason to believe that it will be sufficient for the students to succeed in their academic endeavours.

References


The New JACET8000 and its Academic Vocabulary List

Naoki Sugimori
Ritsumeikan University
naosugi -at- is.ritsumei.ac.jp

Masamichi Mochizuki
Reitaku University
mmotizu -at- reitaku-u.ac.jp

1 Introduction

JACET8000 is an educational word list for Japanese learners of English that was developed by the Basic Words Revision Committee of Japan Association of College English Teachers (JACET). Since its first publication in 2003, it has been widely used as a benchmark for vocabulary teaching in Japan. In 2016, a revision, the New JACET8000, was published. This paper will describe the revision principles and the compilation process of the new JACET8000 word list and its supplementary academic vocabulary list.

2 Principles of Revising the JACET8000

In revising the JACET8000, the committee adopted two principles: 1) the revised word list should be valid as an English word list for Japanese university students to learn; and 2) it should be able to be tested objectively by other researchers and educators. A valid English word list for Japanese university students should reflect not only high-frequency words in large authentic corpora, but also words appearing in the materials these learners generally use in their study of English. Thus, our revision was made by referring to both authentic English corpora and English materials that Japanese learners have been and will be exposed to. With respect to the second principle of testability of the compilation process, we decided to adopt objective criteria for the selection of words. Ishikawa (2012) mentions three objective indicators in compiling a basic word list: frequency, range, and familiarity. We decided to adopt frequency and range information as the resources for the selection of words in the new list. Due to the lack of reliable information, familiarity with the words was not employed in compiling the word list. Considering these two principles, we decided to make an entirely new word list rather than partly modifying the original JACET8000.

2.1 Compiling the New JACET8000

In the first stage, we obtained primary word frequency data from the Corpus of Contemporary American English (COCA) and the British National Corpus (BNC) and created the base list of 15,000 words that appear in all five registers (fiction, magazine, spoken, news, and academic) in both the COCA and the BNC. Next, we created eleven supplementary mini corpora; six consisted of materials that Japanese university students may have used in their study of English, while five contained materials the students will utilize in their studies. The first six mini corpora included authorized junior and senior high school English textbook corpora, public high school English entrance examination corpora, STEP Second to Fifth grade test corpora, National Center English Test corpora, and English monolingual dictionary vocabulary lists. The second five mini corpora included TOEFL, TOEIC, STEP Pre-first grade test corpora, English newspaper corpora, and introductory academic book corpora. In the second stage, we calculated the range indices of words in the first six mini corpora and moved words with range indices of six to four to the highest ranks in the 15,000-word base list. With this adjustment, the first 2,188 words of the New JACET8000 were determined. In the third stage, we selected words following the first 2,188 words to be moved upward in the 15,000-word base list. We referred to the range information from all eleven mini corpora and moved upward 3,553 words which have range indices between eight and two. Through this process, 5,711 words were chosen. In the final stage, we selected the remaining 2,259 words according to the original rankings of the words in the base list. We simply adopted the 2,259 highest ranking words in the base list that had a range index of one or none in the eleven mini corpora. They include such words as aye (1332), bloody (1481), correspondent (1770), constitution (1878), cricket (1977), bind (1988), hence (2004), and shit (2103) (numbers in parentheses represent original rankings in the base list). These words frequently appear in the BNC and COCA but not in the eleven mini corpora. With these adjustments, 8000 words were selected for the new word list.

3 Development of Academic Vocabulary List

Along with the compilation of the New JACET8000, a supplementary academic vocabulary list was also developed in response to a growing demand for teaching EAP in Japanese colleges. For the development of our academic vocabulary list, 52 introductory academic textbooks were collected from eight different academic fields: Agriculture, Biology, Chemistry, Engineering, Humanities, Mathematical and Physical
Sciences, Social Science and Medicine, Dentistry, and Pharmacy. These academic fields were selected based on the classification list of academic fields used for Grants-in-Aid for Scientific Research (Kakenhi) by the Japan Society for the Promotion of Science. In order to build our academic textbook corpus, the text data of these textbooks were electronically extracted and stored in eight sub-corpora according to their academic fields. The total number of words was approximately 11 million, and the total number of different words (type) was about 220,000. Word frequency analysis was performed for each sub-corpora, and the list of high-frequency words which achieved a cumulative text coverage of 97% was generated for each field. In order to select the primary candidates for our academic word list, the distribution of individual word (range) in the eight academic fields was examined, and the words which commonly appear in more than four different fields were selected from the word lists. After excluding the words included in the first basic 2,188 words of the New JACET8000 from the primary candidates, 2,194 words were finally selected as the New JACET8000 Academic Vocabulary List. The words in the list are classified into four levels (Levels 1-4) according to the range of the word in the eight academic fields (Table 1). For example, the Level 1 words appear in all eight academic fields, which suggests these words can be regarded as commonly important in any of the academic fields in our study. Likewise, the Level 2 words appear in seven fields, the Level 3 in six fields, and the Level 4 in five fields. This four-level system allows users of the list to choose which level of words to learn according to their current vocabulary level and needs for vocabulary learning.

<table>
<thead>
<tr>
<th>Vocabulary Level</th>
<th>Range</th>
<th>Number of Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>702</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>403</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>521</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>568</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>2,194</td>
</tr>
</tbody>
</table>

Table 1. New JACET8000 Academic Vocabulary List.

4 Conclusion

The New JACET8000 was developed based on the idea that Japanese college students should be provided with an original English word list tailored for their EFL learning environment. Therefore, our word list aims to contain the words commonly used in the materials that Japanese university students have used and will use in their study of English. However, in order to verify the effectiveness and validity of the list in other EFL/ESL contexts, we need to examine our list by comparing it with other major word lists.

References

JACET Basic Words Revision Committee (Eds.). (2003). JACET list of 8000 basic words (JACET8000). Tokyo: JACET.
Is the Highlighting of the Target Words in Captions Beneficial to Vocabulary Learning? An Eye Tracking Study

Olena Vasylets  
Universitat de Barcelona  
vasylets -at- ub.edu  

Roger Gilabert  
Universitat de Barcelona  
rogergilabert -at- ub.edu  

Marina Ruiz Tada  
Universitat de Barcelona  
marinaruiztada -at- gmail.com  

Ferran Gesa  
Universitat de Barcelona  
ferran.gesa -at- ub.edu  

1 Introduction

The meta-analysis by Montero Perez, Van Den Noortgate and Desmet (2013) reported that captioned videos (L2 audio together with L2 on-screen text) were more effective for vocabulary learning than the no-captions condition. These results, however, must be viewed with caution provided the limited number (N=10) of primary studies reviewed in this analysis. Diverging results have also been obtained in more recent empirical investigations, which found little evidence (Montero Perez, Peters & Desmet, 2015) or even the absence (Bisson, Van Heuven, Conklin & Tunney, 2014) of vocabulary acquisition in the captioning condition. Such conflicting findings clearly show the need for more research into captions. It is also unclear if input enhancement can significantly foster the effectiveness of captions (e.g., Montero Perez, Peters, Clarebout & Desmet, 2014). Besides, to our knowledge, there is only one published study (Montero et al., 2015) which has combined eye tracking with vocabulary tests to explore the role of attention-enhancing techniques in the context of captions.

2 The Current study

The goal of this study is to explore if textual enhancement in captions brings any advantages for vocabulary learning. The participants were N=30 L1 Spanish/Catalan L2 English learners (upper-intermediate L2 proficiency). The group was homogeneous in terms of the vocabulary size (X_Lex, Y_Lex tests by Meara and Miralpeix, 2003). Participants were randomly divided into 3 equal groups. Each group watched the experimental video under one of the following conditions: (i) control condition (L2 audio only); (ii) non-highlighted condition (L1 audio and L2 captions); (iii) highlighted condition (L2 audio and L2 captions in which N=12 key words had been visually highlighted). The data were gathered during individual sessions of data collection. We first tested each subject’s knowledge of the key words. Then each subject was eye tracked while watching, under one of the conditions, a 25-minute episode from a TV series. After the video, we tested if there were gains in the learning of the key words. For the post-tests, we used: (i) a form recognition test; (ii) meaning recall; (iii) meaning recognition (identical to the pre-test); (iv) content comprehension. Except for content comprehension, all the tests were in the combined aural/written mode.

The mixed-design ANOVA showed that there was a significant improvement from the pre- to the post-meaning recognition test, $F(1,27) = 161.042, p = .000$, partial eta squared = .856. There was also a tendency towards a significant interaction between time and group, $F(2,27) = 3.109, p = .061$, partial eta squared = .187. Follow-up analyses showed that there were no differences between the groups at the baseline. But, in the post-test, the highlighted group significantly outperformed the control group ($p = .030$) and also showed a tendency to score higher than the non-highlighted group ($p = .063$); there were no differences between the non-highlighted and control groups.

In the form recognition post-test, the groups differed significantly, $F(2,29) = 8.301, p = .002$, $\eta^2 = .38$. Post-hoc comparisons using the Tukey HSD test indicated that the mean score for the highlighted group differed significantly from both the non-highlighted group ($p = .017$) and the control group ($p = .002$), with the highlighted group always scoring higher; there were no differences between the non-highlighted and control groups. In the meaning recall test, the descriptives showed the highest scores for the highlighted group ($M = 7; SD = 1.24$), followed by the non-highlighted group ($M = 6.51; SD = 2.54$) and the control group ($M = 6.1; SD = 2.18$). However, these differences were not statistically significant, $F(2,29) = .476, p = .627, \eta^2 = .03$. Table 1 summarizes the results of the vocabulary tests.
In the analysis of the eye tracking data, results revealed a significant main effect of the viewing condition for the number of fixations ($F(2,29) = 30.406, p = .000, \eta^2 = .692$) and the number of visits (time between the 1st fixation and exit from the area) ($F(2,29) = 41.227, p = .000, \eta^2 = .753$). Tukey post hoc analyses revealed that the participants had more fixations and more visits in the area of captions under the captioning conditions as compared to the control condition, all $ps = .000$. This result may be taken as evidence that the participants actually read the text when it was available on-screen. However, the two captioning conditions did not significantly differ either in terms of fixation duration, number of fixations, or the number of visits. This could indicate that, when the overall captioning area was taken into account, the participants’ viewing/reading behaviors were largely similar in the two captioning conditions.

The fact that the vocabulary gains of the non-highlighted group were equal to the gains of the control group suggests that the mere presence of the on-screen L2 text does not necessarily confer an advantage in vocabulary acquisition. On the other hand, making the words salient in captions seems to bring about significant learning advantages. These advantages, however, seem to extend only to receptive vocabulary learning. This restriction might be attributed to the shallow level of processing of the salient elements in captions. According to Sharwood Smith (1991), intake and potential learning of an enhanced element depend on the level (deep versus shallow) with which this element is processed in the input. Empirical evidence supports this theory, showing that deeper processing leads to superior information retention and lexical learning (see Leow (2015) for the recent review). Deep processing of the enhanced elements in the context of captions might be difficult to attain because of the speed with which the text lines change on the screen. Additionally, the dynamic images can divert learners’ attention from the captions. Saliency of textual elements might counteract these effects and help attract enough attention to ensure at least the shallow level of processing, which, nevertheless, appears to be sufficient to facilitate the initial receptive stages of lexical learning in the context of captions.

3 Acknowledgments

We would like to thank the GRAL (Language Acquisition Research Group: 2014SGR1089), based at the University of Barcelona, and we would also like to acknowledge the support of the Research Grant FFI2013-47616-P from the Spanish Ministry of Economy and Competitiveness.

References


Word Frequency Dominance and L2 Word Recognition

Yu Tamura  
Nagoya University  
yutamura -at- nagoya-u.jp

Mitsuhiro Morita  
Hiroshima University  
mmorita -at- hiroshima-u.ac.jp

Yosito Nishimura  
Nagoya University  
nishimura -at- nagoya-u.jp

1 Background

The debate over the representation and processing of lexical items, both in the first language (L1) and in the L2, is ongoing. For L1 studies, there are three different positions explaining the mechanism for regularly inflected words. These are the connectionist full-form storage model (MacClelland & Rumelhart, 1981), obligatory decomposition (Taft, 2004) and the dual-route model (Baayen et al., 1997). The processing mechanism and representation of regularly inflected lexical items in the case of L2 learners has yet to be thoroughly investigated. Therefore, the present study investigated how L2 learners of English process and represent regularly inflected words in their mental lexicons from the viewpoint of the word frequency effect.

2 Present study

We conducted a lexical decision task (LDT) to determine the influence of frequency dominance on singular and plural nouns. We operationalized singular dominant nouns as the occurrence of singular forms being twice or more frequent than the corresponding plural forms, and plural dominant nouns as the plural forms occurring twice or more frequently than the singular forms. Overall, singular dominant nouns and plural dominant nouns ($k = 18$ for each type) were chosen on the basis of frequency information derived from the British National Corpus. In addition, 18 nouns whose singular and plural forms have nearly the same frequencies were selected as control items. Cumulative frequency—the sum frequencies of singular and plural forms—of the three types of nouns was controlled. The mean frequencies of the three types of the target nouns are summarized in Table 1. Table 2 shows all the target items used in the LDT.

<table>
<thead>
<tr>
<th>Types</th>
<th>Singular</th>
<th>Plural</th>
<th>Cumulative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singular-dominant</td>
<td>69.865</td>
<td>21.684</td>
<td>91.549</td>
</tr>
<tr>
<td>(25.849)</td>
<td>(10.931)</td>
<td>(34.342)</td>
<td></td>
</tr>
<tr>
<td>Plural-dominant</td>
<td>22.571</td>
<td>69.898</td>
<td>92.469</td>
</tr>
<tr>
<td>(18.661)</td>
<td>(43.345)</td>
<td>(59.779)</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>47.064</td>
<td>43.893</td>
<td>90.958</td>
</tr>
<tr>
<td>(23.202)</td>
<td>(24.664)</td>
<td>(46.185)</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Mean Frequency (per million) and SD (in parentheses) of the Target Nouns.

<table>
<thead>
<tr>
<th>Singular-dominant</th>
<th>Plural-dominant</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>concept</td>
<td>image</td>
<td>topic</td>
</tr>
<tr>
<td>film</td>
<td>ball</td>
<td>rabbit</td>
</tr>
<tr>
<td>science</td>
<td>target</td>
<td>bone</td>
</tr>
<tr>
<td>jacket</td>
<td>video</td>
<td>store</td>
</tr>
<tr>
<td>box</td>
<td>hat</td>
<td>principle</td>
</tr>
<tr>
<td>colour</td>
<td>map</td>
<td>horse</td>
</tr>
<tr>
<td>bar</td>
<td>context</td>
<td>rule</td>
</tr>
<tr>
<td>network</td>
<td>station</td>
<td>function</td>
</tr>
<tr>
<td>college</td>
<td>tower</td>
<td>plant</td>
</tr>
</tbody>
</table>

Table 2. List of Target Items.

In total, 72 Japanese undergraduate students participated in the study. In the task, a fixation point first appeared on the screen for 1,000 ms and then the stimuli were presented. Participants were asked to judge whether or not the presented words were real English words. Only the reaction times (RT) for the correct responses were analyzed. Before analyzing the RT data, outliers (M+3SD of the individual’s RT and responses below 200 ms) were removed, which comprised 1.4% of all the data. The data were then submitted to generalized linear mixed-effects models (GLMMs) with crossed-random
effects (subject and item intercept). The response variable was the row RT data, and the explanatory variables were the two categorical variables: noun type (singular-dominant, plural-dominant, and control), and presentation condition (singular form and plural form). We compared the Gamma distribution and inverse Gaussian distribution together with the identity link function and adopted the latter, which produced the lower AIC.

3 Results

The descriptive statistics of the RT are shown in Table 3. The results of GLMM suggested interaction between the type of noun and the presentation condition. Therefore, we carried out post-hoc multiple comparisons of the RTs. The results of multiple comparisons suggested that the participants judged singular forms faster than plural forms irrespective of frequency dominance. In addition, a comparison of three types of nouns in plural form showed that plural-dominant plurals, which should have demonstrated frequency advantages over the other two types, were the slowest of the three. Similarly, singular-dominant singulars did not demonstrate a frequency advantage over the control items, although they were faster than the plural-dominant singulars.

<table>
<thead>
<tr>
<th>Types</th>
<th>Presentation</th>
<th>M</th>
<th>SD</th>
<th>95%CI</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LL</td>
<td>UL</td>
</tr>
<tr>
<td>Control</td>
<td>plural</td>
<td>838</td>
<td>246</td>
<td>818</td>
<td>858</td>
</tr>
<tr>
<td></td>
<td>singular</td>
<td>765</td>
<td>232</td>
<td>747</td>
<td>783</td>
</tr>
<tr>
<td>Plural-dominant</td>
<td>plural</td>
<td>922</td>
<td>324</td>
<td>896</td>
<td>949</td>
</tr>
<tr>
<td></td>
<td>singular</td>
<td>857</td>
<td>288</td>
<td>834</td>
<td>880</td>
</tr>
<tr>
<td>Singular-dominant</td>
<td>plural</td>
<td>824</td>
<td>280</td>
<td>802</td>
<td>846</td>
</tr>
<tr>
<td></td>
<td>singular</td>
<td>719</td>
<td>212</td>
<td>702</td>
<td>735</td>
</tr>
</tbody>
</table>

Table 3. Descriptive statistics of the RT.

Note. CI = Confidence Interval; LL = lower limit, UL = upper limit

4 Discussion

Given that the plural-dominant plurals were the slowest of all six conditions, it appears that the Japanese L2 learners of English did not directly access the plural forms of the plural-dominant nouns. In other words, the learners were likely to have decomposed the plural forms and accessed the singular forms of the nouns regardless of frequency dominance.

Singular-dominant singulars should have advantages in terms of morphological simplicity as compared with plural forms. We investigated whether they would be accessed faster than other singular forms, given the frequency advantage of singular-dominant singulars. The results were unclear because singular-dominant singulars were accessed faster than the plural-dominant singulars, but not the control singulars. One plausible interpretation is that the Japanese L2 learners of English likely accessed the singular forms of the nouns stored in their mental lexicon. Nonetheless, because the singular-dominant singulars were not accessed faster than the control singulars, the learners may have accessed more abstract lexical entries that include both singular and plural forms.

References


The NGSL, NAWL, TSL, and BSL Word Lists: Corpus-based Word Lists and Tools for Success

Charles Browne
Professor of Applied Linguistics, Meiji Gakuin University
browne -at- ltr.meijigakuin.ac.jp

1 Introduction
This paper introduces four corpus-derived lists of vocabulary words as well as a large and growing number of free online tools designed to help students, teachers, researchers and authors to be able to better utilize and learn from these word lists.

2 New General Service List (NGSL)
Our first list, The New General Service List, or NGSL (Browne, 2013), is a major update of the original General Service List, which was a list of about 2,000 commonly used words based on a corpus of written English compiled by West (1953). Although the original GSL was a remarkable, pre-computer era, corpus-derived list of important high-frequency words for second language learners that has been used for more than 60 years, the corpus it was based on is now considered to be quite dated (most texts in the corpus were published in the 1800s to early 1930s), small by modern standards (the original analysis was done with a corpus of only 2.5 million words), and in need of a clearer definition of what constitutes a “word” within the list.

In February 2013, on the 60th anniversary of West’s publication of the GSL, my colleagues and I published the New General Service List and put up an NGSL website (www.newgeneralservicelist.org) that provides various downloadable versions of the list as well as a growing number of online tools. This list was derived from a carefully selected 273 million-word subsection of the 2-billion-word Cambridge English Corpus (CEC) and initial results indicate that the NGSL provides substantially better coverage for general English texts than West original GSL (see Table 1 below).

3 New Academic Word List (NAWL)
The same is true for our NAWL or New Academic Word List (Browne, Culligan and Phillips, 2013), which is an update of Averil Coxhead’s excellent Academic Word List (AWL, Coxhead, 2000). The NAWL is based on a corpus of 283 million words of academic texts consisting of 248 million words form the academic subsections of the CEC, 3 million words of spoken academic English from the Michigan Corpus of Academic Spoken English (MICASE), and the British Academic Spoken English (BASE) corpus and 36 million words from top-selling academic textbooks. As can be seen in Table 1 below, the NAWL (when combined with the NGSL), provides about 5% more coverage for most academic texts than the original AWL does when combined with the original GSL.

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Size</th>
<th>GSL</th>
<th>NGSL</th>
<th>GSL/AWL</th>
<th>NGSL/NAWL</th>
</tr>
</thead>
<tbody>
<tr>
<td>General</td>
<td>273m</td>
<td>84%</td>
<td>92%</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Academic</td>
<td>283m</td>
<td>x</td>
<td>x</td>
<td>87%</td>
<td>92%</td>
</tr>
</tbody>
</table>

Table 1: NGSL and NAWL Coverage for General and Academic Corpora

About 6 months after the release of our NGSL, another list of high frequency words for general English was published by Brezina and Gablasova (2013), hereafter referred to as the Other New General Service List (ONGSL). Although an excellent bit of research designed to identify a core list of words common among a range of well-known corpora, coverage for the ONGSL is well below that of the NGSL (whose primary purpose is more focused on the pedagogic needs of 2nd language learners than it is on research purposes). This is true not only for general English but even for specialized genres. For example, GSL, NGSL and ONGSL coverage figures for two large corpora of popular business and science publications can be seen in Table 2 below.

<table>
<thead>
<tr>
<th>The Economist</th>
<th>Scientific American</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSL</td>
<td>76.55%</td>
</tr>
<tr>
<td>ONGSL</td>
<td>78.30%</td>
</tr>
<tr>
<td>NGSL</td>
<td><strong>81.75%</strong></td>
</tr>
<tr>
<td>Corpus Size</td>
<td>27.3 million</td>
</tr>
</tbody>
</table>

Table 2: Comparison of the GSL, NGSL, and ONGSL Word Lists for Various Text Genres
4 TOEIC Service List (TSL)

In May of 2016, Dr. Browne and Dr. Culligan published another word list that can be studied in conjunction with the New General Service List, known as the TSL or TOEIC Service List. Designed to help students trying to pass the high stakes TOEIC test and based on a 1.5 million word corpus of TOEIC test materials, the 1200 words of the TSL (when combined with the 2800 words of the NGSL) provides up to 99% coverage of words that appear on the TOEIC test.

5 Business Service List (BSL)

In July of 2016, Dr. Browne and Dr. Culligan published yet another word list that can be studied in conjunction with the New General Service List, known as the BSL or Business Service List. Designed to help students to comprehend general business English texts and based on a 64-million-word corpus of business texts, newspapers, journals and websites, the 1700 words of the BSL (when combined with the 2800 words of the NGSL) provides up to 97% coverage of business English texts.

6 Free online tools and resources

From the very beginning, our focus has been less on publishing academic papers than it has been on creating open-source lists of high-frequency words and a wide range of free online tools that would make these wordlists useful to second language learners of English (as well as teachers, materials developers and researchers), and to this end, we have thus created a dedicated website (www.newgeneralservicelist.org) which gathers all associated NGSL, NAWL, TSL resources in one place. A short list of some of the free resources we provide includes:

- Downloads of various forms of each list (e.g. lemmatized, headwords only, etc)
- Downloads of original definitions written in easy English & Japanese for all wordlists
- Downloads of various research articles published on these wordlists by the authors
- Links to various free flashcard websites such as Quizlet.com and Memrise.com where we have posted our word lists
- Links to free iPhone and Android flashcard apps we have created for the lists
- Links to reliable and valid vocabulary tests for the NGSL and NAWL which help students find the frequency bands they are weakest at and link to Quizlet.com flashcards to help them quickly learn their missing words
- Link to text analysis tools such as VocabProfile, AntwordProfiler and OGTE which allow you to instantly analyze the text coverage provided by our word lists as well as to help you to modify (ie simplify) them for your students

References


Stimulating the Intake of Formulaic Sequences in Texts: The Case of Typographical Enhancement

Helene Stengers
Vrije Universiteit Brussel
helene.stengers -at- vub.ac.be

Julie Deconinck
Vrije Universiteit Brussel
julie.deconinck -at- vub.ac.be

June Eyckmans
Universiteit Gent
june.eyckmans -at- ugent

1 Introduction

In the last couple of decades, linguistics and applied linguistics research have been characterised by a growing interest in the phenomenon of phraseology, also referred to as formulaic language. Formulaic language learning has been proposed as a key element in achieving nativelike production and in the language acquisition process at large (e.g. Boers et al., 2006; Siyanova-Chanturia, Conklin & Van Heuven, 2011; Sonbul, 2015). However, research suggests that acquisition of foreign language formulaic sequences occurs at a very slow pace (e.g. Laufer & Waldman, 2011; Li & Schmitt, 2010; Peters, 2014). Among the explanations for this slow pace are insufficient encounters with the same formulaic sequence in a short span of time, lack of attention given by learners to formulaic sequences that are semantically transparent or consisting of familiar words, and insufficient cognitive processing of the kind that fosters retention. While phraseological teaching approaches and materials are in vogue, the limited amount of instruction time in non-immersive learning contexts is bound to be insufficient to meet the challenge of mastering the formulaic dimension of a foreign language. Therefore, pedagogical interventions aiming to stimulate incidental intake from authentic texts that learners read autonomously, i.e. outside the language classroom, are also called for. This study assesses the effectiveness of a textual enhancement method aiming to direct learners’ attention to the formulaic sequences they encounter in textual input. Even though a given formulaic sequence may occur just once in an authentic text, textual enhancement has the potential to promote the learner’s noticing and thus intake of the formulaic sequence. The following research questions are addressed:

- Does typographic enhancement of formulaic sequences foster their intake?
- Does typographic enhancement of formulaic sequences foster intake of other formulaic sequences from the text beyond those that are enhanced?

2 Method

Selected formulaic sequences were made salient in texts through the use of typographic enhancement by underlining them. Two cohorts of EFL learners (n=38 and n=43), English majors at universities in the Flemish part of Belgium, read two 400-word texts in one of three versions: (a) with 16 target formulaic sequences underlined, (b) with half of these formulaic sequences underlined, and (c) without any underlining. The learners were subsequently asked to identify the formulaic sequences they remembered encountering in the texts. Text version (b) was included to explore whether enhancement of a limited number of formulaic sequences in a text also stimulates intake of others from that text. The texts were checked to be within learners’ reading competence. The EFL learners’ average level of proficiency in the second trial was higher than that of the learners in the first trial. Whether a formulaic sequence encountered in a text was temporarily taken in by the learner or not was operationalized as the learner’s ability to identify the formulaic sequence in a recognition test. The target formulaic sequences included verb-noun collocations, adjective-noun collocations, nominal compounds, and prepositional phrases. In the recognition test, these target phrases were juxtaposed with a synonymous phrase, and the task for the students was to identify the wording they remembered seeing in the text. Per test item, a ‘neither (a) or (b)’ response option was provided to reduce blind guessing.

3 Results

The descriptive statistics for the three conditions’ scores per text and in each trial are presented in Tables 1 and 2. To assess the overall effect of text version on verbatim memory for formulaic sequences, we applied 3X2 ANOVAs to the three groups’ scores on the recognition tests.
The results support the thesis that typographical enhancement can indeed prompt learners to pay attention to the lexical makeup of enhanced formulaic sequences. However, there was no evidence that it also fostered the intake of unenhanced formulaic sequences.

An intervention such as typographical enhancement is relatively simple for teachers and materials developers to implement. However, further research should shed more light on the amount of formulaic language which can be highlighted without compromising the saliency effect of the enhancement, as well as on the impact of typographic enhancement in the domain of reading beyond the episodic memory, i.e. on learners’ development of implicit or procedural knowledge of formulaic sequences.

References


The Role of Vocabulary Learning Strategies in Lexical Progression in an EFL Context

Nuzhat Jafari
University of Bedfordshire
Nuzhat.jafari -at- beds.ac.uk

1 Introduction

This study explores the role of vocabulary learning strategies (VLSs) and other factors behind the lexical progression of adult learners in English as a second language (ESL) context. The lexical progression of ESL learners can derive from learning vocabulary in the class and outside the class. Vocabulary learning, which is a sub-goal of language learning, is essential to achieve other language learning goals (Nation, 2001), since knowledge of vocabulary contributes a very great deal to overall language learning success (Schmitt, 2010).

Research has shown that language learners use more strategies in vocabulary learning than in any other linguistics competences (Schmitt, 1997). VLSs are a part of language learning strategies (Nation, 2001) and can be defined as specific actions taken by language learners to make learning easier, faster, more enjoyable, more self-directed, more effective and more transferable to new situations (Oxford, 1990). VLSs can be described as the process by which information is obtained, stored, retrieved and used (Rubin, 1987; Schmitt, 1997).

During the past few decades, a number of studies have been carried out to describe the use of VLSs in different ESL contexts. However, most of these studies dealt with individual or small numbers of strategies rather than looking at all strategies as a whole (Schmitt, 1997). Additionally, there has hardly been any longitudinal study conducted on curricular and extra-curricular VLSs in Pakistani ESL contexts. To partially fill such a gap in literature, this study aims to explore vocabulary learning strategies in lexical gain in the Pakistani university context. The research questions addressed in this research are:

1. What are the curricular and extra-curricular VLSs adopted by Pakistani tertiary students to learn English vocabulary?
2. What is the impact of the curricular and extra-curricular VLSs on vocabulary gain in this context?

2 Methods

Due to the aims and nature of this project, this study was designed as a large-scale, longitudinal mixed-methods study which combined both quantitative and qualitative techniques within the project. Schmitt (2010) pointed out in this regard that “vocabulary learning is longitudinal and incremental in nature, and only research design with longitudinal element can truly describe it” (p.156). The participants of this study were Pakistani university students, aged between 18-24, and had just started their degree.

Two types of vocabulary test were administered to 578 participants twice as pre- and post-tests with an interval of 52 weeks. The Productive Vocabulary Levels Test (PVLT) (Lauffer & Nation, 1999) containing 72 items was used to examine the learners’ general vocabulary gain. The self-devised Productive Course Vocabulary Test (PCVT) modified from Read (2000) and Nation (2001) was used to assess the learners’ vocabulary progress specific to the course they had taken, and the test contained 30 items. Since it was a subjectively marked test, the marking scheme was informed by a series of expert focus group discussions, and 4 trained raters marked the PCVT tests to ensure inter-rater reliability.

120 semi-structure interviews, 120 structured diary (SD) reports (four weeks) and 578 VLSs questionnaires (e.g., Brown, 1977; Gu & Johnson, 1996; Schmitt, 1997; Harris & Snow, 2004) were administered to explore VLSs adopted by students to learn vocabulary. In the second phase of the data collection, after the post-vocabulary tests, all 578 students were divided into four groups based on their general and course-related vocabulary gain in test scores. These groups were top in both, bottom in both (general and course-related vocabulary), top only in general vocabulary, and top only in course-related vocabulary. Then 30 participants from each group were selected randomly for the interviews and SD reports.

3 Findings

After confirming internal consistency of the tests and questionnaires, Wilcoxon Signed Rank Tests were carried out to examine the learners’ vocabulary gain during the 52 weeks’ period, as measured by the two types of test. The results showed a significant gain in productive vocabulary knowledge within the twelve-month period of this study. Factor analysis was conducted to understand the nature of strategies used by the learners.
The two-factor solution explained a cumulative of 74.23% with factor one contributing variance to 65.84% and factor two to 8.19%. Factor one (77 items) was named as curricular VLSs and factor two (33 items) was named as extra-curricular VLSs. Using the factor structures identified and the learners’ vocabulary test scores, multiple regression analyses were performed to assess the impact of strategy use on the learners’ vocabulary gain. Extra-curricular VLSs (Std.Beta=0.498, p<0.001) best predicted overall general vocabulary gain, and curricular VLSs (Std.Beta=0.159, p<0.001) were also significant positive predictors. Curricular VLSs (Std.Beta=0.682, p<0.001) were the best predictor for overall course-related vocabulary gain, but extra-curricular VLSs (Std.Beta=-0.309, p<0.001) emerged as significant negative predictors of overall course-related vocabulary gain. Kruskal Wallis tests and Post-hoc comparisons with Bonferroni adjustment identified significant differences among the six groups of students in terms of the number of strategies used (Curricular VLSs p<0.001; extra-curricular VLSs p<0.001).

The interviews and DS reports were analysed to triangulate and elaborate on the quantitative results, and they offered in-depth information with the full richness and complexity of views held by the respondents related to adopted VLSs to learn vocabulary. Findings showed that meta-cognitive strategies such as self-initiatives and selective-attention played a significant role in gaining both general and course-related vocabulary. Students who achieved the highest gains in both general and course-related vocabulary mentioned that they applied a wide range of curricular VLSs, e.g., guessing strategies, dictionary strategies, note-taking strategies, memory strategies, and activation strategies to learn English vocabulary. They also indicated their usage of extra-curricular VLSs, e.g., reading English literature, newspaper and magazines, watching English TV, watching and listening to English news, music and movies, and having interaction with native speakers of English.

4 Summary of the results

Participants had significant vocabulary gains in both general and course-related vocabulary tests. Extra-curricular VLSs turned out to be the best predictor of general vocabulary gains whereas curricular VLSs were the best predictor of course-related vocabulary gains. VLSs questionnaire data indicated adopted patterns of vocabulary learning strategies, i.e., higher achievement learners used more VLSs than low-achievement learners. Interviews and Structured diary reports data indicated complex nature of VLSs use, e.g., use of certain VLSs in particular contexts; two or more strategies used in combination. For instance, the students who used dictionary strategies without guessing strategies were less successful as compared to those who combined both (i.e., first guessed the meaning from context, and then used dictionaries just to confirm the meaning). The students who applied only guessing and dictionary strategies had limited vocabulary gains, as compared to those who applied guessing and dictionary strategies and also made notes of these vocabulary items and memorized these items by applying memory strategies. The students who progressed in both general and course-related vocabulary tended to use a variety of strategies in combination, and their balanced and integrated approach seemed the most effective in vocabulary learning in this context.

References

The Vocabulary of Graded Readers

Brandon Kramer
Osaka Jogakuin University &
Junior College
Brandon.L.Kramer -at- gmail.com

Stuart McLean
Osaka Jogakuin University
Kansai University
stuart93 -at- me.com

1 Introduction

Graded readers are designed to make L2 reading accessible and enjoyable by telling stories using a limited range of vocabulary adapted for varying levels of student proficiency. While the books within each series are levelled according to the number of headwords included in their creation, the specific vocabulary used are unique to each series or publisher. To better understand the vocabulary within these texts, this presentation describes the creation of a graded reader corpus, the subsequent extraction of a lemmatized word frequency list, and the results of comparisons between this graded reader vocabulary list with a list produced from a large corpus of general English.

While also dependent on a wide range of other factors, the minimum lexical knowledge requirements for reading comprehension are usually defined as knowing between 95% (Laufer, 1989) to 99% (Carver, 1990) of the words within a text. In order to provide evidence that this minimum lexical coverage has been achieved, researchers and teachers often first measure learners’ vocabulary knowledge using tests which sample from frequency lists built from a large representative corpus, comparing these results with the lexical coverage of those texts using frequency bands of the same corpus.

Are the general corpora used for these analyses representative of the vocabulary in graded readers, however? Using such analyses, for example, previous studies have found that learners need to understand the first 2,000-3,000 word families in order to understand texts written for the most beginner learners (e.g. Webb & Macalister, 2013; McLean, 2014). In addition, some recent studies are casting doubt on the idea that such large corpora are truly representative of learner knowledge, calling into question their use as a proxy measures in this way (see Pinchbeck, McLean, Brown, & Kramer, 2016, at this conference).

Rather than relying on indirect methods such as these, others such as Horst (2005) suggest using corpora built from the graded readers themselves, giving more accurate and precise information about the words students are exposed to. While previous studies have collected a limited number of graded reader texts for analysis, there remains a lack of a sufficiently representative sample of texts including a wide range of publishers, reading levels, and text genres. This paper describes the initial creation of such a corpus and a list of the vocabulary within which attempts to address this gap in previous research.

2 Research questions

To address the previously mentioned gaps the following research questions were created:

1. Which words occur frequently and uniformly throughout a large sample of graded reader texts?
2. How do these words compare with a large corpus of general English?

3 Corpus and list creation

In an effort to be sufficiently representative, texts were collected from a range of publishers and reading levels, here defined as mid-beginner to high-advanced within the Extensive Reading Foundation Grading Scale (based on publisher headword counts), including a mix of both fiction and non-fiction texts. The texts were all manually scanned and digitally prepared for optical character recognition analysis and text extraction. The corpus currently consists of 860 total texts, containing over 5 million running words, and will grow as new graded readers are added. Considering that graded reading texts are intended for beginning learners without such advanced morphological awareness, this study utilizes the lemma unit for the creation of its vocabulary frequency list, which makes fewer assumptions of prior learner knowledge than other units such as the word family by only grouping morphological inflections. This lemmatized list of ER vocabulary was derived using average relative frequency statistics and a minimum range of 3 publisher subcorpora. In order to compare these words with those from a more general corpus of English, a lemmatized frequency list was compiled from available COCA and BNC word type frequency lists.

4 Results

The extracted word list currently contains 5,630 lemmas, and gives some insight into what words learners
will likely encounter when reading or listening to graded texts. It is typically advised to compare the vocabulary found in target texts to a reference corpus in order to infer how many of the running words are likely to be known by the learner. Thus, a direct comparison was made between the newly created graded reader corpus list and a combined COCA/BNC lemmatized frequency list. Although more graded reader text data is continually being added to the corpus, we can see in Table 1 that the frequency of words within these texts is often very different from the frequency within the reference list representing general English use.

<table>
<thead>
<tr>
<th>ER Bands</th>
<th>1k</th>
<th>2k</th>
<th>3k</th>
<th>4k</th>
<th>5k</th>
<th>6k</th>
<th>7k</th>
<th>8k</th>
<th>9k</th>
<th>10k</th>
<th>OFF</th>
<th>#N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>1k</td>
<td>550</td>
<td>217</td>
<td>82</td>
<td>32</td>
<td>13</td>
<td>13</td>
<td>9</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>6</td>
<td>72</td>
</tr>
<tr>
<td>2k</td>
<td>200</td>
<td>222</td>
<td>173</td>
<td>113</td>
<td>79</td>
<td>43</td>
<td>30</td>
<td>20</td>
<td>15</td>
<td>8</td>
<td>44</td>
<td>53</td>
</tr>
<tr>
<td>3k</td>
<td>103</td>
<td>173</td>
<td>149</td>
<td>137</td>
<td>87</td>
<td>71</td>
<td>50</td>
<td>39</td>
<td>34</td>
<td>13</td>
<td>66</td>
<td>78</td>
</tr>
<tr>
<td>4k</td>
<td>40</td>
<td>139</td>
<td>129</td>
<td>125</td>
<td>104</td>
<td>73</td>
<td>74</td>
<td>52</td>
<td>35</td>
<td>27</td>
<td>106</td>
<td>96</td>
</tr>
<tr>
<td>5k</td>
<td>14</td>
<td>52</td>
<td>99</td>
<td>98</td>
<td>97</td>
<td>89</td>
<td>77</td>
<td>60</td>
<td>50</td>
<td>49</td>
<td>184</td>
<td>131</td>
</tr>
<tr>
<td>6k</td>
<td>4</td>
<td>16</td>
<td>33</td>
<td>23</td>
<td>47</td>
<td>43</td>
<td>34</td>
<td>31</td>
<td>29</td>
<td>28</td>
<td>220</td>
<td>122</td>
</tr>
<tr>
<td>Total</td>
<td>911</td>
<td>819</td>
<td>665</td>
<td>528</td>
<td>427</td>
<td>332</td>
<td>274</td>
<td>206</td>
<td>165</td>
<td>125</td>
<td>626</td>
<td>552</td>
</tr>
</tbody>
</table>

Table 1. Overlap Between 1K Lemma Frequency Bands of the GR Vocabulary List and BNC/COCA.

Discussion

A vocabulary list created from graded readers could be used in a variety of ways. It will be argued that the graded reader vocabulary list provides a more meaningful and direct representation of the target construct than lists produced from larger general corpora. Considering the rapid growth of extensive reading and listening in programs throughout the world, such an analysis is a useful step which could lead to better structured curriculums, a better understanding of the vocabulary encountered within them, and more accurate and precise measures of vocabulary knowledge in either formative or summative assessment.

References


Learning and Teaching Collocations through an Online English Collocations Workbook

Adriane Orenha Ottaiano
São Paulo State University
adriane -at- ibilce.unesp.br

1 Introduction

According to Wray (2002, p. 74), ‘an individual’s inventory of holistically stored sequences is heavily influenced by the current patterns of usage in the speech community’ and we store and use theses sequences because they ‘sound right’ to us and also because we desire to sound like others in the speech community. Learning a foreign language may require learning a large repertoire of such current patterns, to be stored in our mental lexicon, retrieved from it, and then be used automatically. Among these patterns, we may include collocations, regarding their relevance in the present sphere of foreign language learning and teaching (Meunier & Granger, 2008; Nesselhauf, 2005; Orenha-Ottaiano, 2012, 2015; Sinclair, 2004; etc.).

In line with this view, this study aims at discussing some corpus-based collocational activities specifically designed to comprise an Online English Collocations Workbook, compiled with the purpose of teaching and learning collocations. The workbook deals with all types of collocations (adjectival, verbal, noun and adverbial) and it already has a game (Memory Game) and an activity (Gap Fill) available online, the latter bringing the collocations presented in the Memory Game contextualized, thus offering different possibilities of memorization of collocations, adjusting to learners’ different learning styles as well as fulfilling their needs.

The pedagogical applications of the Online English Collocations Workbook consists of its contributions to the teaching, learning and awareness of the conventional and collocational aspects of the language.

2 Methodology for collocations extraction

This research first relies on the extraction of collocations from a Translation Learner Corpus (TLC), a parallel corpus made up of university students’ translations in the Portuguese-English direction. The original texts that comprise the corpus are newspaper articles taken from well-known Brazilian newspapers. It has approximately 100,000 words and was compiled at São Paulo State University (UNESP), in Brazil. The collocations were extracted with the help of WordSmith Tools (Scott 2008) and AntConc (Anthony, 2012), which have enabled us to raise the most frequent collocational patterns and some collocational errors. In a second stage, more collocational patterns were extracted, based on the most frequent words from The Corpus of Contemporary American English (Davies 2008-2012), using the Sketch Engine (Kilgarriff et al, 2004), with the aim of ensuring the workbook will offer users access to the most frequent and recurrent.

3 The Online English Collocations Workbook

The Online English Collocations Workbook has the aim of offering an interactive platform to teach and learn collocations, specially designed to native Brazilian Portuguese speakers who are learners of English as a foreign language.

It is comprised of a Memory game and a Gap Fill and will also have other activities to be incorporated in the near future, and also expected to be highly interactive. In the Memory game, the user has to select the number of pairs he or she would like to play (6, 8 or 10).

Both Memory Game and Gap Fill have different levels of difficulty (easy, medium and hard). The user can also choose a category (a specialized area or general language). So far, there are six categories: Academic Language, Business, General, Medicine, Politics and Investigation, as shown in Figure 1. Besides that, the Workbook offers the option of selecting the type of collocations (taxonomy): 1) Verbal collocations; 2) Nominal collocations; 3)
Adjectival collocations; 4) Adverbial and 5) Collocations with phrasal verbs. 

The collocations students worked with in the Memory Game will then be contextualized in the Gap Fill exercise, following the same pattern used before: level of difficulty, category and taxonomy.

Figure 2. Gap Fill

4 Conclusion

We hope to have shed some light on the potential benefits of collocational knowledge in English and on the importance of providing students with a teaching material which may fulfil their needs.

We also expect that the Workbook may enable the target audience to have a more systematized and interactive learning of collocational patterns, filling a gap regarding the lack of phraseological works in Brazil. Furthermore, the aspect of being online will enable it to gain greater visibility and more students and teachers to have access to it.

5 Acknowledgments

I gratefully acknowledge the financial support provided by FAPESP (Foundation for Research Support of the State of São Paulo), Process 2016/04410-7, that made my participation in this conference possible.

References


The Effect of Word Class and Word Length on L2 Vocabulary Attrition

Samuel Barclay
University of Nottingham
samuel.barclay -at- nottingham.ac.uk

1 Introduction

Previous research has shown the importance of intralexical factors in the acquisition of L1 vocabulary. For example, word class and word length have been found to affect the learnability of a lexical item. Research has also helped clarify the effect of intralexical factors on the acquisition of Foreign Language (FL) vocabulary (see Laufer, 1997). Studies have shown, for instance, that nouns are typically acquired more easily than other word classes (e.g. Ellis & Beaton, 1993). However, although much is known about the effect of intralexical factors on the acquisition of FL vocabulary, comparatively less is known about the effect of these factors on the decay of FL knowledge.

2 Why study decay?

Research into the decay of acquired language knowledge may have important implications for language pedagogy and policy (Schmitt, 2010; Van Els & Weltens, 1989). It could offer answers to a number of important pedagogical issues. These include, but are not limited to: identifying levels of competence above which the probability of language retention improves; developing strategies for the reactivation of latent language knowledge; and assessing (and ultimately predicting) the long-term effects of a language course, teaching method, teaching exercise or treatment (Van Els & Weltens, 1989). Furthermore, such research can contribute to the theoretical models of human memory, language change, and the mental lexical, as well as improve our understanding of the interface between acquisition and loss (ibid.). It is therefore important to understand more about this process.

3 Study objectives

In spite of previous research showing differential patterns of acquisition and deterioration of lexical items, we still do not have a clear picture of the effect of part of speech or word length on the decay of FL knowledge. This research therefore aimed to clarify the influence of these two factors on the decay of the recognition and recall of FL lexical knowledge.

4 Methodology

In order to meet the above objectives, participants (international students at UK tertiary institutions) firstly studied previously unknown vocabulary. Participants’ knowledge of the learnt words was assessed immediately after the learning phase in a test of attainment and subsequently in a test of retention. A secondary aim of the study was to investigate the possibility of a relearning advantage of previously learnt items. Thus, in the delayed testing session, the target items and previously unencountered items were studied to determine any relearning benefit attributable to latent knowledge of the target items. The extent of retention of each word as measured by the test of retention was compared to the extent of acquisition as measured by the test of attainment. The data were then analysed to establish the effect of the target intralexical factors on the process of decay.

Electronic flash card software was used to facilitate the acquisition of the target language. The programme Anki was chosen for the current study. This software was selected because its parameters can be manipulated to suit the purposes of a study, and it records user statistics. The latter feature allowed access to data on the number of times each target item was viewed by a participant. This permitted the measurement of the relationship between the number of exposures to a target item and its relative retention.

The instrument contained two tests. The first measured written form recall. Participants were presented with a definition (the same as was used during the learning procedure) and were required to type the equivalent form. Scoring was dichotomous; correctly spelled responses were awarded a score of one, while incorrect responses were given a score of zero. The second test measured written form recognition. It employed a four-choice multiple choice format with three distractors, the key, and an additional I don’t know option. The instrument was administered via computer using Microsoft Excel. If a participant provided the correct form on the test of written form recall, the equivalent item for that word was automatically omitted from the test of written form recognition. This design is based on previous instruments that have recognised a hierarchical...
relationship between the various aspects of word knowledge (e.g. Laufer & Goldstein, 2004).

5 Results

Results are discussed in relation to the number of words retained after the period of reduced input, the possibility of a facilitation effect for previously learnt words, and the effect of word class and word length in this process. Interlexical (e.g. participant L1) and situational (e.g. vocabulary size) factors are also considered and preliminary findings presented. In addition, pedagogical implications of these findings are outlined. More generally, the research aims to highlight the need for increased attention to be paid to the decay of learnt FL lexical knowledge with the objective of developing robust learning procedures and mitigation protocols.

References


Does the NGSL Provide Good Coverage for Authentic Listening Materials Such as TV Shows and Movies? A Quick Look at Star Wars and Friends

Charles Browne
Meiji Gakuin University
browne -at- ltr.meijigakuin.ac.jp

Yosho Miyata
Graduate Student, Meiji Gakuin University

Yuta Kawauuchi
Graduate Student, Meiji Gakuin University

1 Introduction

Research by Nation and others (Hu and Nation, 2000, Kurnia, 2003) suggest that learners need to know at least 98% of the words in a written text in order to be able to adequately comprehend them, and that this level is achieved at about 8-9000 word families for written texts and 6-7000 words for listening texts (Nation 2006). Recent research asking if the numbers are the same for comprehension of listening materials seem to indicate significantly lower thresholds. A study by van Zeeland & Schmitt (2012) and a follow-up replication study by Schmitt, Cobb, Horst and Schmitt (2015) suggest that 95% coverage may be sufficient for listening comprehension, and that this might be able to be achieved with as little as 2-3000 word families.

This paper presents research by two of the lead author’s graduate students at Meiji Gakuin University (the site of this conference), which tried to test the theoretical assertions of the two studies mentioned above with authentic listening texts, specifically with an example of general English listening materials as exemplified by “Friends”, a popular American television show, as well as an example of the specialized genre of science-fiction, through an analysis of the Hollywood blockbuster, Star Wars Episode I. The working hypothesis of these studies was that it might be possible to achieve 95% with less than 3000 words, if a list of high frequency words such as the New General Service List, or NGSL (Browne, 2013, Browne, Culligan and Phillips, 2013) was used in conjunction with a short list of “key words” as identified by a tool such as Tom Cobb’s Keywords Extractor software (http://www.lextutor.ca/key/).

2 Method

In both cases the research followed the same pattern of analysis, using online corpus tools to help calculate coverage provided by the NGSL and a small list of additional words as follows:
1. Download unofficial transcripts of Friends and Star Wars
2. Clean up transcripts (remove stage directions and spelling errors)
3. Analyze coverage provided by NGSL using VocabProfile Tool, removing proper nouns from results
4. Use KeyWord Tool to increase coverage by identifying frequent words particular to the transcript
5. Manual analysis of remaining “off list words” to identify other important words
6. Upload List of important non-NGSL words to free flashcard site Quizlet.com for study

3 Coverage for the TV show Friends (by Yuta Kawauuchi)

When I read the Schmitt, Cobb, Horst, and Schmitt (2015) article that suggested 95% coverage is sufficient for listening comprehension, and that this could be achieved with as little as 2-3000 word families, I wanted to test this hypothesis against something real and authentic. The TV show Friends was chosen for analysis since it is one of the most popular and famous comedy TV shows in the United States.

As can be seen from Table 1 below, the 2800 New General Service List words (Browne, 2013) provided fairly high coverage for Season 1 of Friends (92.59%) but did not achieve the 95% threshold suggested by the research. As a next step towards getting to 95%, the Keyword Extractor Software was used to identify 40 words that occurred more frequently in Friends than in general English These 40 words added about 5% coverage for a total of close to 98%. This means that with well under 3000 words (2800 NGSL + 40 keywords = 2840 words), 95% threshold was easily met, meaning that adequate comprehension of Friends is likely possible with less than 3000 words.

<table>
<thead>
<tr>
<th></th>
<th>NGSL minus Proper Nouns</th>
<th>NGSL + 40 Keywords</th>
<th>Addition Coverage from Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>Season 1 of Friends</td>
<td>92.59%</td>
<td>97.87%</td>
<td>+5.28%</td>
</tr>
</tbody>
</table>

Table 1: Results for Analysis of Season 1 of Friends
In order to make the results of this research useful to others who are studying English and wish to understand *Friends*, I uploaded the list of 40 keywords to the free Quizlet.com flashcard site where the words can easily be studied together with the NGSL flashcards.

4 Coverage for the science fiction blockbuster *Star Wars: Episode 1* (by Yosho Miyata)

My colleague’s results for *Friends* was quite impressive but perhaps not surprising since *Friends* centers on the daily life of a group of friends living in New York and is an example of the kind of general English the NGSL is designed for. I wanted to see if the NGSL could also provide similarly high coverage for more unusual genres such as science-fiction. As can be seen in Table 2 below, NGSL coverage was a bit lower than for *Friends*, still over 90% and the addition of 36 keywords brought the total to 93.73%.

In searching for other ways to increase coverage of *Star Wars*, I noticed that in the results section of VocabProfile, there was a list of words labelled as “Offlist Words”. These were words not included in the list of words used in previous analysis (NGSL, proper nouns and Keywords) but many of them seemed to me to be quite important in understanding *Star Wars*. For example, words such as *astro*, *autopilot*, *battleships*, *podracing*. In the end I chose 37 Offlist words that seemed important to understanding *Star Wars* which increased coverage by 0.7% to 94.43%. This means that with well under 3000 words (2800 NGSL + 36 keywords + 37 offlist words = 2873 words), it is possible to get very close to the 95% threshold even for the genre of science fiction with less than 3000 words.

<table>
<thead>
<tr>
<th></th>
<th>NGSL minus Proper Nouns</th>
<th>NGSL + 36 Keywords</th>
<th>Addition Coverage from Keywords</th>
<th>NGSL + 36 Keywords + 37 Offlist words</th>
<th>Addition Coverage from Offlist Words</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Star Wars 1</em></td>
<td>90.60%</td>
<td>93.73%</td>
<td>+3.13%</td>
<td>94.43%</td>
<td>+0.70%</td>
</tr>
</tbody>
</table>

Table 2: Results for Analysis of *Star Wars: Episode 1*.

In order to make the results of this research useful to others who wish to understand *Star Wars: Episode 1*, I uploaded the list of 36 keywords and 37 Offlist words to the free Quizlet.com flashcard site where the words can easily be studied together with the NGSL flashcards that are already there.

References


Collocation Learning through AWARE Approach

Yu-chieh Hsu
National Chung Cheng University, Taiwan
clych -at- ccu.edu.tw

1 Introduction

According to Nation and Webb (2011), collocation included in multi-word units is crucial to language use and language learning. Research indicates that native and more advanced language learners demonstrate high level of collocational competence (Lewis, 2000). However, for most English learners at intermediate level, their knowledge of collocation is limited, and the lack of collocational competence might bring about problems in their English learning. Previous research focused on the theoretical discussion of collocational competence, examinations of the need for learning collocation, and learners’ use of collocation. Little research has been conducted on how to learn and teach collocation for learners at the intermediate level.

Followed by the research done by Yang and Hendricks (2004) using “AWARE” approach, this research looks at how learners at university level in EFL context use “AWARE” approach to learn collocations and what are their difficulties during their learning process. This research aims to apply “AWARE” approach to teach collocation, hoping to shed some implications on teaching and learning collocation. The framework underlying “AWARE” approach is the role of language awareness in language learning. The participants of this study are 27 learners with intermediate level of proficiency, taking the English for general purpose (EGP) course for a semester (18 weeks) in a university. The data of the study are student interview and student reflection after learning collocation by using “AWARE” approach.

The interview data show that learners do cultivate their collocational awareness after the course, and learners have started to notice words in chunks. The general difficulties among learners are that they find simple verbs appear to be difficult for them in collocation, for example take, get, have, and also some of them have difficulty to tell collocations from phrasal verbs. The practical implication of the research is that collocation teaching is suggested to be included in the language course, and the “AWARE” approach is recommended to be adapted in teaching collocation.

2 Method

The participants of this study are 27 freshmen in General English course. The General English course is leveled, and all of them are intermediate learners. The term “AWARE” approach by Yang and Hendricks (2004) means the steps in process-oriented learning approach is adapted for the participants of this research, which are Awareness-raising, Why learning collocation, Acquiring by explicitly learning through applying different learning strategies, Reflecting on what have been learned, Exhibiting what has been learned. A mini-lesson on collocation is given in the beginning of the course, instructing what is collocation, why learning collocations is important, and also types of collocation. After the mini-lesson, a collocation list organized by units of the course book is provided on the e-learning platform for self-study. Students are constantly asked to do peer-discussion and respond orally in class about the collocation in the unit. After a few weeks, students are asked to write down the collocation in the units every week as in-class exercises (unmarked). The exercises are the collocation from the collocation list, and students need to fill in the missing collocation while given the part of speech and the Chinese translation. After doing this for five times, students are asked to complete the exercise of fill in the blanks of a short story, and the story combined all the previous learned collocations.

According to the framework of the “AWARE” approach, the research aims to answer the questions as follows: 1) What are the participants’ perceptions on learning with the “AWARE” approach and how do they apply those steps in their collocation learning process? 2) What difficulties or problems do they have during the collocation learning process? The interview questions focus on asking whether learners know what collocation is, why do they need to learn collocation; some ways learners prefer to learn collocation, some learning strategies they apply when they study, and the difficulties in learning collocation. Data from both the interview transcripts and students learning reflections were coded to identify the major ideas in students’ perspectives and their identified difficulties.

3 Findings

The findings of this research are briefly summarized into two parts. First, students generally think that
“AWARE” approach helps them in learning collocation. Learners respond positively in terms of the learning as well as indicating collocation helps them a lot in speaking and writing. Besides, most of the students indicate that collocation helps them to produce “economical” languages with more accurate usages and also to sound more “native-like.” For ways to learn collocation, some students think the mini-lesson and the collocation list help them a lot in learning collocation, and some students prefer the in-class discussion and the in-class timed exercises. Most students prefer to learn collocation in context (e.g. in stories and in the reading articles in the course book) as they think it will also assist them in identifying the usages. For most of the learner, they identify themselves as passive learners, and they appreciate the peer-learning and quizzes in class. Second, the difficulties for learners in learning collocations are in recalling which verbs to choose when using it, especially the simple verbs such as take, get, have due to those verbs have similar translations in Mandarin Chinese, and those verbs frequently collocates with other words. Other difficulties learners suggest are that they confuse collocation with phrasal verbs and unable to identify what need to be noticed.

4 Conclusion
The research indicates that learners respond positively in terms of learning collocation via “AWARE” approach. After learning with “AWARE” approach, learners think collocation helps them to use less words and sound more native-like in both written and spoken productions. Learners suggest that simple verbs appear to be their learning problems in learning collocation, and sometimes they find it hard to identify what collocation is; however, learners begin to aware some new and special usages other than simply spare attention on new vocabularies.

The limitation of this research is that due to the limited time before summer vacation, the interviewing data were unable to cover all 27 students taken the course, leaving some student opinions without in-depth investigation. For this research, learners showed that focus on learning collocation improve their language proficiency. Therefore, the implications of the research are language teachers are suggested to include collocation teaching an important part in the language course for learners at intermediate level and also adapt the “AWARE” approach in their instruction.

References


Multiple-choice Vocabulary Test Taking for Monolinguals, Early Bilinguals, and Late Bilinguals

Dawn Lucovich
Tokyo Woman’s Christian University
Lucovich -at- Lab.twcu.ac.jp

1 Introduction

This paper comprises case studies of two monolingual (L1 American English), two early bilinguals (L1 Greek-L2 American English), and two late bilinguals (L1 Japanese-L2 American English), and their performance on the 100-item Vocabulary Size Test (VST) (Nation & Beglar, 2007). This study sought to: 1. qualitatively complement the existing body of quantitative data on vocabulary assessment; 2. employ a “labour-intensive method of testing” to ensure meaningful, non-arbitrary results (Nation, 2012, p. 10); 3. identify how monolingual, early bilingual, and late bilingual test takers approached the test, gauged their own word knowledge, and determined their answers; and, finally, 4. compare the respective participants’ test taking approaches.

2 Methodology

For the purposes of this study, several dimensions--behavioral, socio-cultural, and linguistic--were considered when classifying test takers as monolingual, early, or late bilinguals. This is in accordance with theories of bilingualism (see Hamers & Blanc, 2000). All participants reported currently using English on a daily (or near-daily) basis for both personal and professional purposes.

Participants were individually briefed and then took the monolingual (English) 100-item Vocabulary Size Test (VST) (Nation & Beglar, 2007). No time limit was imposed. Immediately following was a structured retrospective word-knowledge and test taking interview (adapted from Nagy, Herman, & Anderson, 1985) that covered their language learning histories, language use, the VST overall and each item individually (k = 100). The interviews were transcribed, and then coded by the interviewer using provisional coding (Miles & Huberman, 1994).

3 Results

The results of the study found that test takers, regardless of L1, language proficiency (or non-proficiency), or age of acquisition, utilized the same test taking approaches and determination of test answers and in similar proportions (with respect to number of items responded to and total vocabulary size):

- Knowledge-informed (K) answer: Test takers had knowledge of the target word.
- Partial knowledge-informed (PK) answer: Test takers had knowledge of the target word, but this knowledge did not satisfy condition (K).
- Test-informed (T) answer: Test takers had knowledge of a non-target word, a distractor, or item stem, or applied test strategy.
- Informed (IG) guess: Test takers had knowledge and/or applied test strategy, but this knowledge and/or test strategy did not satisfy condition (K) (PK) or (T).
- Non-informed (NG) guess: Test takers lacked knowledge and/or were unable to apply knowledge to the item; their knowledge and/or test strategy did not satisfy condition (K) (PK) (T) or (IG).

4 Discussion

The results of this study imply that age of acquisition or mono/multi-lingual status minimally affect test taking in fluent users of English. They also demonstrate that test takers are not strictly responding with target-word knowledge on tests of written receptive vocabulary size. This is problematic since vocabulary size instruments in particular may be sensitive to score overestimation.

5 Limitations

While this study has contributed qualitative knowledge to the process of vocabulary test taking, there are several limitations to the methodology and interpretation of results. Limitations of this study include the non-transferability of results to other test takers, contexts, languages, and proficiency levels. As this study
specifically targeted fluent users of English, results for English language learners (whether interviewed in their L1 or the target language) may differ drastically for a number of reasons other than fluency, even with test takers of similar age, education, experience, training or metacognitive knowledge of test taking.

6  Conclusion

In the future, since monolingual L1 English and both early and late bilingual users of English show similar test taking patterns and in similar proportions, it would be worth qualitatively investigating test taking on other instruments and with other populations. First, a comparison of the performance on other vocabulary instruments between monolingual L1 English, early and late bilingual users of English, including recall tests, should be conducted. Second, a comparison of the results from this study’s participants with results from returnee early/late bilingual users of English is necessary (Lucovich, in progress). Third, as in this study, test takers with an L1 other than Japanese, and bilingual and multilingual learners and users of languages other than English should be investigated. Finally, and perhaps most importantly, it should be determined how these test taking results might have an effect on approaches to second language instruction and assessment in particular.

References


Principles Guiding the Teaching and Learning of Linking Adverbials in English

Zihan Yin
Southern Cross University
yinzihan1027 -at- gmail.com

Paul Nation
Victoria University of Wellington
paul.nation -at- vuw.ac.nz

1 Introduction

Linking adverbials are important for creating textual cohesion in both written and spoken discourse. While there are reference grammar books describing the usage patterns of linking adverbials (e.g., Biber et al., 1999) and studies investigating learners' difficulty in using these cohesive devices (e.g., Granger and Tyson 1996), there is little discussion on how to effectively teach and learn them.

The present paper draws on a comprehensive two-stage corpus study on all the three aspects of the usage patterns of linking adverbials in five different registers: written academic prose, academic lectures, written news, broadcast news, and conversation. Stage I was a manual analysis, since manual analysis of authentic language data can provide richer and more in-depth information than the quantitative corpus analysis which has been the focus of research to date. Linking adverbials in all the five registers were manually analysed. Possible linking adverbials were identified. Form, meaning, and position of identified linking adverbials were coded in QSR International’s NVIVO 9 software, and differences between linking adverbials and other grammatical categories (e.g., conjunctions) were examined. In total, 67 texts of a total of over 100,000 words were manually analysed. All texts were chosen from the Wellington Corpora of Written and Spoken New Zealand English (WWC and WSC), compiled by the School of Linguistics and Applied Language Studies, Victoria University of Wellington, which provided coverage of the target five registers. Texts were chosen randomly from the relevant sections of the corpora. Stage II was a corpus-based quantitative analysis. The most frequent linking adverbials in written registers identified in the first stage were automatically searched for in larger corpora, i.e., the whole WWC, BNC, and COCA. In the first stage, for a more precise analysis of the linking adverbials and their usage patterns in spoken data, the intonation unit was adopted as the analysis unit and sound files were needed in deciding intonation units and interpreting uses. Thus, in the second stage, automatic search for patterns in spoken data was not possible and the focus was on written data.

Drawing on the research findings from this two-stage corpus study, this paper suggests seven principles to guide the teaching and learning of linking adverbials in the English language classroom:

1. When teaching linking adverbials in the language classroom, all three aspects of usage patterns, i.e., form, meaning, and position, need to be covered to give the learners a rounded picture of the usage of linking adverbials.
2. Frequency patterns of linking adverbials need to be taken into consideration in designing teaching material, with attention being given to high frequency meaning markers and low frequency positions.
3. Register variation also needs to be considered when designing teaching materials and in classroom teaching.
4. Data-driven learning can be introduced to arouse learners’ awareness and increase their motivation.
5. Teachers of a particular English variety background need to be less prescriptive, especially nowadays, when English is used popularly as a lingua franca.
6. Other forms of cohesive devices (e.g., reference, substitution) should also be introduced.
7. The deliberate teaching and learning of cohesive devices should make up around a quarter of the opportunities to learn about cohesive devices. The learners should have plenty of meaning-focused input and output and opportunities for fluency development across the four skills in the appropriate registers and genres.

The principles suggested here provide guidelines for different stages of classroom teaching. First, in lesson planning, language teachers need to incorporate all three aspects of the usage patterns of linking adverbials, as well as their frequencies, register and variety variations and other forms of cohesive devices. Second, in classroom teaching, awareness raising and motivation are catalysts for deliberate learning, for which data-driven learning could provide support. Third, in learning outside the classroom, different modes of learning, i.e., receptive and incidental learning, need to be extended by providing meaning-focused input and output and frequency development opportunities.

References


Explicit Learning, Filling-in and Restructuring of Multi-Word Chunks: Insights into the Development of Phraseological Competence in L2 German

Susanne Rott
Department of Germanic Studies and Linguistics, University of Illinois at Chicago
srott -at- uic.edu

1 Introduction

The development of phraseological competence, broadly defined as the knowledge of how words combine with others and the combinations’ morphosyntactic and lexical restrictions and pragmatic constraints, has been receiving increased attention in second language acquisition (SLA) research and instructional literature. L2 studies assessing learners’ usage of CMWUs suggest that learners sound more proficient when they use collocations (e.g., Boers, Eyckmans, Kappel, Stengers, & Demecheleer, 2006). Likewise, these studies also suggest that a majority of errors at the advanced learning level may not be grammatical but phraseological in nature (Bahns & Eldaw, 1993).

The restrictive nature of conventionalized and formulaic multi-word units presents a particular challenge to second language (L2) learners (e.g., Nesselhauf, 2003). While some restrictions about which words can be combined are semantically motivated, many others appear arbitrary and unpredictable for L2 learners (e.g., Lewis, 1997) and thus cannot be accounted for by grammar rules or the learning of L1 to L2 translation word lists.

2 Review of the literature

Schmitt and Carter (2004) have pointed out that some short formulas may be learned holistically but that others may be acquired incrementally. Naturally, noticing a multi-word chunk and identifying its boundaries and constituents is the first step to establishing a form-meaning connection.

Even though words that combine as multi-word chunks are frequently adjacent in a sentence in English, this is not always the case for morphosyntactically complex languages. For example, in German constituents of multi-word chunks can be separated by multiple words thereby making it difficult for learners to notice and process them together. Studies (e.g., Jones & Haywood, 2004) revealed that awareness for multi-word chunks can be taught successfully but does not immediately translate into use. In fact, multi-word chunks, like individual word learning, might develop from partial to more complete knowledge during multiple, potentially explicit, learning encounters.

3 Research questions

Quantitative analysis:
- Do learners of German establish complete form-meaning connections of conventionalized multi-word units during an explicit learning task? If so, do they retain these form-meaning connections over time (five weeks)?
- If partial form-meaning connections are established, are they retained over time (five weeks)? How much partial knowledge is gained?
- Does language proficiency (second or fourth semester) effect the establishing of form-meaning connections?

Qualitative analysis:
- How much of a partial form-meaning connections can be established and how much can be filled-in and restructured through explicit learning?

4 Method

Participants: 71 beginning and intermediate learners of German
Materials: 28 German multi-word chunks and their English translation
Procedure: four sets of seven chunks were learned explicitly
Assessment: Pre-test, post-test, delayed posttest (five weeks)
Scoring: complete and partial production modified from Barcroft (2008)
5 Results

<table>
<thead>
<tr>
<th>Time of Assessment</th>
<th>n</th>
<th>Partial</th>
<th>Complete</th>
<th>Partial</th>
<th>Complete</th>
<th>Complete</th>
<th>Partial</th>
<th>Complete</th>
<th>Partial</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beginning</td>
<td>35</td>
<td>.31</td>
<td>12.26</td>
<td>2.43</td>
<td>18.14</td>
<td>1.29</td>
<td>16.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.47)</td>
<td>(2.58)</td>
<td>(1.17)</td>
<td>(2.69)</td>
<td>(.86)</td>
<td>(2.67)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intermediate</td>
<td>36</td>
<td>.94</td>
<td>15.03</td>
<td>5.86</td>
<td>18.66</td>
<td>3.22</td>
<td>19.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.63)</td>
<td>(1.95)</td>
<td>(1.84)</td>
<td>(3.14)</td>
<td>(1.12)</td>
<td>(2.49)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Descriptive Statistics of Complete and Partial Knowledge of CMWUs

Note: Maximum Score = 28

Figure 1. Beginning Learner’s Partial Knowledge of CMWUs at Three Times of Measure.

Figure 2. Intermediate Learner’s Partial Knowledge of CMWUs at Three Times of Measure.

References


Towards Appropriate L1 & L2 Scales of English Vocabulary in English-Dominant Contexts: Vocabulary Test Item Difficulty Indicates Lexical Sophistication Scale and Definition of "Word"

Geoffrey G. Pinchbeck
Werklund School of Education, University of Calgary
ggpinchb -at- ucalgary.ca

1 Introduction & background

This study investigates two issues that are important in the design of word lists used in vocabulary research and pedagogy: 1) the source of reference corpus/corpora and 2) word definition. This is part of a larger project with the goal of developing scales of lexical sophistication and lexical development for English-L1 and bilingual learners in English-dominant K-12 contexts in North America.

First, the choice of reference corpus is a key in word list development. Word lists used in applied linguistics and in L2 language pedagogy have been based primarily on word frequency and/or distribution data (see Gries, 2009) of large general corpora that have been compiled purposefully to 'balance' various registers, genres and modalities. However, it is not clear, a priori, which corpora or in what proportions data from different sub-corpora should contribute to word ranks in any given word list (see Biber, 1993).

Second, the question of how word should best be defined in L2 settings has been investigated empirically and debated on theoretical grounds (e.g. Gardner, 2007; Schmitt & Zimmerman, 2002), and word families (i.e. levels 1-6 in Bauer & Nation, 1993) have been widely used in a variety of word-list-based applications. Derivational morphology research on L1-English children reveals that knowledge of word affixes develops throughout the K-12 years (Angelin, 1993), which suggests that definition of word for K-12 word lists might also need to be adapted appropriately to match learner morphological competency.

2 Methods

In a modification of the methodology of Brysbaert and New (2009), the present study employed vocabulary-test word-item difficulty ranks as an external reference criterion upon which candidate word lists could be evaluated. Corpora-derived ranked word lists based on either flemma (roughly inclusive up to Bauer & Nation's levels 1-2) or word family (levels 1-6) were first generated from frequency/dispersion ranks from: 1) the major sub-corpora of the COCA (Davies, 2010), and BNC (Leech, et al., 2002), 2) children's speech (MacWhinney, 2000; Moe, et al., 1982; Murphy, 1957), 3) U.S. school-approved texts (Zeno, et al., 1995), 4) US & UK TV and movie subtitle corpora (Brysbaert & New, 2009; Van Heuven, et al., 2014), 5) the CELEX (Baayen, et al., 1995), and 6) the corpora upon which both new General Service Lists are based. U.K. and U.S. spelling forms were merged into the same word groupings in both the flemma and family based lists.

Corpora frequency/dispersion rank data (20,000+ items) were then compared with vocabulary test word item difficulty rank data (400+ items) for three commercially-available norm-referenced vocabulary tests widely used by K-12 psychologists and English language learner (ELL) specialists in North America (Ballard & Tighe, 2009; Dunn & Dunn, 2015; MacGinitie & MacGinitie, 1992).

3 Results

Using Spearman-rank-correlations and a Hotelling-Williams significance test of correlation differences, the optimal definition of word was first found to depend on the age of the test takers: Word lists using flemma-based grouping word lists showed significantly higher correlations with word ranks derived from younger (≤ 12 y.o.) test-taker data, and also for that of ELL norm-populations, whereas large, word family groupings (i.e. levels 1-6) resulted in consistently higher correlations with item difficulty rankings of older (>17 y.o.) test-takers only. These results are consistent with known derivational-morphology developmental differences.

Genre and modality also showed significant effects: Lists derived from corpora of informal spontaneous speech, TV/movies, and narrative-written texts consistently showed higher correlations with test item difficulty ranks than those derived from formal speech or any non-fiction written-text genres. This pattern was observed for all three independent sets of test data, representing 1) the U.S. general population, 2) U.S. ELL student population, and 3) Canadian K-12 student population, and was also consistent across age groups.
4 Conclusions

These results suggest that the prevailing trend to compile mega-corpora from easily-available electronic written-texts may not be ideal for the creation of either general-service-lists or developmental scales of academic English lexis for North American K-12 learners. Furthermore, these results also support examining the derivational morphology knowledge of the learners prior to using word lists in English language research and pedagogy: instruments that might be designed for younger L1 learners or emerging bilinguals should probably make fewer assumptions regarding learner knowledge of derivational morphology.

5 Acknowledgements

Thanks to Dr. Sari Luoma at Ballard & Tighe Publishers for the IPT vocabulary data and to Rick Norman at Nelson Education Ltd. (Canada) for the Gates-MacGinitie Reading Test (2nd Canadian Ed.) vocabulary data.

References


The Effect of Vocabulary Strategies on Vocabulary Use in Intermediate-level EFL College Students

Cristián Sanhueza
University of Santo Tomás
csanhueza -at- santotomas.cl

1 Introduction

One central concern in the vocabulary research field is the relationship between receptive and productive lexical knowledge and its transfer onto vocabulary use. Breath of lexical knowledge across varied frequency bands has proved to be essential to lexical competence, either in writing or oral production. Poor national results of secondary school students in recent years have placed EFL teaching quality and teacher proficiency in Chile under scrutiny. This 2-year study is set within the field of vocabulary acquisition and learning of English as a foreign language at the university level with teacher-training students. Its main aim is to determine the effects of systematic vocabulary teaching and learning strategies use over the participants’ lexical competence in their writing production.

2 Method

To this end, a quasi-experimental study with pre and post-test was designed. The treatment process consisted of a weekly implementation of one of 4 vocabulary teaching strategies (paraphrasing, second-hand cloze procedure, dicto-comp and ‘speaking for writing’) and 2 vocabulary learning strategies (word cards and thematic record) as project assignments that foster the development of vocabulary use over two academic semesters. In every English writing class, the teacher developed one vocabulary teaching strategy in an alternating fashion, whereas the two vocabulary learning strategies were presented in the form of students’ projects to be monitored by the teacher every three weeks and submitted by the students at the end of each term.

The data elicitation process was carried out by means of different instruments, the Vocabulary Levels Test receptive (Nation, 1990; Schmitt & Clapham, 2001), the Vocabulary Levels Test controlled-productive (Laufer & Nation, 1999), and the Lexical Frequency Profile (Laufer & Nation, 1995; Laufer, 2005, 2012) for free compositions as the pre and post-tests. The participants were 78 intermediate 3rd-year college students, 40 in the treatment group and 38 in the control group.

3 Results

Results from the first year experiment showed that the treatment process did have a positive effect on the participants of the treatment group across different frequency bands in the three tests in relation to the control group, being the receptive knowledge the one that showed more progress.

<table>
<thead>
<tr>
<th>Variable (N=40)</th>
<th>Before</th>
<th>After</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
</tr>
<tr>
<td>R_2000</td>
<td>943,33</td>
<td>56,76</td>
<td>970,00</td>
</tr>
<tr>
<td>R_3000</td>
<td>793,33</td>
<td>127,46</td>
<td>913,33</td>
</tr>
<tr>
<td>R_5000</td>
<td>736,67</td>
<td>143,54</td>
<td>796,67</td>
</tr>
<tr>
<td>R_UWL</td>
<td>906,67</td>
<td>43,89</td>
<td>940,00</td>
</tr>
<tr>
<td>R_10000</td>
<td>353,33</td>
<td>198,89</td>
<td>503,33</td>
</tr>
</tbody>
</table>

Table 1. VLT Test Receptive Results in the Treatment Group.

Also, interesting developments were observed in the controlled productive knowledge and free vocabulary use in the treatment group. Significant gain is observed in band PC_5000 and PC_10000, consistent with the growth obtained in the VLT receptive test.
growth obtained in the use in the treatment group.

Table 1. VLT Test Receptive Results in the Treatment Group, being the receptive knowledge the one that showed more progress.

<table>
<thead>
<tr>
<th>Variable (N=40)</th>
<th>Before</th>
<th></th>
<th>After</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PC_2000</td>
<td>888,89</td>
<td>107,98</td>
<td>850,00</td>
<td>94,61</td>
</tr>
<tr>
<td>PC_3000</td>
<td>544,44</td>
<td>179,16</td>
<td>550,00</td>
<td>172,63</td>
</tr>
<tr>
<td>PC_5000</td>
<td>266,67</td>
<td>165,22</td>
<td>483,33</td>
<td>185,28</td>
</tr>
<tr>
<td>PC_UWL</td>
<td>577,78</td>
<td>91,47</td>
<td>588,89</td>
<td>192,81</td>
</tr>
<tr>
<td>PC_10000</td>
<td>88,89</td>
<td>102,10</td>
<td>366,67</td>
<td>211,47</td>
</tr>
</tbody>
</table>

Table 2. VLT Controlled-productive Test Results in the Treatment Group.

With regards to the results of the free composition, it is possible to observe significant gains in the total number of words produced by the participants in the treatment group. This translates as gains in the total number of words produced in band 1000, in the word families and the sum of band 1000 and 2000.

4 Conclusion

It is possible to conclude that the treatment group presents significant gains in the three tests applied and in the different frequency bands of the VLT tests receptive and controlled-productive and indicators of the LFP in the free composition. In response to the main aim of the study, the systematic teaching of vocabulary, in contrast with the incidental one, does affect positively in the participants’ lexical competence. Additionally, we can say that the improvement in the treatment group is attributable to the intervention process, since the control group, working under the same conditions (with the exception of the lexical treatment) mostly maintained, in two frequency bands moderately improved and in other two its performance decreased. Nevertheless, the incidence and development of the different lexical knowledges and use differ amongst them. The receptive knowledge is the one that progresses the most in the treatment group, however, this knowledge does not transfer easily onto the controlled-productive one. In fact, the controlled-productive and the free lexical use are the ones that develop the slowest.

References


1 Introduction

Acquiring basic high-frequency words is essential for ESL/EFL learners. Learning new words in context (contextualizing) and learning new words out of context (decontextualizing) are “complementary ways of learning” (Nation, 2013, p. 441). Decontextualizing such as learning with word lists or word cards is especially effective for novice learners to establish form-and-meaning connections.

After ESL/EFL learners have learned hundreds of high-frequency words, they can make use of these words in order to expand their vocabulary. Attaching a known word to a target word has been found to be effective for retrieval and retention of the meaning of the target. Kasahara (2010, 2011) revealed that learning a known-and-unknown word combination was more effective in retention and retrieval of the meaning of the target unknown word than learning a single unknown word in isolation. The known word can help learners to put the target words into their mental lexicons in the encoding phase, and the known word can help them to limit the scope of meaning search in the decoding phase.

Most high school students in Japan have learned about 1,000 words in junior high school (Kasahara, 2006), and they can use these high-frequency words as known cues in order to enlarge their vocabulary. They are supposed to learn another 3000 words at high school. Consequently, it would be helpful for them to have a list of known (1000 word-level) and unknown (2000-4000 word levels) word combinations. The purpose of this study is to create such a list of two-word combinations for intermediate English learners who already know about 1,000 basic English words.

2 Materials

We used The JACET 8000 (JACET, 2003) in order to identify known words and target unknown words for Japanese high school learners of English. The JACET 8000, which is based on British National Corpus (BNC) and English materials published in Japan, has been considered to be a reliable vocabulary list for Japanese learners of English for more than a decade. We decided to use the most frequent 1000 words as the known words, or the nodes in two-word combinations, and to employ the following 3000 words from 2000-word level to 4000 word-level as the unknown target words, or the collocates in the two-word combination list.

The Corpus of Contemporary American English (COCA) was used to select known-and-unknown word combinations because it is the largest freely-available English corpus, including 450 million words. COCA includes various discourse genres such as spoken, fiction, popular magazines, newspaper and academic texts. Therefore, COCA is thought to be a reliable reflection of modern American English, which is the most commonly taught in English lessons in Japan. In addition, COCA has useful collocation searching functions, which make it possible to find frequent collocates to a target word.

3 Selecting known-and-unknown combinations

First, we extracted 860 known words from the most frequent 1000 words on The JACET 8000 by eliminating determiners, pronouns and prepositions. Then, using the collocation searching functions, we selected words that collocate with the known words. The searching functions of COCA can find words that appear to the left or the right of the corresponding known words. We selected collocates based on two criteria that make them useful for vocabulary learning for Japanese high school students. The first criterion was frequency of combinations: we chose the combinations that appeared 10 or more times in COCA. The second one was the strength of combination between a node and a collocate. The index we depended on was mutual information (MI) scores, which are automatically calculated in the searching function of COCA. If a collocate has a high MI score with its node, it is highly probable to see these two words together in various contexts. Schmitt (2010) shows that “a commonly cited threshold for statistical significance for MI is 3” (p. 131). Therefore, combinations whose MI scores were 3 or higher were selected. Finally, among these collocates that satisfied the two criteria above, we singled out the collocates that ranged from 2000-word level to 4000-word level in The JACET 8000.

As a result, we obtained two types of combination lists: one is a list where the known words come to the right side of the unknown words; the other one is a list whose order is the opposite. We named the former the
known-unknown list (Kn-Un list) and the latter the unknown-known list (Un-Kn list).

4 Screening

We continued by discarding combinations that we judged were not suitable to include the final lists. We adopted the types of combinations whose learning effect was validated in Kasahara (2013), such as noun + noun, adjective + noun, verb + noun and verb + adverb. We also included the combinations like adverb + verb and adverb + adjective because they can work as independent chunks. However, we eliminated other types of combinations, which are thought to be parts of larger chunks such as noun + verb, verb + adjective, noun + adjective, noun + adverb, adverb + noun, adjective + adjective and adverb + adverb, etc. We further eliminated combinations including undesirable nodes for learning, such as auxiliary verbs (can, may, will…), quantifiers (much, many…) or interrogatives (what, who, where, when…). This process left us with approximately 20,000 combinations on Kn-Un list and around 23,000 combinations on Un-Kn list.

Then, more detailed screening was conducted: we checked the combinations one by one and got rid of combinations that were true to one of the following conditions.
(a) a combination that is part of a larger chunk (e.g. prevent people [from doing…])
(b) a combination that includes an archaic word (ye) or a colloquial word (grandma) or a vulgar word
(c) a combination that denotes a proper noun (Bottle Mountain)
(d) a combination that denotes the name of a product or trade mark
(e) a combination that denotes a technical term (storm cell)
(f) a combination that denotes the name of a species (passion fruit)

5 Results

The detailed screening provided us with 17,000 items on Kn-Un list and 20,000 items on Un-Kn list.

<table>
<thead>
<tr>
<th>Known</th>
<th>Unknown</th>
<th>COCA</th>
<th>MI</th>
<th>JACET (K)</th>
<th>JACET (U)</th>
<th>PoS (K)</th>
<th>PoS (U)</th>
</tr>
</thead>
<tbody>
<tr>
<td>present</td>
<td>article</td>
<td>155</td>
<td>4.5</td>
<td>399</td>
<td>1002</td>
<td>a</td>
<td>n</td>
</tr>
</tbody>
</table>

Table 1. An Example Item from Kn-Un List

<table>
<thead>
<tr>
<th>Unknown</th>
<th>known</th>
<th>COCA</th>
<th>MI</th>
<th>JACET (U)</th>
<th>JACET (K)</th>
<th>PoS (U)</th>
<th>PoS (K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cultural</td>
<td>activity</td>
<td>68</td>
<td>3.9</td>
<td>1003</td>
<td>619</td>
<td>a</td>
<td>n</td>
</tr>
</tbody>
</table>

Notes: COCA = frequency in COCA; JACET = frequency order in JACET; (K) = Known; (U) = Unknown

Table 2. An Example Item from Un-Kn List

6 Discussion and conclusion

The two combination lists above can be a helpful tool for learners of English who know basic high-frequency words, such as high school students in Japan, to increase their English vocabulary. If they already know 1000 most frequent words in the JACET 8000, they can use these words as known cues that can be attached to the next 3000 words. The connection between the known words and new words help them to transfer the meanings of new words in their mental lexicons, and to retrieve the meanings of the target words when they see it again.

However, these lists include the combinations whose effectiveness has not been demonstrated in prior research (adverb + verb and adverb + adjective combinations). Moreover, how to use the list effectively has not been researched. These problems should be clarified in further studies.

References


Starting from Scratch: Issues in Developing a New Diagnostic Vocabulary Test

Benjamin Kremmel
University of Nottingham, UK
University of Innsbruck, Austria
benjamin.kremmel -at- nottingham.ac.uk

1 Introduction

Vocabulary testing is deeply rooted in conventionalized traditions. Several vocabulary tests of written receptive word knowledge have been developed and suggested for different purposes, such as the Vocabulary Levels Test (VLT) (Nation, 1990; Schmitt, Schmitt & Clapham, 2001), the Vocabulary Size Test (Nation & Beglar, 2007) or the CATSS (Laufer & Goldstein, 2004). Most of these have been commonly labelled measures of vocabulary breadth, suggesting that this is a well-defined construct, useful for practitioners in their daily work with language learners. However, the interpretation of scores generated by tests of vocabulary size is not as straightforward as it seems. This is largely due to taking the underlying assumptions of existing tests for granted and merely reproducing them in new tests, with most tests still conceptualized as paper-pencil tests, using word families as the counting unit, recycling old test formats without criticality, and using outdated frequency lists for sampling. To outline the process of developing a new computer-adaptive measure of vocabulary size, which has attempted to start from a clean slate and base its development decisions on empirically-founded rationales, this extended abstract focuses on three key concerns in these stages.

2 Test formats

The item format a test uses plays a crucial role as it directly affects the measurement. However, there has been little research into how well different test item formats represent a learner’s vocabulary knowledge. For example, does the item format show that a learner can use the target word when reading?

A study that investigated four commonly-used item formats compared the behaviour of two recognition formats (VLT and 4-option multiple choice) and two recall formats (with and without a context sentence) by administering 36 vocabulary items in the different formats in a Latin square design to native and non-native speakers of English. The word knowledge of the participants was afterwards verified in face-to-face interviews and written meaning recall measures, probing whether their test scores derived from guessing, partial knowledge, or mastery of different word knowledge aspects of the target items. The results show that the various test formats do behave differently. The recognition formats consistently gave credit to learners for words they did not know, while recall formats underestimated learner knowledge. On this basis, it may be argued that test scores need to be interpreted in conjunction with an understanding of how various item formats perform in order for those scores to be interpreted meaningfully. Multiple choice formats are tentatively recommended as the most useful item type for this purpose because of their systematicity in measurement error.

3 Counting units

The next assumption addressed is the counting unit. Most vocabulary size tests have used word families, but there are reasons to think that this may not be the best unit. Even if learners know one or more members of a word family, they do not necessarily know all of its members. A study was therefore conducted that compared 99 Austrian EFL learners’ knowledge of root forms with their knowledge of the entire word family. Findings show that the word family is potentially limiting for the interpretability of scores as learners only managed to connect the base word for which they knew the meaning to its derivative forms in 73% of the cases. Given the inappropriateness of word families as a counting unit, I suggest the lemma as alternative counting unit. This is supported by a study that related the scores of a lemma-based vocabulary test to reading proficiency. It found that a change in counting unit (from word family to lemma) does not necessarily imply the need for a complete overhaul of the size of the total sampling population of vocabulary items.
4 Frequency banding

The third assumption is that the vocabulary continuum should be divided into bands or levels of exactly 1,000. Current research, however, indicates that a one-size-fits-all approach to frequency division may not be the best solution. Based on coverage figures of frequency lists from four different corpora, it could be suggested that more narrowly defined bands (500 units per band) might be more useful for higher frequencies, while wider bands (2,000) might be perfectly workable for lower frequencies, particularly for tests aimed at beginner to intermediate language learners, who are most often the target audience of vocabulary tests.

5 Implications and outlook

Incorporating the findings from the studies outlined above, it appears that a new diagnostic test of vocabulary knowledge that uses up-to-date lemmatized frequency lists for item sampling, explores a new approach to frequency banding and employs an item format that has been selected on the grounds of an empirical comparison of different formats appears a potentially useful addition to the current repertoire of vocabulary assessment tools. In addition, it seems worth exploring how such a test could be implemented in an online test environment to probe the success of the computer-adaptiveness of such a measure.

References


Going with the Flow: The Specialised Vocabulary of Plumbing

Averil Coxhead
Victoria University of Wellington
Averil.coxhead -at- vuw.ac.nz

1 Introduction

Vocabulary studies in English for Specific Purposes have tended to focus on university or secondary level educational contexts, including well-known and widely used lists of general, specialised and technical vocabulary (see, for example, Coxhead, 2000; Simpson-Vlach & Ellis, 2010; Liu, 2012; Gardner & Davies, 2014; Greene & Coxhead, 2015). In contrast, areas such as plumbing and carpentry have received little attention, yet these areas are key for employment for second or foreign language speakers in many countries around the world. It is fairly well established that specialised or technical vocabulary includes a range of vocabulary, from general everyday vocabulary used with a specialised meaning, such as cost in accounting (see Sutarsyah, Nation and Kennedy, 1994) and flow in plumbing, through to much more technical lexis, such as anti-scour, hot-dip, and P-trap. Plumbers know well that flow has a technical meaning in the trade whereas people outside the field might not know that flow not only has a technical meaning, it is particularly important. It refers to the flow of water or other liquids into a drinkable water supply from somewhere it should not be coming from: it is good to know that a plumber knows that. This corpus-based investigation into the nature of the vocabulary of plumbing considers the role of frequency in specialised vocabulary, based on the division of high, mid & low frequency words (see Schmitt & Schmitt, 2012; Nation, 2013) and the vocabulary load of these specialised texts (Nation, 2006). The study is part of a larger research project on Language in the Trades. The research is important if we are to understand the demands of technical texts in the trades and can support learners and teachers with ways to identify and build trades-based vocabulary into their programmes of instruction.

2 Methodology

A corpus of approximately 565,881 running words of written Plumbing texts which are used in a Polytechnic in New Zealand was developed for this study. This corpus is part of a larger corpus of Language in the Trades, which includes three other trades: Fabrication, Carpentry, and Automotive Engineering. These Plumbing corpus included workbooks, unit standards, and instruction manuals from two levels of study in the institution. The Range programme (Heatley, Nation & Coxhead, 2002) was used to identify the vocabulary in the corpus with Nation’s (2006; 2013) BNC/COCA 25,000 word lists, developed using the British National Corpus and later on the Contemporary Corpus of American English (CoCA) (Davies, 2008). Items in the corpus which occurred outside the BNC/COCA 25,000 and supplementary lists (proper nouns, marginal words, abbreviations, and so on) were then either added to the existing BNC lists or to a new BNC/COCA-style base list of plumbing. Accounting for all the words in the corpus in this way allows for a full analysis of the lexis in the corpus.

3 Results and Discussion

The corpus analysis suggested that there are a number of lexical items relating to plumbing which are outside Nation's BNC/COCA lists as well as within the lists. The plumbing-augmented high frequency BNC lists (1,000-3,000) covered 86.6% of the plumbing corpus (see Table 1). Unsurprisingly, the first 1,000 word families of the BNC had the highest coverage at 63.56%, followed by the second 1,000 at 14.52%, dropping to 8.52% for the third 1,000 word family list. Like other texts, then, these plumbing-specific texts contain a large amount of high frequency lexis, including items which are closely related to plumbing, such as fixings, pipe, pressure, and drain. The Mid-Frequency lists (4,000-8,000) covered 7.29% of the corpus and included words such as outlet, valve, sewer, gully and effluent. Proper nouns (for example, Buchan, Eco, Kelvin, Legionella, Newton, Pascal), marginal words, compounds (e.g. backflow and stormwater), and abbreviations lists covered 3.85% of the corpus. Low frequency vocabulary, (9,000-25,000 word families) covered 1.28% and included items such as ingress, galvanic, reticulated, vitrified, outfall, and ballcock. The combination of high frequency and mid-frequency vocabulary plus proper nouns, marginal words, compounds and plumbing-based vocabulary covered almost 98% of the Plumbing corpus. These figures are similar to Nation’s (2006)
findings of 8,000-9,000 plus proper nouns for novels, newspapers and written academic texts. The vocabulary items from the word lists in the examples and their relationship to plumbing were checked through the use of the words in the Plumbing corpus, expert judgments by Plumbing tutors, online dictionaries and expert websites. Further analysis using the new plumbing-based BNC/COCA lists was then carried out, leading to the development of a specialised word list of plumbing (Coxhead, under review) and further insights into the frequency levels of specialised vocabulary.

For learners and teachers, this kind of analysis is helpful for understanding more about the nature and frequency of lexical items in specialised texts. More research on plumbing vocabulary in spoken texts is underway and is also important, given the apprenticeship model of learning for the trades, where learners and teachers rely heavily on talk for developing understanding of the trade.

4 Acknowledgements

I acknowledge the financial support for this project (NPF14-010) by Ako Aotearoa, and colleagues in the Languages in the Trades (LATTE) project.

References

The Effects of Different Learning Conditions on Implicit and Explicit Knowledge of English Collocations

Mark Toomer  
Victoria University of Wellington  
mark.toomer-at-vuw.ac.nz

Irina Elgort  
Victoria University of Wellington  
irina.elgort-at-vuw.ac.nz

1 Introduction
Second-language learners commonly misuse multi-word units (MWUs), which is due partly to their lack of exposure to MWUs. MWU reading studies among higher-proficiency second-language speakers have found frequency effects: participants read high-frequency MWUs more quickly than low-frequency MWUs and non-MWUs. A second cause of misuse of MWUs is lack of salience (N. Ellis, 2002), with learners often failing to recognise MWUs as lexical units, treating them as individual words instead (Bishop, 2004; Boers & Lindstromberg, 2009).

Acquisition of MWUs is frequently conceptualised as occurring through deliberate learning or incidental learning. A number of studies have shown that collocations can be learned incidentally and incrementally in context, especially through exposure to repeated occurrences of the target phrases (e.g., Sonbul & Schmitt, 2013). A second incidental learning method is a brief externally-induced focus on target words in the context of meaning-focused activities (Laufer, 2005). Techniques such as typographic enhancement and glossing can make target vocabulary more salient, although very few studies have investigated these techniques with MWUs. Furthermore, few studies have examined how learning conditions affect the development of procedural knowledge. In two such experiments, Sonbul and Schmitt (2013) found that both native speakers and advanced non-native speakers gained significant explicit knowledge of two-word, low-frequency technical collocations (measured in gap-fill tasks), but not their implicit knowledge. In their study, the target collocations were presented either in a deliberate condition or one of two incidental learning conditions, in one learning session.

2 Overview of current study
The present research extends the findings of Sonbul and Schmitt’s (2013) experiments by exposing participants to collocations in incidental contextual learning conditions over three sessions, with the expectation that the increased exposure would lead to more robust and fluent processing of the target collocations, in addition to improving their controlled recognition and retrieval. Sixty-two advanced adult ESL speakers of various L1s were exposed to nine occurrences of fifteen English medical collocations (e.g., stone heart, gap periods) in 500-word texts, in three learning sessions over two consecutive days. Three learning treatments were implemented: reading-only, bolding and bolding-plus-glossing. The learning was assessed in two tests of declarative knowledge of the target collocations and one test of procedural knowledge. Declarative knowledge was measured in a cued recall test, in which participants saw a glossed definition of each target collocation and were asked to supply the first words of the collocations in sentence gap-fills. Declarative knowledge was also measured in a recognition test using the same sentences as the cued recall test but with multiple-choice answers provided. Procedural knowledge (i.e., the ability to fluently access collocational knowledge under time pressure) was tested in a primed LDT, in which the first word of each collocation was the prime, presented for 150 milliseconds, and the second word was the target. Two-word control phrases were constructed by replacing the first word in each target collocation with a synonym. For example, rock heart was the control phrase for the target collocation stone heart. Procedural knowledge of the target collocations was operationalised as a priming effect, i.e., faster reaction times to the same word (e.g., heart) presented in the target collocation (stone heart) compared to the control phrase (rock heart). The immediate post-tests were run on the second day of the experiment; the delayed post-tests, attended by 45 participants, were run two weeks later.

3 Results
The preliminary results from the immediate and delayed declarative tests showed that multiple encounters with the target collocations in supportive contexts produced a significant amount of learning in the three treatment groups (combined). Furthermore, participants in the bolding-plus-glossing condition did not display significantly greater declarative knowledge in any post-tests. However, participants in the bolding
condition showed significantly greater declarative knowledge in the immediate cued recall and immediate recognition post-tests than those in the reading-only condition, and greater declarative knowledge in the delayed recognition test. These results largely support previous findings that repeated exposure to typographically enhanced lexical collocations promotes declarative collocational knowledge.

Interestingly, the preliminary results of the LDT tests showed that the reading-only condition was the only condition in the immediate and delayed post-tests in which the reaction times for the target collocations were faster than the times for the control phrases. The implication of this finding is that repeated exposure to lexical collocations in informative reading contexts (with no typographic enhancement) may lead to long-term procedural knowledge of collocations at a faster rate than exposure to typographically enhanced collocations. The discovery of a priming effect in the present experiment, where Sonbul and Schmitt (2013) found none, may be explained by the extra day of reading sessions (two days vs. one) that includes overnight memory consolidation (Lindsay & Gaskell, 2010); the greater number of sessions (three sessions vs. one); and more exposures to the target collocations (nine vs. three).

The different learning conditions in the present study promoted varying types of processing of the target collocations and different types of collocational knowledge (procedural and declarative). The reading-only condition facilitated the development of procedural knowledge through the repeated co-occurrence of the two component words of the target collocations in meaningful contexts. In the bolding condition, learners received support in explicitly identifying collocations as lexical units, within the context of a sentence, without creating additional processing demands. This resulted in the most effective development of declarative collocational knowledge compared to the other two conditions. The bolding-plus-glossing condition may have provided additional information about the meanings of the collocations but drew the participants’ attention away from the context and created an extra processing burden. This could have negatively affected the participants’ ability to commit collocations to long-term memory. Thus, this condition was less effective than the bolding condition in establishing the declarative collocational knowledge needed for their recognition and retrieval.

Applying the results of the study to a teaching context, it is possible that procedural knowledge of target collocations can be acquired incidentally through the use of texts such as graded readers, which could be manipulated to include multiple instances of the MWUs. Repetitions of both single words and MWUs in graded readers have previously been found to produce gains in receptive declarative knowledge of the target items (Waring & Takaki, 2003; Webb, Newton, & Chang, 2013).

References


Lindsay, S., & M. G. Gaskell. (2010). A complementary systems account of word learning in L1 and L2. Language Learning, 60, 45–63.


What Does It Mean? Compulsory Testing in a Japanese University Context

Myles Grogan
Kansai University
myles1000sei -at- gmail.com

1 Introduction
Whilst many would argue that vocabulary is a cornerstone to language learning (e.g. Folse, 2004), to what extent should it influence grading? This presentation looks at grading in the Japanese university context, with emphasis on vocabulary learning. It follows the introduction of a weekly university-wide multiple choice vocabulary quiz as part of the overall assessment scheme. The author—a teacher within the context—explores the impact of that testing on student performance, and the reaction of students to the tests in terms of how it affects their learning experience. The circumstances described are of particular interest in that the testing was imposed on teachers from a higher authority. The author has sought to investigate what claims can be made as a result of this change.

2 Background
The study focuses on a large private university in Western Japan. In this context, teachers may teach in different departments using the same material, although level and needs may vary. Grading activity is left in large part up to the instructor. While percentages for parts of the grade were decided (e.g. midterms worth 20%), the content of the grades was left to individual instructors. A midterm assessment may be a speaking test for one teacher, a presentation for another, or a paper-based test in a different class. In the spring of 2015, however, a course of weekly vocabulary testing was introduced to the first-year English language program. This was seen by some in the administration as a first step in creating a unified curriculum, with further change to follow. Assessment based on this vocabulary quiz is worth 20% of each student’s final grade.

Testing was based on sections of the JACET wordlist (see, for example, Uemura & Ishikawa, 2004). A special edition of the wordlist was produced for purchase by the students, based on the first 3,000 items in the wordlist. The content to be studied was the same regardless of class level. Students were required to study prescribed sections for weekly multiple-choice tests, and classes began with a test on those items. Sections contained slightly less than 40 headwords, as well as related words (including copyright with copy, for example).

In-class quizzes contained 10 items per class, and students took the quiz in 12 classes during the 15-week semester. The items consisted of an English sentence with the target word replaced with a Japanese translation. Students selected the best choice from four English words to replace the Japanese word. For test security, several forms of the quiz were prepared for each week. After class, teachers passed the completed papers to a third party. This party returned the grades at the end of the semester after classes were completed.

3 Procedures
The investigation was done in the second semester of the 2015-16 academic year, after students were used to the format. The project aimed to look at gains in the target vocabulary over one semester using a pre- and posttest procedure. A test was constructed using all items from one of the forms of the quiz in the first class of the semester, before actual in-class testing began. The same test was administered in the last class of the semester, and results were compared.

In addition to the quantitative data, students’ attitudes to the quizzes were sampled using a Narrative Frames model (Barkhuizen & Wette, 2008).

4 Findings
Results were analyzed and a small gain in vocabulary was found, equal to about 5 words per person on just under a quarter of the material. Under perfect conditions, this would translate to a gain of about 20 words per person over the 15 classes in the semester. Although initially disappointing, the trend seemed more positive than has been found elsewhere in Japanese students after regular high stakes testing ceases (McLean, Hogg, & Kramer, 2014; Okamoto, 2007).
Review of student responses showed that students generally saw the tests as worthwhile, and students valued vocabulary learning. Students wanted more regular feedback about their weekly score, however. Opposing views were seen with regard to test format and content, however, with some students finding the material or format too easy. Others saw this as a good thing in terms of lowering their performance anxiety.

5 Recommendations based on findings

At present, the material to be learned stands alone, with no apparent connection to required textbooks or classtime given over to instruction. Making a more explicit link between the required texts and the JACET Word List would be beneficial. It is also suggested that this assessment system would benefit from matching with pedagogical approaches to vocabulary learning and associated metacognitive skills (Mizumoto & Takeuchi, 2009). Findings from the narrative frame suggest that more direct feedback to the students concerning their scores would also be beneficial. Finally, since many students found the material to be easy, allowing students the opportunity to test out of the system would create a more genuine sense of achievement and a clear goal to attain. Following this, options for moving beyond this basic level can be explored.

6 Conclusions

Questions remain concerning the return on investment in terms of student and teacher time, the infrastructure development to get books and grade tests, and the validity of the assessment in the wider context of English language learning and use.

However, although the gain in vocabulary from this required part of the course was in some ways disappointingly small, it is also a step in the right direction. Even a small gain in vocabulary is contrary to the trend in some university settings. Definite improvements could be made with additional attention to pedagogy in vocabulary, rather than concentrating solely on its assessment.

7 Acknowledgements

Special thanks for the cooperation of Stuart McLean in gathering and analyzing additional data for this project.

References


Meeting the Challenges of Researching L2 Collocational Competence in Domain Specific Language

Birgit Henriksen
University of Copenhagen

Pete Westbrook
University of Copenhagen

birgit -at- hum.ku.dk

1 Introduction

An increasing number of universities across Europe are introducing English-medium instruction (EMI) study programmes, especially at the Master’s level. This has led to a situation where tenured lecturers are teaching high-level academic courses to students in a language which is the L2 for both parties. To respond to the challenges of EMI, the University of Copenhagen has introduced a certification procedure involving an oral proficiency test of lecturers’ English (TOEPAS, http://cip.ku.dk/english/certification). Our study draws on data from 12 mini-lectures recorded from this test with lecturers from two different university departments (Large Animal Sciences and Mathematics).

Inspired by research which has shown that even advanced L2 learners experience problems with producing consistently correct and appropriate collocations (e.g. Nesselhauf, 2005), the present study investigates the lecturers’ L2 collocational competence. Henriksen (2013) has argued that conflicting results across previous collocation studies may be due to differences in the methodological approaches taken. In addition, providing a comprehensive description of collocational use in domain specific language has posed a range of new research challenges. The focus of our study is therefore on finding ways of meeting the many challenges of researching collocational use in domain specific language.

2 Collocations

Collocations are generally defined as two or three word combinations, i.e., “conventionalized recurring word combinations exhibiting more or less restrictedness, more or less semantic opacity and a certain degree of predictability for native speakers” (Handl, 2009:70). They can be divided into lexical collocations, i.e., combinations between nouns, verbs, adjectives and adverbs (strong argument, take sides, truly happy) and grammatical collocations, i.e., combinations which include prepositions (afraid of, present to, craving for). L2 learners' collocational competence has been widely studied (e.g. Barfield & Gyllstad, 2009; Henriksen, 2013), but many of these studies have relied on written corpora data of student essays, and have investigated only specific types of general collocations often limited to the lexical verb-noun and adjective-noun constructions. Little research has, however, been conducted on the overall collocational competence of non-native academic staff in oral language production.

The main reason for focusing on collocations in relation to EMI language use is that collocations typically have a highly referential function (Howarth 1998), as opposed to the social or pragmatic functions of other types of formulaic units, and tend to be very genre specific. As such, collocations are often a characteristic feature of technical subLANGUAGES (Ananiadou & McNaught, 1995). Similarly, mastery of collocations may be a hallmark of academic language which emphasizes clarity, precision, and lack of ambiguity (Howarth, 1998). Even very advanced L2 users seem to have problems with using collocations (Henriksen 2013; Nesselhauf, 2005; Laufer & Waldman, 2011), and, apart from leading to unfortunate misunderstandings, advanced non-native speakers’ collocational deviations may signal a lack of academic expertise (Henriksen, 2013:37). Because of the complexity of professional academic discourse found in the certification data, the authors found it necessary to make the distinction between general, academic and domain specific collocations in line with the single word item distinctions brought out by Nation and Hwang (1995).

A number of studies have provided empirical evidence suggesting a positive link between formulaic language and spoken fluency (e.g. Wood, 2010). It could therefore be assumed that collocational competence will also increase L2 learners’ fluency. However, in a recent study of oral data produced by non-native university lecturers, no correlation was found between collocational use and various fluency measures (Westbrook, 2015). In line with a number of previous studies, this project only focused on general lexical and grammatical collocations, disregarding both academic and technical collocations, which are a prominent feature of academic, domain specific language. It was hypothesized that a relationship between fluency and collocational competence may have been established, if a broader range of collocational types had been included in the analysis. Westbrook (2015), who investigated the same data set as the present study, found significant differences in results between including and not including domain-specific collocations into the
3 The research issues addressed in the study

Providing a more comprehensive description of collocational use in domain specific language, however, poses a range of serious research challenges in relation to identifying, quantifying, and distinguishing between general, academic and domain specific collocations. Trying to delimit what does and does not constitute a collocation combined with distinguishing between general, academic, and domain specific collocations proved to be an extremely complex task in this context. A number of methodological questions will be raised and exemplified in the analysis, with a focus on the following issues:

- How does an electronic search for collocations compare with a manual search?
- What difference does it make if we focus on specific types of lexical collocations compared to analysing all types of collocations, including the different grammatical types?
- What are the challenges in analysing complex embedded collocations?
- What are the specific challenges in including domain specific collocations in the analysis?

In the analysis, we present a possible approach to these challenges, exploring the effect of different methodological choices on the collocational competence profiles of our informants. The study reports on the differences found between measures of general, academic, and technical collocations and discusses the differences in collocations used across the two academic disciplines studied. The study highlights the need to reach a more shared agreement on applying the same criteria and coding procedures if we want to compare results across studies. It also emphasizes the importance of studying academic language use across different disciplines in order to detect cross-disciplinary differences and similarities in academic language.

References


The Cross-Sectional Development of Collocational Knowledge of Japanese Senior High School Students

David Coulson  
Ritsumeikan University  
coulson -at- fc.ritsumei.ac.jp

Naoaki Tanaka  
Ritsumeikan University  
nktnk0 -at- gmail.com

1 Background

This study investigated collocational knowledge growth in EFL senior-high school students. The students in all three years of a senior high school in Japan gave three responses to stimulus word on a test of collocational knowledge and the development of their knowledge was examined in the light of the total number of correct responses, word frequency, and word class they produced. The result showed that the second and the third graders generated statistically the same amount of correct responses while the number was significantly larger than the first graders. That is, collocational knowledge growth stopped developing from the second to the third year. We propose that the main factor for the third graders’ stagnation of development is insufficient exposure to high-frequency words.

Nation (2013) argues that collocation is use-based knowledge for knowing words. Collocational knowledge is necessary for language use, and a lack of collocational knowledge hinders it. Pawley and Syder (1983) also argued for this phenomenon and suggested many words are stored firstly as individual words (or unanalyzed chunks) and then as larger analyzed-chunks after many repetitions. This process underpins the development of fluency and language use (e.g. Ellis, 2001; Wray, 2002). Finally, Tono (2007) examined middle and high school learners’ growth of lexical collocation through analyzing the learner corpus based on English-free writing by junior and senior high students.

2 Research questions

Overall, the reviews of collocational studies in Japan indicate that English education at SHS does not put emphasis on collocation in terms of textbooks and teachers. To respond to these issues, this study raises the question expressed below:

Research question 1: To what degree does the crucial aspect of collocational knowledge of SHSs develop under the current educational settings?

Research question 2: What is characteristic of SHSs’ response patterns?

3 Subjects and the study

Subjects were 113 EFL student at a Super English Language High School. There were 39 first-year, 36 second-year, and 37 third-year students. A test to assess productive knowledge of collocation was developed following LexCombi (Barfield, 2009). This is an association task comprised of 30 stimulus words, and testees are required to give three collocational responses for each stimulus word. The criteria for inclusion of words were the following:

1. High-frequency, academic and low-frequency vocabulary (1K, 2K, AWL, and Off-List words) were included.
2. Words which were thought likely to produce common and dominant responses.
3. Words unlikely to elicit the same responses as those to other words.

A database for scoring was created based on the database by Barfield, (2009) and additionally on the New Oxford Collocation Dictionary (NOCD) published in 2009. Inflected verb responses were accepted if the database includes the basic present tense form of the verb. Uninflected verbs were accepted if the database includes inflected forms of the verb. Derivational words were not counted as appropriate collocates.

Erroneous spellings were tolerated as long as they are identifiable by comparing them with the correct form and finding a near match.

4 Results

The reliability of the test was confirmed (Cronbach α = 0.75). For tokens, a one-way ANOVA analysis confirmed that there is a statistically significant difference among three ($F(2, 109) = 8.27, p = .000, η^2 = .389$). Next, a multiple comparison test (Tukey HSD) clarified that the score of Y1 is significantly different from that of Y2 ($p = .008$) and Y3 ($p = .001$), but not between Y2 and Y3. The same procedure was taken toward the
types. It was found that there is a significant difference among three years. \( F(2, 109) = 10.17, p = .005, \eta^2 = .432 \).

### Table 1. Total Scores of Tokens and Types.

<table>
<thead>
<tr>
<th></th>
<th>Tokens</th>
<th></th>
<th>Types</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Y1</td>
<td>33.90</td>
<td>7.27</td>
<td>26.95</td>
</tr>
<tr>
<td>Y2</td>
<td>39.80*</td>
<td>7.93</td>
<td>32.44*</td>
</tr>
<tr>
<td>Y3</td>
<td>41.19*</td>
<td>9.62</td>
<td>32.56**</td>
</tr>
</tbody>
</table>

Overall, comparison of the scores and types reveals that collocational organization of Y3 and Y2 are statistically alike, and both significantly different from Y1. This fact (Y3 > Y2 > Y1) can be witnessed in word frequency while the proportion of less frequent production was relatively small in all grades. This indicates that Y2 and Y3 can access the similar amount of collocational links and also can similarly produce less frequent collocates. The finding differs from the hypothesis (cf., p. 14) based on their estimated development of vocabulary size.

### 5 Conclusion

The main finding of this study is that although there was a significant increase in the mean score of the first-year and second-year students, the collocational skill appeared to cease between the second and third years. Also, this study suggests the need to reconsider the relation between vocabulary size and organization since the current result exhibited the distinctive developmental pattern compared to previous research. Comparison with similar research findings (Coulson, 2014) which indicate a similar cessation of word knowledge development will be noted, and implications for improvement to vocabulary instruction for learners at the secondary level will be offered.

### References


Introduction
Since Paul Meara’s 1987 paper suggesting vocabulary was no longer neglected, vocabulary research has boomed. Today vocabulary research and pedagogy is vibrant, but this is not to say that all research strands have proved productive. I have been researching and thinking about vocabulary now for very many years, and have seen a number of trends unfold. This presentation will give my ‘big picture’ perspective on the development of vocabulary studies since Meara’s article.

Advances in research
A selection of important advances include:
- A consensus that 95%+ coverage is required
- The rough establishment of vocabulary size targets for doing things in English (but mostly reading)
- A consensus that around 10 exposures for incidental learning
- A better understanding of the vocabulary required for academic and specific purposes
- The development of word lists (although of widely varying quality)
- The use of psycholinguistic measurement techniques
- Several key messages seem to be filtering down to practitioners:
  o Formulaic language is important
  o Knowing a word is more than just matching its form and meaning
  o Extensive reading as important for vocabulary growth

Remaining gaps
Vocabulary scholars can congratulate themselves on these advances. However, I see there are still many gaps and challenges to address, and this presentation will focus on what did not quite work out in vocabulary research, and where we need to go in the future, e.g.:
- A principled approach to vocabulary selection and recycling is typically not used in textbook writing
- Although vocabulary/reading research is buoyant, there is much less effective research on the other skills. We do not have a much of an idea of how much and what type of vocabulary is required for the productive skills, and vocabulary/listening research is just beginning
- Unlike incidental learning, we do not know the number of repetitions that are necessary for learning from intentional learning, or which factors affect this and to what degree
- Vocabulary tests are almost without exception not validated adequately
- The availability of corpora and the ease of the Internet means that much material (e.g. tests and word lists) are disseminated without proper vetting and validation
- There have also been some major research strands that have not really delivered in terms of concrete pedagogical applications. Among these, association and strategy research seem to have hit virtual dead ends.

This presentation will outline my views of where vocabulary research needs to focus its energies in the near and mid-term future.
Vocabulary in English for Chinese Medical Purposes: What Can a Corpus Tell Us?

Cailing Lu

Victoria University of Wellington
Cailing.Lu -at- vuw.ac.nz

1 Introduction

With the development of Traditional Chinese Medicine (TCM) and its recognition over the world, there is a greater need for students of this discipline to acquire English-language literacy in their specialized field. As found in a needs analysis of an English for Chinese medical purposes course (Lu, 2014), vocabulary learning was perceived as one of the most challenging aspects in students’ specialized English-language learning course. At the same time, it is becoming increasingly more common for institutions in English speaking countries, such as New Zealand and Australia, to include TCM as a degree course. For instance, University of Technology, Sydney provides the Bachelor of Health Science in Traditional Chinese Medicine. Vocabulary learning also poses challenges to learners from these degree courses. To meet the vocabulary needs of these TCM learners, it is of great importance to understand the lexical demands in the specific target settings. Further, having a heavy load of specialized or technical vocabulary is one of the prominent features of English for Specific Purposes (ESP) (Nation, 2013). The present study, therefore, sets out to investigate the vocabulary loads necessary for reasonable comprehension of the widely-used textbooks and journal articles in this specialized field, which are the two prominent genres that students are likely to read in their target situation.

2 Method

This study investigated the vocabulary loads of English-medium textbooks and journal articles in Traditional Chinese medicine. Vocabulary load was the number of 1000-word-family lists needed by counting the frequency-based BNC/COCA word lists from the first 1000 word families (Nation, 2006). The lexical coverage of each 1000 word family in the target texts was calculated and added until coverage reached 98%, which is the minimal vocabulary necessary for reasonable comprehension (Schmitt, Jiang, & Grabe, 2011).

Similar to other medical students who proceed from pre-clinical theory-based years to clinical settings (Wette & Hawken, 2016), TCM learners usually start with the TCM theories and then proceed to clinical practices. Textbooks used at these two stages differ considerably from each other. For this reason, this study adopted a new way to categorize the textbooks by dividing the textbooks into the theory-orientated textbook corpus and the practice-orientated textbook corpus. The vocabulary loads of both textbook corpora and the journal articles corpus were then measured against Nations’ (2012) BNC/COCA lists of word families from the 1,000 to 25,000 word levels with four additional lists. The words that appear in the category ‘Not in the lists’ by Range (Heatley, Nation, & Coxhead, 2002) were examined and categorized.

Technical vocabulary is a special group of vocabulary that are fairly high frequency words in a specialist domain but might be low frequency words in other domains (Chung & Nation, 2003; Nation, 2013). The lexical items which occur in the TCM texts but are not included in any of the BNC/COCA lists were thus made into the TCM technical word lists. In order to reach the 98% coverage, two TCM technical word lists were compiled and added to the existing base word lists. To be specific, one TCM word list includes all loan words that borrowed directly from Chinese, for example, qi and yang; and the other TCM word list contains words that do not belong to any of the word lists mentioned above. Latin names of herbal medicines (e.g., *Rhizoma, Atractylodis*) are typical examples. In this study, the total lexical coverage of each 1000 word family was calculated and added to proper nouns, abbreviations, apparent compounds, and technical vocabulary.

3 Results

Table 1 shows the preliminary results of the overall vocabulary loads of the English-medium Chinese medical textbooks and journal articles in the field of Traditional Chinese Medicine. Along the BNC/COCA word-frequency scale, the TCM textbooks involved varying vocabulary thresholds ranging from the first 11,000 most frequent word families to 14,000 word families. Among them, practice-based textbooks have the largest vocabulary load whereas the journal articles have the smallest vocabulary load, indicating that the
practice-based textbooks are the most lexically demanding while the journal articles are the least lexically demanding.

<table>
<thead>
<tr>
<th>Text Type</th>
<th>Vocabulary Load</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theory-based textbooks</td>
<td>13,000</td>
</tr>
<tr>
<td>Practice-based textbooks</td>
<td>14,000</td>
</tr>
<tr>
<td>Journal articles</td>
<td>11,000</td>
</tr>
</tbody>
</table>

Table 1. Vocabulary Loads of the TCM Textbooks and Journal Articles at 98% Lexical Coverage Including Abbreviations, Apparent Compounds, and Technical Vocabulary.

A close look at the most lexical demanding practice-based textbooks show that the first BNC/COCA 1000 word families accounted for 53.89% of the total tokens. This is much lower than the average first BNC/COCA 1000 word families of 78 – 81% estimated by Nation (2006). However, the combined coverage of the two technical word lists accounted for 9.12%. This is higher than that of the second 1000-word-family list. The following paragraph is an example text from the practice-based textbooks with TCM technical vocabulary underlined.

This condition is very chronic and occurs on a background of Spleen Yang deficiency and often Kidney Yang deficiency. In such cases one should add herbs to tonify Spleen and Kidney Yang such as Baizhu Rhizoma Atractylodis macrocephalae, Huangqi Radix Astragali membranacei, Yinyanghuo Herba Epimedii and Duzhong Cortex Eucommiae ulmoidis.

The findings suggest that TCM learners have a rather heavy vocabulary burden, especially for technical vocabulary. It is thus better to direct TCM learners’ attention to technical vocabulary early in their studies.

References


1 Introduction
The present study looks into the under-researched area of collocation use with learners from different proficiency levels by looking into the type of collocations and the learners’ elaboration on their collocation use. (Bestgen & Granger, 2015; Durrant & Schmitt, 2009; Li & Schmitt, 2009). 194 students from three different years of study in a Chinese university were asked to write a composition on the same topic, and 42 of them were invited to post-writing interviews on collocation use. The compositions were used to retrieve collocations, and the interviews intended to explore learner’s reflections on collocation use. Three types of two-word collocation, i.e., verb+noun, adjective+verb, noun+noun, were extracted from the writings manually and used for statistical analysis of the frequency of occurrence and mutual information (MI) in the reference corpus of Corpus of Contemporary American English (COCA). The post-writing interviews covered the questions exploring the source of collocation learning and the type of knowledge learners have when they felt confident using collocations in writing.

2 Results
The results show discouraging evidence that learners fail to show consistent improvement in the use of collocation in terms of frequency and association strength with the increase of proficiency both at the group level and individual level. The overwhelming majority of the collocations in learner writing are below the MI threshold, which means they appear more likely due to chance and show low “exclusiveness” between collocates. These collocates are either erroneous or creative combinations. The error analysis into this group of collocations shows that there are very few errors of spelling, morphology, and word coinage, where most of which are creative combinations with no evident grammatical errors. The pervasiveness of the creative combinations was later confirmed in the post-writing interviews when learners expressed their habits of combining two grammatically possible words as collocations since their confidence in using collocations were greatly dependent on grammatical correctness than other aspects of collocational knowledge.

3 Conclusion
By analyzing the most frequent collocations and strong collocations, the present study substantiates former studies in that learners rely on a group of highly frequent collocations in their writing. However, they are able to use collocations with great heterogeneity and little overlap between strong collocations in learners writing at different proficiency levels. In the post-writing interviews, learners expressed their cautiousness in using novel collocations learnt in college and reliance on the high-frequency collocations learnt from high school. The present study also found out that learners at an advanced stage of learner need both awareness for collocations, and more importantly, awareness for low-frequency collocations with high MI scores. It is important to introduce the notion of association strength to advanced learners and help them understand the exclusive relation between collocates.

References
A Fresh Look at Measuring Depth of Vocabulary Knowledge

John Read
University of Auckland
ja.read -at- auckland.ac.nz

1 Introduction

An enduring distinction in second language vocabulary testing is that between breadth and depth of vocabulary knowledge. The construct of breadth, or vocabulary size, has been relatively easy to define, and to operationalise as tests of the form-meaning link. On the other hand, depth of vocabulary knowledge has proved to be more problematic, both to define as a distinct construct and also to measure, despite various efforts at construct definition over the years. A recent review article by Schmitt (2014) makes it clear that uncertainty about the depth construct persists to this day, despite the existence of a substantial body of empirical research which has sought to measure it as a distinct dimension of vocabulary knowledge.

2 The word associates format

Read’s (1993, 1998) two word associates tests have been widely used as research tools and the 1998 version in particular is regarded by many writers as the standard measure of depth of vocabulary knowledge. However, this version has limited value for any operational purpose because it was designed specifically to measure knowledge of adjectives as target words and the actual words were taken from an ill-defined domain of high-frequency adjectives.

The starting point for the present project, then, is to develop a new test format which samples from an inventory of academic vocabulary across the major lexical word classes. The design of the test takes account of research since the 1990s on the word associates format (Read, 2012), notably by Schmitt, Ng & Garras (2011); Schmitt’s (2014) comprehensive review of the literature on breadth and depth of knowledge; and the work by Meara (2009) and his associates exploring word association formats to gain insights into the state of the mental lexicon, rather than knowledge of individual words.

3 The project

Rather than assessing general academic vocabulary, the test in its present form targets Business English vocabulary, drawing on the Commerce sub-corpus of the British National Corpus. A new corpus analysis tool called GraphColl (Brezina, McEnery & Wattam, 2015) has been used to identify high-frequency collocations, since the syntagmatic relationship between words is a key component of the word associates format. GraphColl can analyse and display multiple-order collocations among networks of words in a corpus, and thus it provides a more empirical basis for writing test items than the subjective judgements used to develop items for previous tests. The current focus is on refining the conceptual basis of the test design and developing a prototype format. Its viability is being investigated by reference to expert informants in Business to confirm that the correct answers to the test items (based on corpus evidence) do represent “deep” knowledge of vocabulary in their field, of the kind that undergraduate Business students should be expected to acquire. This will lead to trials with learners to evaluate the measurement qualities of the new test.

References


A Framework for Grading Vocabulary: Combining Corpus Data and Teacher Judgments

Veronica Benigno
Pearson English
veronica.benigno –at– pearson.com

1 Introduction

Vocabulary is a crucial component of communicative competence. Although much research has been conducted to define vocabulary knowledge both in terms of quantity (size) and quality (depth), a reference framework which establishes what and how many words learners should master at increasing proficiency levels is still missing and existing estimates on vocabulary size (Hazenberg & Hulstijn, 1996; Nation, 2006) are sometimes conflicting. The dominating approaches to vocabulary teaching and assessment have used corpus frequency as the main selecting criterion to produce ordered lists of single words to serve as teaching or assessment targets. Such lists have taken the lemma or the word family as unit of counting, without an indication of what and when different word meanings should be learned, and given exclusive relevance to the frequency of occurrence of words in the input, without insight into their communicative usefulness (Stubbs, 2002). For instance, the different meanings of the word date as “day of the month”, “romantic engagement” and “sticky brown fruit” have the same rank in a corpus frequency list – somewhat contradicting research showing that vocabulary learning is a gradual process which generally goes from basic/useful words and meanings to more complex/specialized ones (Brent, 2009).

2 The Global Scale of English vocabulary inventory

The CEFR (Council of Europe, 2001) is a language-neutral framework which describes what learners can do at increasing proficiency levels. The Council of Europe made a recommendation to all member states to create, for each language, inventories of lexical and grammatical words known as “Reference Level Descriptions” – RLD- (http://www.coe.int/t/dg4/linguistic/DNR_EN.asp). The RLDs are defined as “inventories of the linguistic realisations of general notions, acts of discourse and specific notions / lexical elements and morphosyntactic elements” which are characteristic of each level (Council of Europe, 2005, p.5). Pearson has carried out a research project to create a graded vocabulary inventory aligned to the Global Scale of English (GSE, Pearson metric to measure English proficiency) and the CEFR. The GSE vocabulary inventory outlines the lexical exponents needed to acquire the competences described in the framework, with the ultimate goal of increasing the efficiency of language learners in achieving their communication goals. The inventory is freely available at http://www.english.com/gse/teacher-toolkit/user/vocabulary. It was produced by combining frequency data (based on corpus investigation of L1 language use) with qualitative information about usefulness of vocabulary (provided by language teachers). It indicates how many and what word meanings learners should be able to understand at increasing proficiency in order to communicate efficiently in the target language. It is of very large size, including more than 20,000 lemmas, 36,000 word meanings, 80,000 collocations, and 7,000 phrases.

3 Data and methodology

The creation of the GSE Vocabulary inventory consisted of four main steps: corpus analysis; semantic annotation; teacher ratings; and link to the Global Scale of English and the CEFR. The first methodological step required corpus analysis and computational processing of the data. Three corpora which differ in size and content were analysed: LCN (Longman Corpus Network), a corpus of 330 million words used as the basis for all Longman dictionaries; UKWAC (http://wacky.sslmit.unibo.it/doku.php?id=corpora; Baroni et al., 2009), a 2 billion word corpus constructed by crawling the web. This corpus is freely accessible, POS-tagged, lemmatized, and searchable using a commercial interface such as SketchEngine; and COCA (http://corpus.byu.edu/coca/), a corpus of contemporary American English of 450 million words. For our study, we selected COCA spoken component only (about 90 million words). The decision to investigate three different corpora was motivated by the need to compensate for the limitations of using a traditional corpus (e.g., a more limited size) or a web-based corpus only (e.g. unbalanced and less controlled). The Web has evolved into a powerful resource for corpus analysis because of its size, authenticity, linguistic and socio-linguistic variety, and up-to-dateness. Our selected data
A sample consisted of the top 10,000 lemmas occurring in the three corpora and all entries included in LASDE (Longman Active Study Dictionary of English – an intermediate learner’s dictionary), accounting for a total of about 25,000 lemmas.

In a second step the obtained frequency list including more than 20,000 lemmas (corresponding to 36,000 word meanings) was semantically annotated using the Council of Europe categorization in Specific Notions, General Notions, and Functions found in the Vantage Specifications (van Ek & Trim, 2001). A team of expert lexicographers manually assigned tags to each word meaning so to organize vocabulary around topic and subtopics, e.g. “fork” was tagged as topic “Food and drinks”, subtopic “Kitchen equipment, utensils and appliances”; topic “Holidays, travel, and transportation”, subtopic “Road or rail network”; topic “Sports, hobbies and interests”, subtopic “Gardening”.

In a third step, vocabulary was rated by a pool of 19 teachers who were asked to rate word meanings based on the principle of usefulness and using a pre-defined scale ranging from 1 to 5. Each word meaning was rated by 10 teachers. Teachers were provided with a written briefing and received online training.

In a final step, the information on the frequency of words (extracted from corpus) and on the usefulness of words (provided by teachers) was combined to produce a weighted algorithm to rank vocabulary. The data were fit onto a model of vocabulary learning based on current research evidence on vocabulary size so that each word meaning could be linked to the Global Scale of English and the CEFR.

4 Final considerations

Our study seeks to align the learning, teaching, and assessment practice by providing a clear description of the intended vocabulary goals for adult learners of general English. The GSE Vocabulary inventory uses a probabilistic model to determine what vocabulary (counted as word meanings) gives learners the best chance of successfully communicating in English.

References


Can Knowledge of Multi-Word Expressions be Fostered through a Fluency Workshop?

Haidee Thomson
Muroran Institute of Technology & Victoria University of Wellington
haidee.thomson -at- vuw.ac.nz

1 Introduction

Multi-word expressions (Siyanova-Chanturia & Martinez, 2015) constitute a broad category of frequent and familiar word sequence, such as I think I will, would you like to. Multi-word expressions (MWEs) are prevalent throughout language and often associated with fluency (Boers et. al., 2006). So what types of learning activities promote the use of multi-word expressions while building spoken fluency? Nation & Newton (2009) claim that to develop fluency, activities should be meaning-focused, use familiar language, and be performed under time pressure. This study replicates and expands on a 6-week fluency workshop (Wood, 2009) in an ESL context, which claimed to build oral fluency through a sequence of language learning activities focused around target expressions. Oral fluency is often neglected in English learning in Japan, so if proven to be effective in this EFL context, the fluency workshop could be a valuable resource for teachers. Therefore this action research (the first of three studies) seeks to answer the following questions:

1) Does the fluency workshop increase spoken fluency in the EFL context of Japan?

2) Are fluency gains (if any) attributable to the increased use of target expressions from the fluency workshop?

2 Participants and materials

Participants (N=73) came from intact compulsory English classes in an engineering university in Japan. In general, the students’ listening and speaking abilities in English could be described as low, with 67% reporting a TOEIC score range of 10-400. Participation was voluntary and consent collected after information was given explaining the study.

Materials for the experimental condition were chosen and developed to help build basic ability to communicate using English in particular travel related situations. A model dialogue for each situation was created by the researcher and the vocabulary profile checked using lextutor (Cobb, 1994), to ensure predominant use of high frequency vocabulary.

Each model dialogue included 10 target expressions (four words in length) which were either chosen from within the text or integrated into it, resulting in a total of 30 target expressions. Common usage was checked using frequency data from the spoken Corpus of Global Web-Based English (Davies, 2013). The model dialogues were audio-recorded and used for listening activities in class. The target expressions were read, heard, written, and spoken multiple times in classroom activities over six weeks, as listed in Table 1.

| 1 | Listen with gist questions |
| 2 | Mark pauses |
| 3 | Phrase instruction |
| 4 | Shadowing |
| 5 | Dictogloss |
| 6 | Mingle jigsaw |
| 7 | Role-play |
| 8 | Decreasing time role-play |
| 9 | Record role-play |
| 10 | Free related situation role-play |

Table 1. Fluency Workshop Activities in Sequential Order.

The control condition involved speed reading and linked skills progression on engineering related topics. Students in this condition experienced the pre and post-tests, but they were not otherwise deliberately exposed to the target multi-word expressions for this study.

3 Methodology

In order to test the influence of the fluency workshop, participants were given pre and post-tests. 1) A MWE cloze test to measure productive knowledge of the 30 MWEs, 2) A recorded role-play to measure spoken fluency and use of MWEs, and 3) The Listening Vocabulary Levels Test (McLean et. al., 2015), an external test to measure aural vocabulary knowledge for the most frequent 3,000 words in English. The first two tests were internal to this study, measuring target MWEs, while the Listening Vocabulary Levels Test was chosen...
for potential triangulation with these internal measures.

4 Results and discussion

The experimental and control group scores did not differ significantly in any of the pre-tests, suggesting comparability on these measures. After six weeks, the same tests were administered. There was no difference between the groups in average aural vocabulary scores pre \( (N=73) \) \( M=57.33, SD=6.03 \) and post-test \( M=58.30, SD=5.87, t(72)=1.71, p>.05, r=.08 \). However, results from the MWEs cloze test showed that productive knowledge of target expressions increased significantly more for the experimental participants \( (n=44) \) \( M=7.63, SD=4.12 \) than for the control participants \( (n=29) \) \( M=2.55, SD=3.01, t(71)=-5.71, p<.05, r=.58 \).

There were 23 dialogue recordings produced by participants working with the same partners in both the pre-test and the post-test. The audio was transcribed for each of these dyad interactions. Fluency was measured by speech rate (syllables uttered per minute). The experimental group significantly increased their speech rate \( (M=47.03, SD=16.39) \) in the post-test \( (M=57.95, SD=18.49), t(14)=-3.77, p<.05, r=-0.29 \), however, the increase was not significantly greater than the control group \( F(1,21)=0.454, p=.51, \eta^2=.02 \). In Wood (2009), the speech rate increased by 13.8%. In comparison, the experimental group’s speech rate in this study increased by 28.3%. This suggests that the fluency workshop can build spoken fluency, but further investigation is necessary.

In the experimental group, MWE use \( (M=6.20, SD=4.07) \) significantly increased in the post-test role-play \( (M=11.00, SD=7.34), F(1,14)=4.81, p<.05 \), but it did not differ significantly from the control group \( t(21)=-.877, p>.05; r=.20 \). It is interesting that, while the experimental group was able to produce more MWEs in the cloze test, it did not extend to greater use in conversation when compared with the control group. This could reflect the incremental nature of learning MWEs and perhaps suggests that writing knowledge precedes oral knowledge. It could also be a reflection of the test type, in that the role-play did not force the use of the target MWEs whereas the cloze test did.

In the experimental group, the mean length of MWE significantly increased from pre-test \( (M=2.03, SD=0.12) \) to post-test \( (M=2.29, SD=0.42), t(14)=-2.32, p<.05 \), with a medium effect size of \( r=0.37 \), but the difference in growth between the groups was not significant \( t(21)=-0.04, p>.05 \). Wood (2009) found that the mean length of formulaic sequences increased by 40.7%. In comparison the mean length of MWEs increased by 12.6% in this study. The results show that the speech rate positively correlates with an increase in the length of MWEs, \( r=.37, n=23, p<.05 \) (one tailed), suggesting that fluency gains to some degree can be attributed to the use of longer target MWEs from the fluency workshop.

This preliminary study is not without limitations. First, the numbers for analysis of the spoken dialogues need to be increased in order to see the fluency differences at statistically powerful levels. This is the first of three studies which will improve on and replicate the previous ones in order to discover ways to build fluency and unpack the relationship between fluency and the use of multi-word expressions.

5 Acknowledgements

This research was made possible with funding and support from Muroran Institute of Technology and Victoria University of Wellington. Many thanks go to Dr Averil Coxhead and Associate Professor Frank Boers from Victoria University of Wellington for their supervision of this continuing PhD research project.

References


Advanced Learners’ Lexical Inferencing: Investigating the Effects of Depth of Vocabulary Knowledge and Target Word Part-of-Speech

Elin Nylander
Lund University
elinnlyander4 -at- gmail.com

Henrik Gyllstad
Lund University
Henrik.gyllstad -at- englund.lu.se

1 Introduction

Irrespective of vocabulary size, learners occasionally need to compensate for their lack of word meaning knowledge during reading. A frequently used strategy in such cases is lexical inferencing, i.e., the ability to guess the meaning of unknown words in a running text using different types of knowledge and textual cues (Haastrup, 1991). Drawing on a taxonomy of inferencing sources proposed by Bengel and Paribakht (2004), the present study explores the relationship between advanced learners’ lexical inferencing behaviour and their vocabulary depth. The study aimed to examine whether learners with greater vocabulary depth made use of different types of knowledge and cues and whether they were more successful when inferring word meanings than learners with less lexical depth. An additional aim was to investigate whether the part-of-speech of an unknown word may influence the inferencing success.

2 Background

Even though both vocabulary size and depth are expected to play a role in inferencing, vocabulary depth has been demonstrated to make a stronger contribution to reading performance. Qian (1999) found that depth of vocabulary knowledge was not only a better predictor of L2 reading comprehension but that it also made a unique contribution beyond the contribution made by vocabulary size. Studies investigating the relationship between vocabulary depth and lexical inferencing have found that learners with greater vocabulary depth use more context-based lexical inferencing strategies more frequently than those with a weaker depth of vocabulary knowledge, who in turn use more local-word-based strategies (Nassaji, 2004). It has also been found that inferencing success rates are significantly higher for participants with a deep lexical repertoire than for those with less depth of vocabulary knowledge (Nassaji, 2004; Qian, 2005). The present study adds to the body of literature on the relationship between vocabulary depth and lexical inferencing behaviour of EFL learners. Methodologically, the study attempts to overcome some flaws in previous inferencing studies, e.g. to restrict participants to reporting inferencing only in writing (Ehsanzadeh, 2012), to allow reporting of inferencing only in the target language (Nassaji, 2003), and to use rather small groups of participants (Qian, 2005). In the present study, these shortcomings are addressed by the use of a decent size homogeneous L1 background participant group of EFL learners, who can verbalize their thoughts in the L1 or L2, in a spoken think-aloud procedure. The following RQs guided the study:

- What is the relationship between Swedish EFL university students’ lexical inferencing behaviour and their depth of vocabulary knowledge?
- Do the participants in this study evincing a deep lexical repertoire employ different knowledge sources than the informants with less vocabulary depth?
- Does the part-of-speech of a target word affect learners’ lexical inferencing processes?

3 Methods and materials

In the present study, two instruments were used: a 40-item vocabulary depth test (Word Associates Test, WAT) (Read, 1993) and a lexical inferencing task (LIT). The latter made use of a short text (250 words) specially designed for the study, containing 12 target words. As the participants were advanced learners of English, the task needed to contain quite infrequent target words. The words were therefore found within the 13K or 14K frequency bands of Paul Nation’s BNC-based frequency list (Nation, n.d.). A pilot study was used to verify that target words were expected to be unknown to the intended participants. In order to examine the relationship between inferential success and the part-of-speech of a target word, half of the target words were adjectives and half were verbs. The participants were L1 Swedish adult undergraduate EFL learners. Initially, 45 participants took the WAT in paper-and-pencil format. Out of these, 20 were selected to do the LIT. This selection of participants was based on the student’s WAT scores, where the aim was to include individuals with a range of depth scores, enabling a comparison between their depth and
inferencing behaviour. In individual audio-recorded sessions, they were asked to verbally infer the meaning of the target words and explain their reasoning in an introspective think-aloud procedure. The participants shared the researchers’ L1, and could therefore use Swedish, English or a mix of both in the LIT. It was assumed that this would allow them to express themselves as freely as possible, and thus generate valid data.

4 Results

Firstly, a correlation analysis indicated a positive relationship between inferencing task success scores and WAT-based vocabulary depth scores for the omnibus group, at  \( r_{(20)} = + .52, p < .05 \). In a subsequent analysis based on WAT scores, the participants were divided into a lexically skilled (LS) (\( n = 10 \)) and a less lexically skilled group (LLS) (\( n = 10 \)). The groups’ descriptive score distributions on the two tasks are shown in Table 1. A Mann Whitney U test revealed a statistical difference in mean vocabulary depth, LS (\( Median = 149 \)) and LSS (\( Median = 128 \)), \( U = -3.180, p < .001 \).

Both groups used more context-based inferences (~55%) than word-based ones (~35%). The LS group produced more inferences than the LLS group (55% vs 45%) and in terms of overall success rates, the LS group was more successful (~50%) when employing knowledge inferencing sources than the LLS group (~26%). The part-of-speech of a target word had a significant impact on the learners’ inference scores, with the score for verbs (\( M = 3.45, SD = 1.43 \)) being significantly higher than that for adjectives (\( M = 2.50, SD = 1.88 \)) for the omnibus group in a paired-samples t-test, \( t_{(19)} = 2.17, p = .043 \).

5 Concluding discussion

The results corroborate earlier results that hold that learners with greater lexical depth are more successful inferencers of unknown words than learners with lower depth (Qian, 2005; Ehsanzadeh, 2012). Moreover, the more lexically skilled group made more inferences overall, and were more successful, both for more word-centered and more contextual inferences, but in contrast to some earlier observations (Nassaji, 2004; Qian, 2005) they did not draw more on contextual clues than the less lexically skilled group. The fact that the participants scored higher in their inferences on verbs than adjectives is interesting, and Wesche & Paribakht (2010) have argued that nouns and verbs are more likely to attract inferencing attempts than other word classes. However, more analysis and further studies are needed to tease out the reasons for these results.

References


Teacher Language Knowledge as Applied to Lexical Error Analysis and Design of Vocabulary Instruction

Anne Ediger
Hunter College, City University of New York
aediger -at- hunter.cuny.edu

Christine Rosalia
Hunter College, City University of New York
crosalia -at- hunter.cuny.edu

1 Introduction

ES/FL teachers require various types of knowledge about language (KAL) (Fillmore & Snow, 2002) and are expected to use knowledge of grammar and vocabulary in their teaching practice (Bigelow & Ranney, 2005). Thus, novice teachers should have an understanding of grammar and vocabulary in order to develop effective language teaching curricula, diagnose and give feedback on student errors, design instruction to meet student needs, and have a shared language and knowledge base for discussing learner performance with colleagues (Andrews, 2007).

Considerable research now shows that at least in terms of grammar, teachers’ knowledge is not always transferrable to other areas of classroom teaching (Bartels, 2005), and teachers have difficulty using it for lesson planning (Baecher, Farnsworth, & Ediger, 2014; Baecher, Ediger, Farnsworth, in press), error correction, and using linguistic metalanguage for explaining how the language is used (Andrews & McNeill, 2005). What can we say about teachers’ ability to apply their lexical knowledge to error correction and determining the learning needs of their students?

While a number of researchers have investigated the grammar knowledge teachers need (Andrews, 2007; Bartels, 2005; among others), and the issue of correcting students’ grammar errors has spawned extensive discussion (see Ferris, 2002, among others), little research has investigated lexical errors in student writing (Hemchua & Schmitt, 2006), much less what teachers’ knowledge of vocabulary should consist of for addressing such errors. At the same time, lexical errors are often the most irritating (Santos, 1988), and more difficult to tolerate outside classrooms than syntactic errors (Carter, 1998).

2 The Data

Part of a larger study of novice teachers’ grammatical and lexical knowledge, data from 60 MA in TESOL teacher candidates’ end-of-course Essay Analysis fieldwork assignments in a large urban university in the United States were analyzed, investigating their understandings of English lexical information as evidenced through their identification, analysis, feedback, and design of language learning instructional activities for addressing lexical errors in their ESL students’ writing.

3 Research Questions

- Are our teacher candidates (TCs) able to identify their students’ errors as a) errors, and b) whether they are lexical errors?
- Can our TCs analyze/explain why these are errors (i.e., not target-like) to students?
- Can our TCs design instruction for their students, based on their understanding of the students’ error types and their analysis?

4 Results

A detailed analysis of the data showed degrees of both success and difficulty in the teachers’ knowledge, explanations, and treatments of lexico-grammatical errors. The difficulties include misdiagnosing lexical errors, inappropriately treating lexical errors as grammar concepts, failing to note that errors may simultaneously be both lexically- and grammatically-based, treating an error in ways inappropriate for its specific context, failing to recognize related facets of an error throughout the student’s writing, or over- or under-explaining errors. Candidates also showed a lack of awareness of appropriate online and other sources of lexical information for instructional purposes (such as collocational dictionaries, concordancing software, and searchable language corpora). An analysis of the teachers’ difficulties designing appropriate instruction to match students’ lexical needs suggests the need for explicit instruction in identifying and analyzing errors, wording explanations, and designing follow-up instruction, as well as several ways teacher educators can
foster an understanding of lexical knowledge more effectively in the preparation of ES/FL teachers.

Figure 1. Accuracy of Teachers’ Identification, Analysis, and Treatment of ESL Students’ Lexical Errors.

References


The Use of Psycholinguistic Formulaic Language in Japanese Speakers of English

Stephen Cutler
Cardiff University
stephen.cutler -at- hotmail.com

Introduction
Formulaic sequences are thought to be a significant feature of native speaker language, but research (e.g. Paquot and Granger, 2012) suggests that intermediate and advanced L2 learners do not use them to nearly the same extent as native speakers. Most such research however has tended to focus on sequences that are considered to be formulaic ‘in the language’ (such as idioms and high frequency multiword units). Wray (2008) makes the distinction between such externally-defined sequences and those which may be psycholinguistic units in the lexicon of the individual speaker. Taking the latter speaker-internal approach to formulaicity, recent work by Cordier (2013) suggests that formulaic sequences may be more prevalent in the speech of advanced L2 speakers than previously thought. She defined psycholinguistic formulaic sequences to be multiword units that present a processing advantage to the individual speaker – either because they are stored holistically or because they are processed automatically as a unit. She found that on average 27% of the speech used by a group of advanced L2 speakers of French consisted of such formulaic sequences.

The current study uses a similar definition and identification process to investigate the use of psycholinguistic formulaic sequences in the speech of intermediate/advanced Japanese Speakers of English (JSE). Its aim is to appraise the effectiveness of this method of identification and to explore the practical and theoretical issues associated with investigating formulaicity in the speech of L2 speakers.

1 Identification of formulaic sequences
The identification process adopted in the study starts with the idea that any sequence stored or processed holistically should be delivered fluently in speech. The first (necessary but not sufficient) criterion for formulaicity therefore was that a candidate sequence should not contain any disfluency (here defined as pauses >0.25s, hesitation markers, repetitions and reformulations). Since not all fluent runs are necessarily formulaic, a further necessary criterion (applied after the first) was that the sequence should show some sign of unity. The diagnostic criteria applied here were: grammatical or semantic irregularity; a holistic form/meaning; and/or, the likelihood of having been received as input (or used frequently) in the form observed. An additional criterion to strengthen he case for formulaicity was that of multiple usage by the participant.

2 Method and participants
Eight intermediate and advanced adult Japanese Speakers of English (JSE), with varying degrees of overseas English experience, were recruited. To provide a point of comparison, two native speakers of English also took part. All participants performed the same two speaking tasks: a structured interview of around 4 minutes about the participant’s work, and the narration of a story based on a sequence of pictures. The tasks were recorded and transcribed and all signs of relevant disfluency in the speech stream were noted. On the basis of these, each sample text was then segmented into fluent ‘runs’. Sequences within the runs therefore satisfied the first criterion (fluency). Formulaic sequences were then identified on the basis of the diagnostic criteria.

Measures of formulaicity were calculated, including FS% (the percentage of syllables that were part of a formulaic sequence) and ANR (the average number of syllables within a formulaic sequence per run). To aid analysis, several temporal fluency measures were also calculated for each task performance, including SR (Speech Rate, in syllables spoken per minute) and MLR (the mean length, in syllables, of each fluent run).

3 Results
Overall, 4798 words (6340 syllables) were spoken by the 8 participants over the two tasks and 663 formulaic sequences were identified (around 20% of which were repetitions). These contained 1685 words (2285 syllables). Types of sequences identified included: Referential (e.g. verb phrases – have to deal with, noun phrases – book stores, time complements – last year); Meta-discursive (e.g. fillers – you know, hedges – some kind of); and ‘Sentence builders’ (e.g. I think __, I’m not good at __). A summary of the quantitative measures
of formulaic sequence usage and fluency for each participant (using pseudonyms) are given in Table 1.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Sex/Age</th>
<th>TOEIC</th>
<th>FS %</th>
<th>ANR</th>
<th>SR (syll/min)</th>
<th>MLR (syll)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Junko</td>
<td>F-40+</td>
<td>650</td>
<td>30.9%</td>
<td>0.81</td>
<td>70.9</td>
<td>2.54</td>
</tr>
<tr>
<td>Eri</td>
<td>F-50+</td>
<td>735</td>
<td>29.6%</td>
<td>0.84</td>
<td>83.6</td>
<td>2.82</td>
</tr>
<tr>
<td>Wataru</td>
<td>M-40+</td>
<td>-</td>
<td>40.3%</td>
<td>1.44</td>
<td>97.0</td>
<td>3.50</td>
</tr>
<tr>
<td>Sachi</td>
<td>F-40+</td>
<td>865</td>
<td>36.0%</td>
<td>1.78</td>
<td>115.7</td>
<td>4.96</td>
</tr>
<tr>
<td>Kanae</td>
<td>F-30+</td>
<td>940</td>
<td>35.6%</td>
<td>1.58</td>
<td>123.4</td>
<td>4.44</td>
</tr>
<tr>
<td>Mami</td>
<td>F-30+</td>
<td>-</td>
<td>33.8%</td>
<td>1.81</td>
<td>127.3</td>
<td>5.34</td>
</tr>
<tr>
<td>Yayoi</td>
<td>F-40+</td>
<td>975</td>
<td>31.9%</td>
<td>2.21</td>
<td>148.3</td>
<td>6.80</td>
</tr>
<tr>
<td>Yoko</td>
<td>F-40+</td>
<td>960</td>
<td>38.5%</td>
<td>2.63</td>
<td>175.9</td>
<td>6.85</td>
</tr>
</tbody>
</table>

Table 1: Summary Fluency and Formulaicity of Participants over Both Tasks

Overall, mean FS% was 34.6% and mean ANR was 1.64. Comparing the two different tasks, average FS% for the interview task (38.1%) was higher than for the picture narration task (31.0%) with similar task differences being reflected in the other measures too. As expected, average formulaicity (FS% = 47.3%, ANR = 4.27) and fluency (SR = 189) for the two native speakers were both higher than for the JSE.

4 Discussion points

The FS% figures suggest that psycholinguistic formulaic sequences are a significant part (30-40%) of the speech of these JSEs. Indeed these figures are higher than Cordier found for her advanced French students. This may be due in part to the 0.25s pause cut-off used here (rather than the 0.2s used by Cordier) and also possibly to the language difference. In addition, although the same process was used, there was still potential for differences in interpretation - both in applying some of the diagnostic criteria and in deciding what should be included as part of the formulaic sequences (e.g. all of *at the bottom of the sea* or just the frame).

The results show a different pattern in the two formulaicity measures across the participants. For ANR, the average number of formulaic syllables per fluent run seem to increase in line with fluency. Indeed the two participants with considerable experience of living overseas (Yayoi and Yoko) had much higher ANR figures, (although not yet as high as the native speakers). For FS% however, there was not such a clear pattern. For example, Wataru had a high ratio of formulaic syllables in his speech even though he spoke quite hesitantly, possibly indicating the use of lengthy pause time between runs to plan and retrieve formulaic sequences. This highlights the complexity of the relationship between fluency and use of formulaic sequences and suggests that drawing on a combination of measures (including average pause time) is necessary to better understand the interplay between holistic storage, retrieval and automatic processing.

The study also provided examples illustrating the potentially dynamic and context-based nature of formulaicity in the individual speaker (as suggested by Ellis 2012). For example, Junko in her interview appeared to at first construct the name of her department in English (*PR unit*) but then use it regularly in a formulaic way. Also highlighted was the slightly arbitrary nature of using the word as a defining feature for identification (e.g. *test takers* was included but not *examinees*).

5 Conclusion

As a first study to use these particular identification criteria on L2 English speakers, the results provide a useful insight into the prevalence of psycholinguistic formulaic sequences in intermediate/advanced JSE and the practical challenges in identifying these. It confirms the sensitivity of the results to task, identification method and measures of formulaicity used. The use of disfluency provided a clearly quantifiable starting point for identification, but further refinement of the diagnostic criteria may be required.

References


Exploring the Effects of Learner-generated Multimedia Annotations on Vocabulary Acquisition

Di Zou  
The Hong Kong Polytechnic University  
dizou daisy -at- gmail.com

Haoran Xie  
The Education University of Hong Kong  
h r x i e 2 -at- gmail.com

Fu Lee Wang  
The Caritas Institute of Higher Education  
pwang -at- cihe.edu.hk

1 Introduction

Language learners are often encouraged to keep notes of unfamiliar words that they encounter while reading, and go over them periodically to obtain a strong and durable memory trace (Hulstijn, 2001). As one of the most widely adopted word learning strategies, such practice of creating personalized word lists, however, does not normally involve multimodality. Word lists are normally comprised of only the target words and simplified definitions, with sample phrases or sentences sometimes included. Few learners try to, or are advised to, integrate multimedia (e.g., images, videos, animation, audio, etc.) into their word lists, even though the effects of multimedia annotations on vocabulary acquisition have been widely acknowledged (Chun & Plass, 1996; Nation, 2001; Abraham, 2008).

However, the scopes of these studies are limited to the annotations that are provided to learners, while the field of learner-generated annotations has rarely been explored. This research, therefore, aims to investigate the effectiveness of four types of learner-generated annotations in promoting the initial learning and retention of ten target words. So far as we know, this is the first attempt to examine the employment of multimedia in creating word lists.

2 Methods

One hundred and fifty-one non-English major freshmen, who were intermediate language learners, participated in the study. They were randomly divided into four groups, each of which did one of the following four tasks. Task 1: Create a word list with textual annotations; Task 2: Create a word list with textual plus audio annotations; Task 3: Create a word list with textual plus Pictual annotations; and Task 4: Create a word list with textual, audio, plus Pictual annotations. As shown in Table 1, the first type of annotations includes definitions of the target words and original example sentences generated by the participants; the second type includes definitions, original sentences, and audio recordings of the words; the third type includes definitions, original sentences and images that represent meanings of the words; and the fourth type includes definitions, original sentences, audio recordings and images. The participants were encouraged to use online dictionaries and sources to generate multimedia annotations for the target words.

<table>
<thead>
<tr>
<th>Type</th>
<th>Text (definitions of the target words + original example sentences)</th>
<th>Audio recordings of the pronunciations of the target words</th>
<th>Images that depict the meanings of the target words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 1</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type 2</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Type 3</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Type 4</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Table 1. Four Types of Annotations.

The ten target words were burglary, dash, detest, inflammation, rake, scribble, shatter, shiver, tumble, and wrath. The results of the pre-test, which asked the participants to write down the meanings of these target words, showed that they had almost zero knowledge of these words.

Both immediate and delayed post-tests were conducted. The participants were asked to recall the meanings of the target words and generate original sentences. Following Zou’s (2016) scoring criteria, a meaning was graded zero if it was completely incorrect, a half score if it was semantically acceptable, and a full score if it was completely correct. A sentence was graded zero if the semantic context for the target word was completely inappropriate, a half score if the context was appropriate but the target word was used ungrammatically, and a full score if the sentence was semantically and grammatically correct. Blind scoring was employed.

Six participants from each group were also interviewed to further tap into their learning processes. Open-
ended questions were asked, examples of which are as follows: (1) What did you pay special attention to in the course of task completion? Why? (2) What did you find helpful for your task completion and word learning? Why? (3) For which target words were you able to recall the meaning and usage? What do you think helped you learn these words?

3 Results and discussions

The results confirmed the superiority of multimedia annotations over simple text annotations, as the scores of the participants who did the task of creating a word list with textual annotations were much lower than the scores of the participants who did the other three tasks of creating word lists with multimedia annotations in the immediate and delayed posttests. Among the three tasks which involved multimedia annotations, the forth task was significantly more effective than the other two according to the one-way ANOVA test, while no significant difference was found between task 2 and task 3.

<table>
<thead>
<tr>
<th>Type</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 1: text</td>
<td>31</td>
<td>.12</td>
<td>.36</td>
<td>12.15</td>
<td>4.67</td>
</tr>
<tr>
<td>Type 2: text + audio recordings</td>
<td>32</td>
<td>.13</td>
<td>.38</td>
<td>14.09</td>
<td>4.22</td>
</tr>
<tr>
<td>Type 3: text + images</td>
<td>32</td>
<td>.10</td>
<td>.35</td>
<td>14.33</td>
<td>4.59</td>
</tr>
<tr>
<td>Type 4: text + audio recordings + images</td>
<td>32</td>
<td>.11</td>
<td>.31</td>
<td>15.89</td>
<td>4.48</td>
</tr>
</tbody>
</table>

Table 2. Descriptive statistics of the participants’ scores in the pretest, and immediate and delayed post-tests.

Such results, with supporting evidence from the interview transcripts, can be explained from the perspectives of the notions of depth-of-processing and elaboration (Craik & Lockhart, 1972; Craik & Tulving, 1975). As the generation of word lists with multimedia annotations involved using and re-using various knowledge aspects of the target words, greater depth of processing and more elaboration were induced. Also, greater chances for the participants to build up networks of the target words were created, so a greater complexity of connections was available, assisting the participants in processing the target words from diverse dimensions and making use of more complex connections to build up networks regarding them. The tasks involving the generation of word lists with multimedia annotations were, therefore, more effective in promoting word learning than the task of generating a word list with simple textual annotations.

4 Acknowledgements

The work described in this paper was fully supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (UGC/FDS11/E06/14).

References


Profiling Lexical Diversity in College-Level Writing

Melanie C. González  
Salem State University  
mgonzalez-at-salemstate.edu

Millene Alves  
Salem State University  
m_alves4-at-salemstate.edu

1 Introduction

Over the years, vocabulary-related studies in multilingual (ML) writing scholarship consistently cite the positive influence of lexical diversity on college-level writing quality (Crossley & McNamara, 2009; Engber, 1995; Friginal, Li, & Weigle, 2014; Johnson, Acevedo, & Mercado, 2013). These findings have resulted in the inclusion of criteria such as “the sophisticated use of vocabulary” and “a variety and range of vocabulary” on holistic rubrics to assess L2 writing proficiency and, subsequently, college writing readiness (ETS, 2014). However, little is known about what vocabulary items actually contribute to the lexical diversity of a text. As a result, writing instructors aiming to help ML writers to improve their productive vocabulary often must rely on their intuition to determine which words to target for instructional activities.

2 Literature review

A first step to profiling lexical diversity is to examine the lexical frequency of the words writers produce. Lexical frequency aids researchers and teachers alike to quantify learners’ vocabulary size and profile the lexical sophistication of written texts. Studies have indicated that as ML writers grow their language proficiency, they begin to produce more words from the lower frequency bands in their writings (Laufer & Ravenhorst-Kalovski, 2010; Meara & Bell, 2001).

In terms of the intersection between lexical frequency and lexical diversity, there is an assumption within the field that a text that produces more low-frequency words would earn a higher score of lexical diversity than a similar text that uses more high-frequency terms (Daller, Van Hout, & Treffers-Daller, 2003). However, there is scarce, but preliminary evidence that lexical frequency does not always correlate well to lexical diversity. Laufer (1994) found no significant correlation between learners’ increase in the use of lower-frequency terms and their lexical diversity scores. Johnson, Mercado, and Acevedo’s (2013) study, however, indicated that the use of lower-frequency words, albeit from the 4,000-5,000 word frequency bands, did facilitate writing score. This incongruity suggests that learners do not necessarily need to draw from the low-frequency bands in order to achieve the lexical diversity necessary for proficient L2 writing.

To add to these previous findings, the examination of college writers’ written lexical frequency profiles and how they intersect with lexical diversity has the potential to lend insight into possible gaps in their productive vocabulary knowledge. Furthermore, looking at what frequency bands are represented in ML texts as compared to what their monolingual English-speaking (MES) peers can help to answer the question of what words to use instructional time on.

Given lack of studies profiling lexical diversity in academic compositions, this study took a first step in hopes of filling this gap via the following two research questions: (1) How do the lexical frequency profiles of advanced ML writers’ college-level compositions compare to those of their MES peers? (2) What frequency level(s) is a significant contributor to lexical diversity in college-level compositions?

3 Methods

A small corpus of 119 narrative, final draft essays produced by 54 ML and 65 MES writers enrolled in first-year composition courses was gathered. The mean length of the texts in the corpus was 618.78 total words.

The study utilized Lextutor’s BNC-COCA25 lexical frequency profiler and the MTLD index of lexical diversity available in the Coh-Metrix 3.0 textual profiler to quantify the variables. The number of high-, mid-, and low-frequency words were then tallied to complete the lexical frequency variable. An ANOVA was run to determine if there were significant differences between the lexical frequency and diversity profiles of ML and MES texts. A multiple regression then determined which category of lexical frequency significantly facilitated greater lexical diversity within the corpus.

4 Results

For research question one, ML writers significantly produced more high-frequency words than their MES peers ($F_{2,117}=54.13$, $p<.00$). By contrast, MES writers used more mid-frequency words in their writing
(F2,117=15.12, p<.00). MES writers also exhibited greater diversity in their writing than their ML peers (F2,117=5.06, p<.05). There were no significant differences between the groups’ use of low-frequency terms (see Table 1).

<table>
<thead>
<tr>
<th>Designation</th>
<th>M</th>
</tr>
</thead>
</table>
| High-frequency  | MES   | 468.55
|                 | ML    | 487.17*|
| Mid-frequency   | MES   | 14.45*|
|                 | ML    | 7.09   |
| Low-frequency   | MES   | 3.20   |
|                 | ML    | 2.40   |
| Lexical diversity | MES | 79.95* |
|                 | ML    | 69.54  |

Table 1. Descriptive results. * Significant at p<.05

The multiple regression analysis for research question two revealed that lexical frequency explained about 27% of the variation in lexical diversity score (MTLD; (F2,117= 4.75, p < .05) and that mid-frequency vocabulary was the only significant predictor of lexical diversity (beta = .93, p < .05). Therefore, for every .93 increase in lexical diversity score, there is an uptick in the likelihood of a mid-frequency vocabulary term being utilized.

4 Discussion, pedagogical implications, and limitations

The outcome that ML writers use more high-frequency words than their MES peers align with the results of precedent research (Meara & Bell, 2001). However, the outcome that MES writers made more use of mid-frequency vocabulary terms than ML writers and that these are the terms that significantly contributed to lexical diversity is a noteworthy finding and points towards a possible gap in ML writers’ productive lexicon. The fact that each groups’ use of low-frequency terms did not significantly differ appears to validate this finding. Instructural applications of the results of this study include teaching writers how to effectively find and select synonyms and hypernyms for high-frequency terms to diversity their texts’ vocabulary.

Outside of population generalizability as a general consequence of research, the fundamental limitations of the methodology of this study relate to the challenges that text length, task topic, and writing genre pose for any study of lexical diversity. The writing samples were all narrative, composed on the same topic, and were relatively short (between 500 and 900 words in length). The present study chose to constrain the corpus in this manner in order to control for these variables and gain a clearer picture of differences between the frequency and diversity profiles. However, with a wider range of genre, topic, and word count available to participants, results may shift.

References


Short-term Study Abroad and the Lex30: A Replication Study

Tom Caton
Nakamura Junior College  Cardiff University  PhD student
tomcaton901 -at- hotmail.com

1 Background

This study looks at the findings of a replication experiment carried out in early 2016 as part of ongoing PhD research investigating changes in vocabulary knowledge during short-term Study Abroad (SA).

Short-term SA programs can give opportunities for language use and can improve participants’ overall language proficiency. However, there has been little research on how such programs can affect different kinds of language knowledge, especially lexical knowledge. With short programs measuring changes in language ability have been problematical as established tests, such as TOEIC, often lack sufficient sensitivity. Forms of assessment which focus on a particular aspect of language and knowledge are perhaps more likely to reveal subtle changes taking place over a short duration than more general testing methods. Assessment techniques, including measures of both receptive and productive vocabulary size (Beglar and Nation 2014) might offer a more accurate way to measure levels of vocabulary knowledge than existing mass market tests.

2 Lex30 longitudinal studies

Fitzpatrick and Clenton (2010) looked at the reliability and construct validity of the Lex30, a test of productive vocabulary knowledge, in their paper, ‘The challenge of validation: Assessing the performance of a test of productive vocabulary’. They presented a strong argument for the validity of the test as a research tool by comparing it with similar tests and looking at changes in test performance over time. This study attempts to replicate part of their research in which the authors investigate the Lex30’s ability to detect improvement in learners’ vocabulary knowledge. In the original paper a longitudinal study of 40 Japanese L1 participants attending a six-week English language improvement course is described along with the results of pre and post Lex30 testing. The authors found that that there was a significant increase in the number of infrequent words produced suggesting that it was possible to measure an increase in vocabulary knowledge in weeks rather than years.

This study also carries out a replication of part of Fitzpatrick’s (2003) PhD thesis, ‘Eliciting and measuring productive vocabulary using word association techniques and frequency bands,’ which further builds on work done with Paul Meara on the introduction of the Lex30 productive vocabulary test (Meara and Fitzpatrick 2000). In her thesis she explores how productive vocabulary can be elicited and measured by using word association techniques and word frequency lists and demonstrates how the Lex30 can collect useful data in an efficient way, thereby avoiding many of the problems that had plagued previous attempts at designing similar tests (ie; Laufer and Nation 1995, 1999). For our replication study and for the purposes of a comparison with the research carried out by Fitzpatrick and Clenton (2010) a second longitudinal experiment, carried out with a group of 19 Japanese students studying in Britain over a 4-week period, is described. Using an alternative scoring system measuring the number of infrequent words as a percentage of the total number of words produced by an individual instead of using a simple infrequent word total score, it was found that the difference in the mean Lex30 scores between test time one and two was not significant, remaining relatively stable over the 4-week period. These results suggest that either the number of infrequent words in the subjects’ productive lexicons has not increased over the study period or perhaps that the Lex30 test is not sensitive or sophisticated enough to pick up any increases within such a short period.

3 The replication study

This experiment attempts to replicate both these previous studies by looking at 38 Japanese L1 university students participating in a 17-day short-term SA programme. A paper and pen version of the Lex30 test was administered before and just after the programme with a time limit of 15 minutes given to complete the test on each occasion. After completion individual responses were processed according to protocol laid down in Meara and Fitzpatrick (2000) and lemmatized so that inflectional suffixes (plural forms, past tenses, comparatives) and frequent regular derivational affixes (-able, -ly) were counted as base forms of these words. Two scores were calculated for each participant, corresponding to the scoring systems used in the original papers: firstly, a raw score (which simply totaled the number of infrequent words produced that did
not appear on JACET 8000’s wordlist of 1000 most frequently occurring words) was awarded. Secondly, a score which expressed this raw number of infrequent words as a percentage of the total number of responses produced by each individual, was calculated.

4 Results

The replication experiment reproduced the results of the two previous longitudinal studies with mixed success. Comparisons with Clenton and Fitzpatrick (2010) were encouraging. Using a similar number of participants, similar test administration and protocols and most importantly the same Lex30 scoring system using a raw count of infrequently occurring words, the results that were obtained were broadly equivalent. In both studies the difference between pre and post test means was significant with the t value being slightly higher with the replication. The correlation for both test scores was close with figures of 0.809 (p < 0.0001) for Fitzpatrick and Clenton’s study and 0.747 (p < 0.0001) for the replication. With Fitzpatrick’s 2003 study comparisons were more complex as the raw Lex30 test scores had to be converted into percentage forms before they could be properly analyzed. In both studies the number of infrequent words produced by most participants increased between the two test times. However, the total number of responses increased over the same period at a much greater rate. The result of this was that the percentage of infrequent words as a proportion of the total number of words produced by each participant actually decreased suggesting that no significant improvement in language proficiency had taken place over the study period.

In conclusion there seems to be some doubt about what kind of scoring system is appropriate for use with the Lex30 test. Are simple Lex30 infrequent word scores sufficient to make a judgment of a subject’s improvement in productive vocabulary knowledge over the course of a short-term SA programme? In addition what account should be taken of the influence of the total size of each individual corpus that is produced? Further exploration into how data from productive vocabulary tests such as the Lex30 can be processed will perhaps help formulate new and better-balanced marking schemes for the future.

References


A Time for Multi-State Models of Vocabulary Acquisition?

Rob Waring
Notre Dame Seishin University
Waring.rob -at- gmail.com

1 Introduction
This paper develops a paradigm for assessing huge amounts of words, even whole lexicons, to see how they develop over time by looking at how words change their state of knowledge.

2 Problems with classical vocabulary testing
In SVLA, the standard classical paradigm tends to assess a person's lexical knowledge with tests, either receptive or productive, and from insensitive to sensitive degrees and to assess breadth and/or depth, among others. Building a picture of the development of a whole lexicon in this way is problematic as each test is assessing discrete and independent aspects of lexical knowledge and we need to make arbitrary decisions about how these discrete scores fit together. Moreover, this method does not allow us to test large amounts of vocabulary as each test is time intensive as well as very difficult to make.

3 Vocabulary Knowledge Scales
Vocabulary knowledge scales (e.g., Wesche & Paribakht, 1996) attempt to resolve some of these problems with a self-reporting tool, such as the VKS. However, their scale also presents problems, not least of which are the multiple knowledge sources being assessed. Other problems include the arbitrary scale of development which is both linear and non-linear, the very suspicious scoring system, and lack of clarity over what a score of 2.5, for example, might mean. Nevertheless, scales of these types can provide a platform for assessing vocabulary in new ways, especially if the scales are developed using a uniform aspect of knowledge such as the following.

0 I do not understand (the meaning of) this word
1 I understand (the meaning of) this word a little
2 I understand (the meaning of) this word quite well
3 I understand (the meaning of) this word very well

4 Multi-state models of vocabulary acquisition
We can re-conceptualize the points on a VKS not as being along a mathematical ordinal scale of growth, but as nominal data and of knowledge being in various 'states' which exist independently of all others, but nevertheless interact, as seen in the example in Figure 1.

Figure 1. An example of a metacognitive multi-state model with 5 states.

Multi-state models of vocabulary acquisition require subjects to state their lexical knowledge on two or more pre-determined scales, each of a single knowledge type, like the one above. If we have a scale of 0 to 3 for "understanding" and another for "use", a subject could tell us that a certain word might be (2,1) (i.e., a word's "meaning is quite well understood" (level 2), but it "can't be used easily, e.g., in a sentence" (level
1)), while other words might be (3,1) or (2,0). From this, we can build a matrix of knowledge for the whole set of words being tested which show how words move from state to state over time. In Figure 2, the word $a$ stays in the same state (0,0) at $t_1$ and $t_2$, but word $e$ has moved from (2,0) to (3,3).

![Figure 2. Example matrices showing the movement of words from $t_1$ to $t_2$.](image)

With some very clever math, it would be possible to map vocabulary knowledge of many types in multiple dimensions over time. Figure 3 shows how we can graphically represent these data along two axes (e.g., one for ‘understanding’ and one for ‘use’) allowing us to build scatterplot representations that visually represent how a lexicon changes and fluctuates over time while allowing us to visualize vocabulary knowledge in new and exciting ways.

![Figure 3: A sample representation of how a lexicon might be viewed using multi-state models](image)

### 5 The flexibility of multi-state models

While additional data may need to be collected to verify the self-reporting is accurate, multi-state models of vocabulary development still have several advantages. 1) As each scale is simple, subjects (individuals or groups) can respond to hundreds of items in short sessions, allowing us to collect huge amounts of data over time, especially if collected digitally. 2) The same design can be used with any type of declarative vocabulary knowledge by varying the scale descriptors, e.g., spelling or collocational awareness. 3) Researchers can vary which words are assessed—individual words of any frequency level, or phrases or even different meaning senses, derivatives, and so on.

### References

i-lex: An Improved Method of Assessing L2 Learner Ability to See Connections between Words?

Ian Munby
Hokkai Gakuen University
ianm-at-jin.hokkai-s-u.ac.jp

1 Background

Knowing a word’s associations is considered an aspect of word knowledge (Richards, 1976; Nation, 2001). It follows that L2 learner ability to see connections between words may improve with gains in vocabulary knowledge and proficiency (Meara, 1983). Meara (1996) also suggests that WATs (word association tests) may measure the degree of organization of the L2 learner lexicom. Existing WATs seek to measure learner ability to make associations between words using a variety of formats. These can be broadly divided into receptive WATs and productive WATs, yet both types are susceptible to design weaknesses which may undermine results. Receptive WATs (e.g. The Word Associates Test, Read, 1993) may allow testees to guess word associations among sets through elimination processes without knowing why the TWs (target words) are associated. With productive WATs, where the task is to supply free associations to TWs, Schmitt (1998) observes that scoring responses with lists of word association norms is problematic because results depend on the cue words and norms lists used.

2 Aims

The aim of this study is to circumvent these weaknesses with a new WAT inspired by Meara (1994), who mused upon possible uses of a Spanish word association norms list. He suggests presenting learners with the three most common associates of a cue word and asking them to supply the missing word. The instructions of this new WAT, named i-lex, are as follows: What word is associated with the following sets of 3 words? Example: drink, red, glass > w ______ [4]. The first letter, “w”, is given for you and the word has 4 letters [4]. The answer is wine. From this aim, the following three research questions were formulated:

RQ1 Does i-lex distinguish between native and non-native speakers?
RQ2 Is there a significant, positive correlation between learner i-lex scores & vocabulary test scores?
RQ3 Does i-lex satisfy initial, basic scrutiny for validity?

3 Method

The three cue words (CWs) in each test item are chosen from the five most common associates to each TW on the Edinburgh Associative Thesaurus, or EAT (Kiss et al, 1973). Additionally, to enable participation by elementary level testees, all the TWs and CWs are selected from the BNC 1K range. For details of the remaining design criteria, see Munby (2013). A total of three trial versions of i-lex were conducted in order to: (i) estimate appropriate time limits for a 50-item test, (ii) identify and remove problematic items, such as items which no subjects could answer, and (iii) sort the items in order of difficulty from the easiest to the most difficult. The experimental group comprised 98 Japanese learners of English at four different universities who ranged in level from elementary to intermediate. There were 25 native speakers of English in the control group. After completing i-lex (25 minutes), the non-native group completed a kanji translation test adapted from Webb (2008) in 20 minutes. The task is to write English translations for a series of 160 single words of varying levels of word frequency written in L2 (Japanese).

4 Results

With regard to RQ1, on average, native speakers outperform non-natives (see Table 1). A one-tailed unpaired t-test confirms that the difference between the means of the two groups is significant at \( t = 11.153, \) \( p < 0.0001 \)
Table 1. A Comparison of the Means and Standard Deviations of all Test Scores for the Native Group (n=25) and the Non-Native Group (n=98).

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Hi</th>
<th>Low</th>
<th>Max</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>i-lex (natives)</td>
<td>41.76</td>
<td>4.28</td>
<td>49</td>
<td>33</td>
<td>50</td>
<td>0.960</td>
</tr>
<tr>
<td>i-lex (non-natives)</td>
<td>25.00</td>
<td>7.23</td>
<td>42</td>
<td>6</td>
<td>50</td>
<td>0.990</td>
</tr>
<tr>
<td>Translation test</td>
<td>96.26</td>
<td>20.2</td>
<td>145</td>
<td>44</td>
<td>160</td>
<td>0.990</td>
</tr>
</tbody>
</table>

With regard to RQ2, Pearson correlations among i-lex scores and translation test scores were .729 (1-sided p value, significant at p < 0.01). This indicates that the ability of this group of subjects to see links between highly frequent English words is related to their vocabulary knowledge. With regard to RQ3, there was an issue with one item (47 boy, face, girl > baby) wherein 19 non-natives provided the non-scoring, or non-target response body, compared with only nine who supplied the intended TW. Since five native speakers also provided body it represented a challenge to the validity of the test.

5 i-lex version 2

Due to this flaw, item # 47 was replaced with a new item and i-lex was relaunched in a second study (i-lex version 2) with new participants and an additional proficiency countermeasure: The New Vocabulary Levels Test, Levels 1-5, (McLean & Kramer, 2015). Data collection is currently in progress and results will be reported at the conference.

6 Conclusion

The results of the first study (Munby, 2013) indicate that native speakers generally obtain higher i-lex scores than learners. Further, learner i-lex scores correlate positively and significantly with vocabulary test scores, suggesting that, with gains in vocabulary knowledge or proficiency, the ability of learners of English to recognize and produce word associations generally improves.

References


Tracking Vocabulary Growth Over Time in an International School Context

Averil Coxhead
Victoria University of Wellington
Averil.coxhead @ vuw.ac.nz

TJ Boutorwick
Victoria University of Wellington
Tj.boutorwick @ vuw.ac.nz

1 Introduction

Recent studies in the Danish (Henriksen & Danelund, forthcoming) and Taiwanese (Webb & Chang, 2012) contexts with teenage and older second language learners suggest that low vocabulary knowledge and growth are a major concern for research and pedagogy. The present study focuses on the vocabulary growth of secondary school students from 2009-2015 in an international school in Germany, using Nation’s (1993; Schmitt, Schmitt & Clapham, 2001) Vocabulary Levels Test (VLT). The study is part of a larger project on vocabulary in an international school context in Germany, which includes teacher talk data and collections of some student writing.

In the international school classrooms, students come from a wide variety of language backgrounds, including native speakers of German who enrol at the school for an English-medium education. At this school, English was the medium of instruction until late 2015 when it changed to a bilingual education model. Each year, Grade Six (10-11 years old) students are tested using the VLT and the results are used to determine whether and what level of support a student might need to help them deal with their reading texts and subject studies over the coming year. Non-native speakers of English who score low on the VLT are offered English as an Additional Language support to help improve their proficiency.

In this study, a total of 471 students took the VLT, including 306 non-native speakers of English (NNS), 103 non-native speakers who needed EAL support based on their VLT scores (NNSEAL), and 59 native speakers of English (NS). These students sat a monolingual version of the VLT at the end of the school year from 2009 to 2015, starting at Grade Six and continuing each year through to 2015. This means that a student entering Grade Six in 2010 took the test in 2010, again in 2011, again in 2012 and so on. This data enables tracking of vocabulary development over time, across the levels of the VLT (2000, 3000, 5000, academic and 10,000) for the students in the study.

In this abstract, we present the data from these three groups over each section of the VLT for students who took the test for the first time in 2011 and have repeated it yearly through to 2015. As far as possible, we have tried to maintain complete data sets, but this population of students is fairly mobile as parents and their children move schools, cities, countries, and even continents perhaps several times in the course of a secondary school education. The N-size for the three groups from the NNSEAL, NNS and NS groups are presented in Table 1.

<table>
<thead>
<tr>
<th>Year</th>
<th>NNSEAL</th>
<th>NNS</th>
<th>NS</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>5</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>2012</td>
<td>10</td>
<td>27</td>
<td>8</td>
</tr>
<tr>
<td>2013</td>
<td>13</td>
<td>31</td>
<td>11</td>
</tr>
<tr>
<td>2014</td>
<td>13</td>
<td>36</td>
<td>12</td>
</tr>
<tr>
<td>2015</td>
<td>10</td>
<td>37</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 1. Number of Participants from Each Group in Each Year of Testing from 2011-2015.
Note that there are more NNS at each time of testing than in either of the other groups.

2 Results and discussion

Mastery in the VLT is set at 26/30 items being correct in the test (Schmitt, Schmitt & Clapham, 2001). Our results showed that the NS group attained mastery at the 2,000 and 3,000 levels in the first year of testing (2011), and that their scores for the 5,000 and Academic levels of the test attained mastery in the third year of taking the test (2013). The NNS cohort also attained mastery of the 2,000 level from the first year of testing, and this group also attained mastery of the 3,000 level in 2013, the Academic level in 2015 and mastery of the 5,000 level (26.9) in 2015. The NNSEAL group attained mastery of the 2,000 in the fourth year of testing (27.9) in 2014. This group attained mastery of the 5,000 and Academic in 2015 (both scores were 26.9). The 10,000 level section of the test is not reported here, because it some scores were not available for the 2011 and 2015 data sets.

The group which showed the greatest gain in VLT scores over the five years in the study was the NNSEAL group. This result is perhaps unsurprising, given that this group had the most to learn in the first place. Table
2 shows the gains in scores at each level of the test, for each of the groups.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Group</th>
<th>2,000 VLT level</th>
<th>3,000 VLT level</th>
<th>5,000 VLT level</th>
<th>10,000 VLT level</th>
<th>Academic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-native spekers EAL</td>
<td>18.4</td>
<td>20.2</td>
<td>18.9</td>
<td>4.4</td>
<td>21.4</td>
<td></td>
</tr>
<tr>
<td>Non-native spekers</td>
<td>3.3</td>
<td>8.8</td>
<td>13.1</td>
<td>5.1</td>
<td>12.6</td>
<td></td>
</tr>
<tr>
<td>Native speakers</td>
<td>0.3</td>
<td>2.5</td>
<td>5.0</td>
<td>5.4</td>
<td>7.7</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Word Gains in the VLT for the Three Groups from 2011-2015.

These findings suggest that the academic secondary school environment supports learning of vocabulary at all levels of the VLT, but particularly at the 5,000 and Academic levels for the NNS and NNSEAL groups. Clearly, more data and a wider range of data need to be gathered to find out more about vocabulary in these contexts.

3 Acknowledgements

We wish to acknowledge funding from the Faculty of Humanities and Social Sciences, Victoria University of Wellington, New Zealand, co-researcher who is anonymous, and the participants in the International School.

References


Effects of Within-session Repeated Retrieval on Second Language Vocabulary Learning

Tatsuya Nakata
Kansai University
nakata-at-kansai-u.ac.jp

1 Literature review

Previous studies demonstrate that retrieval, or the process of accessing information about a second language (L2) word from memory, is beneficial for L2 vocabulary acquisition (e.g., Barcroft, 2007). Provided that retrieval increases learning, some may wonder whether increasing retrieval frequency within one learning session (hereafter referred to as within-session repeated retrieval) increases L2 lexical acquisition. There exist conflicting views regarding whether within-session repeated retrieval facilitates learning. On one hand, usage-based models of language acquisition (e.g., Tomasello, 2003), according to which frequency is one of the major factors affecting acquisition, suggest that increasing retrieval frequency in one learning session increases learning. The spacing effect and retrieval effort hypothesis, in contrast, suggest that within-session repeated retrieval may not be beneficial.

First, according to the spacing effect (e.g., Bahrick & Phelps, 1987; Nakata, 2015; Nakata & Webb, in press), the positive effects of within-session repeated retrieval are short-lived. This is supported by L2 vocabulary research on within-session repeated input (van Zeeland & Schmitt, 2013) as well as non-L2 vocabulary research on within-session repeated retrieval (Rohrer et al., 2005). Second, according to the retrieval effort hypothesis (Pyc & Rawson, 2009), difficult and effortful retrievals enhance learning more than easy and effortless retrievals. When learners practice retrieval repeatedly within one learning session, retrieval will be progressively easier. As a result, the retrieval effort hypothesis suggests that the benefits of retrievals decrease as a function of retrieval frequency. This predicts that when learners practice retrieval repeatedly within one learning session, most learning perhaps occurs in the first few retrievals, and increasing within-session retrieval frequency beyond a certain level does not provide a benefit that justifies its cost (i.e., increased study time). Hence, the retrieval effort hypothesis suggests that within-session repeated retrieval may not necessarily be efficient.

2 Research questions and hypotheses

With the conflicting views regarding the effects of within-session repeated retrieval in mind, the present study investigated the effects of within-session repeated retrieval on L2 vocabulary acquisition. The following two research questions were examined:

RQ1: Does increasing within-session retrieval frequency increase effectiveness for L2 vocabulary acquisition?

RQ2: Does increasing within-session retrieval frequency increase efficiency for L2 vocabulary acquisition?

Based on the review of literature, the following two hypotheses were formulated:

Hypothesis 1: Within-session repeated retrieval is more effective than fewer retrievals in the short term, but not in the long term.

Hypothesis 2: Within-session repeated retrieval is less efficient than fewer retrievals in both the short and long term.

3 Method

The participants were 98 first-year Japanese students at a technical college in Japan who were randomly assigned to one of the four retrieval frequency groups: Retrieval 1, 3, 5, and 7. The participants studied 16 English-Japanese word pairs (e.g., husk – 皮) using one of the following four retrieval frequency levels (i.e., 1, 3, 5, or 7). Vocabulary learning was measured by productive and receptive recall posttests given at three retention intervals: immediately, one week, and four weeks after the treatment. Responses on the pretest and posttest were scored using the following two procedures: strict and sensitive.

4 Results

The productive and receptive posttest scores were entered into a two-way 4 (retrieval frequency: 1 / 3 / 5 / 7)
The analysis found that regardless of the posttest, scoring system, or posttest timing, 5 and 7 retrievals led to significantly better learning than 1 and 3 retrievals. No significant difference, however, was found between 1 and 3 retrievals on the one hand and between 5 and 7 retrievals on the other. The findings suggest the following order on all posttests: 5 and 7 retrievals > 1 and 3 retrievals.

In order to investigate the efficiency of the four retrieval frequency levels, efficiency scores were calculated by dividing the posttest score by the study time. The efficiency scores were entered into a two-way 4 (retrieval frequency: 1 / 3 / 5 / 7) x 3 (posttest timing: immediate / one-week delayed / four-week delayed) ANOVA. The analysis revealed that 1 retrieval was the most efficient among the four retrieval frequency levels.

5 Discussion
Contrary to Hypothesis 1, the benefits of within-session repeated retrieval were observed at least four weeks after the treatment. The results were possibly attributed to the positive effects of increased frequency outweighing the negative effects of massing practice.

The findings of this study were in line with Hypothesis 2. Theoretically, the findings support the retrieval effort hypothesis, according to which the benefits of repeated retrieval decrease as a function of retrieval frequency. From a pedagogical perspective, the present study suggests that if learners do not have enough time, it may be desirable to practice retrieval only once. At the same time, if efficiency is not an issue, increasing retrieval frequency within one learning session may be preferable.

References


Exploiting a Teacher-Made Wikipedia Corpus for Vocabulary Learning

Marlise Horst
Concordia University
marlise -at- education.concordia.ca

1 Introduction

The idea that corpus analyses provide powerful insights into the lexical forms, meanings and uses that are frequent in a language—and therefore important for learners to know—is no longer new. The idea of putting corpora and the tools of analysis directly into the hands of learners so that they can make lexical discoveries for themselves is also familiar. Indeed, an impressive body of current experimental research attests to the benefits of engaging in such data-driven vocabulary learning activities (Cobb & Boulton, under review). The idea that teachers might also usefully do corpus work has been put forward (e.g. Timmis, 2015), but research on teacher-made corpora and their ideas for applying corpus finding appears to be scarce. My investigation addresses this underexplored area.

Teachers of English in a graduate course at a North American university worked collaboratively to create a 1-million-word corpus of Wikipedia texts and then explored its teaching potential. The investigation of their experience begins with the following process-related questions: How did the teachers decide on the kind of corpus to collect? How was it organized, gathered and compiled? How did they experience using corpus tools? What problems and eventual solutions emerged? The second part of the research considers how the teachers explored the corpus, their findings, and their ideas for using the corpus in the language classroom.

2 Building the corpus

In a discussion where the participants proposed various kinds of corpus they might collect (songs, academic texts, graded readers, etc.), they concluded that a corpus of Wikipedia texts would have considerable pedagogical value. The corpus they eventually created (the Concordia Corpus of Wikipedia or ConCoW) is made up of 12 subcorpora, one for each of Wikipedia’s 12 theme areas or ‘portals’. The portals address the following topic areas: culture, geography, health, history, mathematics, nature, people, philosophy, reference, religion, society, and technology. Each participant was assigned one of these portals and was responsible for gathering and compiling texts to create a sub-corpus of approximately 88,000 words. The total length of the 12-part ConCoW corpus is 1,053,751 words.

Compiling the corpus involved using randomizing software to select entries within a given Wikipedia portal. An initial problem centered on randomly selected entries that proved to be a set of links to other entries rather than actual texts. This was resolved by agreeing upon a scheme that led to the random selection of one of the links on such pages. Following corpus-building procedures described in course readings, the teachers decided that once an entry had been selected, reference lists, bracketed links to related topics, and excerpts in foreign languages should be removed. They also opted to remove non-text material such as pictures and charts but to retain subheadings and picture captions. Entries under 300 words in length were not included in the corpus; this decision was motivated by the fact that many entries under 300 words proved to be ‘stubs’ that request Wikipedia user to contribute material on a particular topic. Since such texts are unlikely to be used in activities with language learners, the teachers felt they should not be represented in a pedagogical corpus. No other restrictions on length of entries were set. As a result, lengths of the 347 entries that make up ConCoW vary considerably. The mean length is 3037 words (SD = 3043).

The teachers found the corpus building process somewhat tedious but generally unproblematic; using software tools to analyze the corpus proved to be more challenging. Once the appropriate tool was identified (usually with guidance from the instructor), many participants narrowed the scope of their investigations considerably in order to make amounts of data more manageable.

3 Exploring the corpus

The teachers explored the Wikipedia corpus in an assignment that asked them to investigate a question of interest and to link findings to their ESL classroom practice. They had been familiarized with a variety of online corpus tools available at Cobb’s Lextutor website, and the ConCoW was added to the bank of searchable corpora at Lextutor’s concordancing link (http://www.lextutor.ca/conc/eng/). Teachers also worked with offline tools such as AntWordProfiler (http://www.laurenceanthony.net/software/antwordprofiler/). Questions they investigated are as follows:
• Is knowledge of academic vocabulary useful for reading Wikipedia? How well does the Academic Word List (Coxhead, 2000) ‘cover’ the text of Wikipedia entries?
• How frequently will ESL learners meet high-frequency phrasal verbs (as identified by Gardner and Davies, 2007) in Wikipedia entries?
• How neutral is the language of Wikipedia entries? More specifically, how frequent are evaluative expressions such as surprisingly, amazing and undoubtedly? How do selected ConCoW entries compare to earlier versions of these entries?
• Which forms and meanings of five highly frequent AWL words (analyse, consist, involve, period, source) occur most frequently? What are the most frequently occurring collocations for each?
• How many English words would learners need to know to read the entire Wikipedia corpus (assuming they would need to know at least 98% of the words)? Are there differences across subcorpora?
• How important is it to understand the proper names that occur in Wikipedia? What proportion of the corpus do they represent?
• What verbs are used with the nouns life, society, time, children and money? The nouns were selected from a study by Laufer and Waldman (2011) that investigated learner writing.
• To what extent does the vocabulary of a novel (e.g. Verne’s Journey to the Centre of the Earth) overlap with the vocabulary of the Wikipedia subcorpus that corresponds to the topic of the novel (e.g. Geography)?
• What are the most frequent specialist words (keywords) found in two subcorpora that can be expected to differ substantially from each other in terms of content – Nature and Religion?
• How often will learners meet verbs in the regular –ed past in Wikipedia entries? Of particular interest is the use of –ed past in contexts that may push the acquisition of lexical aspect features that are known to be difficult (Collins et al., 2009).

It is worth noting that although course readings had drawn the teachers’ attention to a wide range of topics in applied corpus research including studies of grammar, pragmatics, and discourse, all of their questions (with possible exception of the last one) focus on vocabulary. Three of the questions explore collocations.

Overall, the teachers’ corpus investigations tended to confirm findings of other corpus studies that had addressed similar questions, but there were also new insights. For example, a comparison of old and new versions of the same Wikipedia entry showed that the more recent version was relatively neutral in tone with fewer instances of coloured language and evaluative expressions. Generally, findings were seen as useful with direct applications to classroom teaching. In the case of the aforementioned example, it is easy enough to envision a data-driven learning activity in which ESL students in an academic writing class are asked to compare coloured and neutral versions of the same Wikipedia entry and identify the vocabulary that makes the difference.

References
Two Empirical Examinations of the Effect of Guessing on VST Scores

Jeffrey Stewart  
Kyushu Sangyo University  
jeffjrstewart -at- gmail.com

Stuart McLean  
Osaka Jogakuin University  
stuart93 -at- me.com

Brandon Kramer  
Osaka Jogakuin University & Kansai University  
Brandon.L.Kramer -at- gmail.com

1 Introduction

The Vocabulary Size Test (VST) was created to provide a reliable estimate of learners’ written receptive vocabulary size, measuring from the most frequent fourteen 1000-word families of the spoken subsection of the British National Corpus (Nation & Beglar, 2007). While Beglar (2010) and Elgort (2013) recommend that users should limit the amount of the test taken to only slightly above a student’s level, Nation (2012), Karami (2012), and Coxhead, Nation, and Sim (2014) argue that learners should take every level. However, there are concerns that correct responses on lower-frequency levels could largely be attributed to guesses rather than vocabulary knowledge (Stewart, 2014). Due to this, there have been calls to determine which sections of the VST are appropriate for given levels of learner proficiency; once mean scores on a level approach 25%, scores may cease to operate as indicators of vocabulary knowledge. This presentation reports on two recent empirical studies regarding the effect of guessing on VST scores, for the purpose of determining these cut points.

2 Study 1

In the first study, as suggested by Stewart (2014), we analyzed a data set of 3,373 Japanese university students’ responses to the first eight levels of the original VST under the 3PL model, in order to determine the minimum expected score on the test for learners of low ability, examined the proportion of low-level students’ scores on the lowest frequency level tested attributable to guessing under the 3PL model, and conducted a model fit comparison to determine whether the 3PL model offers a significantly better description of the data than the Rasch model. The results indicate that a substantial portion of lower-level learners’ scores on items testing low-frequency words can be attributed to guessing. For further information on this study, please refer to McLean, Kramer & Stewart (2015).

3 Study 2

A drawback of the first study was that only the first eight levels of the 14,000-word version of the VST were tested, and presumably, guessing effects would be most prevalent for levels of the VST measuring knowledge of lower-frequency words. In the second study, 196 Japanese university students took both a version of the VST with items from all 14 levels, and, prior to that, an English-Japanese translation test of those same words, in order to compare the VST to an alternate measure of written receptive vocabulary knowledge. 70 words on the original 14K version of the VST were tested using the translation test, followed by the corresponding original VST items. In order to reflect the possibility that the VST rewards partial knowledge, responses on the translational test reflecting partial knowledge were marked correct. As seen in Figure 1, although the smaller sample size precluded an analysis under the 3PL model, it is possible to compare item difficulties between the two tests.

Figure 1. Mean VST / Passive Recall Item Difficulty Comparisons by Word Level
At the 10,000-word level, average item difficulty for the VST approaches 0.25 (25%), while average item difficulty for the same words tested on the test of passive recall fall to below 0.05. Given the relatively high English proficiency of many of the Japanese learners tested in this study (students of a prestigious private university), the results indicate that levels of the VST above the 9K level may be too difficult for most Japanese learners.

4 Conclusion

Scores on the VST items were on average 78% higher than the translational measure of receptive knowledge, and the correlation between the two tests of the same words was low at 0.65. These results support Beglar (2010) and Elgort’s (2013) position that students should not sit every level of the test. While the VST is useful in distinguishing learners of varying written receptive vocabulary knowledge, we recommend using caution when interpreting vocabulary size under the assumption a number of correctly-answered multiple-choice questions is directly equivalent to an equal number of known words.

References


Effects of Watching Captioned TV Series on Vocabulary Acquisition in High-School EFL Learners

Ferran Gesa
Universitat de Barcelona
ferran.gesa -at- ub.edu

Imma Miralpeix
Universitat de Barcelona
imiralpeix -at- ub.edu

1 Introduction

Paivio’s Dual Coding Theory (Paivio, 1986) postulates that verbal and non-verbal stimuli are processed by two different systems that interact and that the activation of both systems results in greater depth of processing and better information recall. In light of this theory, the simultaneous presentation of different types of input should then enhance the acquisition of a second language. Bimodal input (consisting of text and sound) has been shown to result in better learning (Bird & Williams, 2002), and research so far indicates that multimodal input (consisting of text and video) can also be advantageous for this purpose (e.g. Price, 1983; Baltova, 1999).

In foreign language learning, when exposure amounts to a few hours of formal teaching and input is not usually varied in genres and registers, maximising the contact with real language input is crucial to facilitate learning of the target language. In this respect, multimedia materials can be a good way to promote learning.

Research has shown that multimodal input can help in developing L2 listening comprehension and, to a lesser extent, vocabulary acquisition (Montero Perez et al., 2013). However, most of the research available has been conducted using short clips to which participants were exposed once or twice (e.g. Yuksel & Tanriverdi, 2009) and this practice may not be enough for substantial gains to occur. Very few longitudinal studies have been performed on the possible benefits of sustained exposure to multimodal input, such as TV series (e.g. Rodgers, 2013). Furthermore, the majority of studies published on the topic have been conducted with adult language learners, whose proficiency levels tend to be rather advanced (e.g. Etemadi, 2012). Therefore, there is a lack of research on how regular exposure to multimodal input over extended periods of time can help learners of younger ages and at different proficiency levels.

2 Research questions

The present study aims at filling these gaps by exposing high-school EFL learners to subtitled TV series for two academic terms by means of a pedagogical intervention, which was especially devised to answer the following research questions:

- Does sustained exposure to L2 subtitled TV series lead to an improvement in vocabulary learning?
- Does learners’ proficiency level have an effect on vocabulary learning from L2 subtitled TV series?

3 Methodology

This longitudinal study involved two groups of EFL learners: an experimental group (EG; N=33) and a control group (CG; N=31). They were all Catalan / Spanish bilinguals at Grade 10 (mean age = 15 years) in a state-funded school in Catalonia (Spain). In each of the groups, there were low-intermediate (N=30 in total) and upper-intermediate (N=34 in total) students.

The pedagogical intervention consisted of 15 sessions (one per week over two terms) in which learners in the EG were pre-taught a series of target words (TWs) (N=5 per episode) through a focus-on-form activity and then watched an L2 subtitled 25-minute episode of a TV series. Immediately afterwards, they took a vocabulary test which asked them to recall the form of the TWs and recognize their meaning. The CG followed the same procedure, but students did not watch the TV episode. All students also took a vocabulary pre-test and a post-test at the beginning and end of each term which assessed forms and meanings of all the TWs introduced that term.

4 Results

Regarding the first research question, a mixed between-within ANOVA with scores from the immediate tests showed a substantial main effect for time \( F(14, 49)=13.15, p=.000, \) partial eta squared=.790, although students’ performance across the fifteen sessions seems to be episode-dependent. The main effect comparing the two conditions (viewing TV series or not) was not significant \( F(1, 62)=.649, p=.423, \) partial eta squared=.010, suggesting that participants performed in a similar way in the immediate vocabulary tests.
regardless of seeing the TV series or not. However, independent-samples $t$-tests with gains (i.e. the pre-test score subtracted from the score of the post-test), revealed that students in the EG were able to recall a significantly higher number of TWs’ forms than learners in the CG both in the first term [EG – $M$=8.70, $SD$=5.41; CG – $M$=6.10, $SD$=4.45; $t$(58)=2.03, $p$=.047] and in the second term [EG – $M$=7.80, $SD$=4.24; CG – $M$=4.92, $SD$=3.79; $t$(53)=2.63, $p$=.011]. Additionally, a Mann-Whitney $U$ Test showed that the number of TW meanings that students could recall in the first academic term was greater for those participants in the EG ($Md$=4, $N$=30) than for those in the CG ($Md$=2, $N$=30) ($U$=288.5, $z$=-2.405, $p$=.016), although these differences were not found in the second term.

If the learners’ proficiency levels are taken into account, gains only reached statistical significance at the upper-intermediate level. In the first academic term, differences were found in the number of word meanings that learners were able to recall [EG – $M$=9.06, $SD$=5.37; CG – $M$=4.06, $SD$=2.84; $t$(22.465)=3.315, $p$=.003]. In the next term, significant differences were not just found in word meaning recall [EG – $M$=7.19, $SD$=3.31; CG – $M$=3.77, $SD$=2.59; $t$(27)=3.041, $p$=.005], but also in the number of word forms that upper-intermediate students remembered [EG – $M$=9.19, $SD$=4.35; CG – $M$=5.77, $SD$=2; $t$(21.986)=2.643, $p$=.011].

5 Discussion and conclusion

The study shows that sustained exposure to multimodal input can favour vocabulary acquisition as learners in the EG obtained greater gains than participants in the CG, this difference being significant in most cases when several episodes had been watched. Therefore, these findings would confirm Paivio’s Dual Coding Theory (Paivio, 1986) in relation to vocabulary acquisition. Simultaneous exposure to text and video seems to lead to better learning and this is shown in delayed testing (as opposed to immediate testing); there were no differences in the scores participants obtained in the immediate vocabulary tests, when learners were asked about those TWs they had just been introduced to in the same session. Eventually, upper-intermediate learners appear to benefit more from such practice than lower-intermediate learners, hinting to the possible existence of a proficiency threshold after which learners take more advantage of being exposed to L2 subtitled TV series.

Our results seem then to corroborate previous findings by showing that exposure to multimodal input leads to better learning of target vocabulary, but they make evident that longitudinal data may be necessary to see the extent of the benefits. The study also adds to previous research by suggesting that various aspects of word knowledge may be differently enhanced by this practice.

6 Acknowledgements

Funding for this project was provided by the Spanish Ministry of Economy and Competitiveness (FFI2013-47616-P and BES-2014-068089) and the Generalitat de Catalunya (2014SGR1089). We are grateful to the high-school teachers and students who participated in this study (school in Mataró, Spain).

References


The Teachers’ U.S. Corpus (TUSC)

Eric Dwyer  
Florida International University  
eric.dwyer -at- fiu.edu

S. J. Ehsanzadeh  
Florida International University  
sehsa002 -at- fiu.edu

1 Introduction

Over the last half century or so, researchers of young English language learners (ELLs) in the U.S. have suggested low high school graduation rates relate to students’ failure to master academic language. Recent curricular shifts, known in the U.S. as “Common Core”, suggest that reading for information cuts across all content areas—particularly Math, Science, and Social Studies—not just Language Arts. Primary school teachers, as well as secondary English Language Arts teachers or Reading coaches, are often charged with assisting ELLs’ acquisition of complex language promoting critical thinking. This shift in literacy development perspectives suggest that all content teachers—especially those at secondary levels—should more emphatically address their texts’ vocabulary.

Meanwhile, researchers have attempted to create general and academic word lists (Browne, Culligan, & Phillips, 2013; Coxhead, 2000; Gardner & Davies, 2013; West, 1953); however, to our knowledge, no corpus specifically meeting needs of primary and secondary school teachers and their students has been compiled. Our study is such an attempt, hopefully bringing attention to language levied upon children in U.S. schools.

To this end, we have conducted a content analysis of 47 textbooks, all volunteered by colleague teachers in Miami’s public schools, ranging in all grades from kindergarten to high school in four domains—Language Arts, Mathematics, Science, and Social Studies—which have been scanned, converted to text, and concordanced into a 2.5 million-word corpus affectionately called the Teachers’ U.S. Corpus, or TUSC. The texts were also tagged with respect to one of four grade level chunks: kindergarten to 2nd grade (early primary grades), 3rd to 5th grade (later primary grades), 6th grade to 8th grade (middle school), and 9th to 12th grades (high school).

2 Preliminary results

While creating a school language corpus and establishing useful corresponding lists are our principal objectives, we find comparing our corpus to predecessors (e.g., Coxhead’s AWL) important in distinguishing language written for college and that written for kids. As a result, applying the TUSC to Cobb’s (2016) LexTutor word profiler, which sorts vocabulary into numerous categories such as West’s most frequently implemented words and Coxhead’s AWL, may yield results that help us understand the degree to which teacher language for kids corresponds to adult language needed for college.

Preliminary analyses are striking, as noted in Table 1. As an example, our corpus indicates most kids’ texts implement so-called academic vocabulary around 6% of the time, not the 10% Coxhead reported. Such results raise questions regarding levels of academic vocabulary intensity. In other words, we wonder if students relying exclusively on school texts for college readiness are receiving sufficient academic vocabulary input to prepare them for college.

In addition, Language Arts textbooks implemented the smallest proportions of academic vocabulary with respect to Coxhead’s AWL. As Language Arts teachers are responsible by law for garnering professional development for ELLs and working with them in detailed fashion regarding their language development, these results suggest that energy regarding academic language may be currently misplaced.

In fact, the other 3 subject areas yielded more, with Math generating the most across all grades, including a dynamic upswing in AWL use in secondary school. Indeed, data regarding Math (often touted as a universal language) indicates that Math teachers should undergo substantial professional development regarding vocabulary instruction. Similarly, Social Studies, where cross-cultural interactions are discussed and critical thinking skills are often fostered, should be a go-to subject area where substantial swaths of academic vocabulary may be offered and practiced in context.

Off-list vocabulary—i.e., vocabulary not tagged in any aforementioned list—was heavier in earlier primary grades than it was in secondary grades, suggesting that school-based vocabulary is complex for ELLs when they enter school at early ages. However, the LexTutor-tabulated data, particularly that concerning off-list words, suggests that kids’ vocabulary differs substantially from that intended for adults. In light of such, we implemented a concordance to unveil our own K1, K2, and academic words lists.
Table 1. Proportions of K1 and K2 (West), AWL (Coxhead), and Off-list words in the TUSC, as analyzed through LexTutor (Cobb).

<table>
<thead>
<tr>
<th>Subject</th>
<th>K1</th>
<th>K2</th>
<th>AWL</th>
<th>Off-list</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>grades K-2</td>
<td>60.64</td>
<td>16.56</td>
<td>5.17</td>
<td>17.63</td>
</tr>
<tr>
<td>grades 3-5</td>
<td>58.70</td>
<td>15.57</td>
<td>5.54</td>
<td>20.19</td>
</tr>
<tr>
<td>grades 6-8</td>
<td>68.56</td>
<td>11.12</td>
<td>7.95</td>
<td>12.63</td>
</tr>
<tr>
<td>grades 9-12</td>
<td>80.08</td>
<td>5.02</td>
<td>9.78</td>
<td>5.11</td>
</tr>
<tr>
<td>Overall</td>
<td>70.70</td>
<td>9.97</td>
<td>8.00</td>
<td>11.23</td>
</tr>
<tr>
<td>Science</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>grades K-2</td>
<td>77.50</td>
<td>7.54</td>
<td>3.69</td>
<td>11.25</td>
</tr>
<tr>
<td>grades 3-5</td>
<td>62.10</td>
<td>23.81</td>
<td>3.94</td>
<td>10.15</td>
</tr>
<tr>
<td>grades 6-8</td>
<td>75.49</td>
<td>7.03</td>
<td>6.70</td>
<td>10.77</td>
</tr>
<tr>
<td>grades 9-12</td>
<td>67.96</td>
<td>17.05</td>
<td>4.50</td>
<td>10.49</td>
</tr>
<tr>
<td>Overall</td>
<td>67.96</td>
<td>17.05</td>
<td>4.50</td>
<td>10.49</td>
</tr>
<tr>
<td>Language Arts</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>grades K-2</td>
<td>78.57</td>
<td>11.80</td>
<td>1.73</td>
<td>7.90</td>
</tr>
<tr>
<td>grades 3-5</td>
<td>82.22</td>
<td>6.21</td>
<td>4.44</td>
<td>7.11</td>
</tr>
<tr>
<td>grades 6-8</td>
<td>79.58</td>
<td>7.21</td>
<td>3.44</td>
<td>9.50</td>
</tr>
<tr>
<td>grades 9-12</td>
<td>83.59</td>
<td>6.34</td>
<td>5.48</td>
<td>4.45</td>
</tr>
<tr>
<td>Overall</td>
<td>82.70</td>
<td>6.44</td>
<td>4.88</td>
<td>5.96</td>
</tr>
<tr>
<td>Social Studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>grades K-2</td>
<td>85.14</td>
<td>6.10</td>
<td>4.91</td>
<td>3.84</td>
</tr>
<tr>
<td>grades 3-5</td>
<td>83.09</td>
<td>5.79</td>
<td>6.31</td>
<td>4.81</td>
</tr>
<tr>
<td>grades 6-8</td>
<td>79.56</td>
<td>6.10</td>
<td>5.91</td>
<td>8.41</td>
</tr>
<tr>
<td>grades 9-12</td>
<td>81.69</td>
<td>6.00</td>
<td>5.86</td>
<td>6.44</td>
</tr>
<tr>
<td>Overall</td>
<td>81.69</td>
<td>6.00</td>
<td>5.86</td>
<td>6.44</td>
</tr>
</tbody>
</table>

Preliminary versions, based on 1 million words across all domains and grades, are found at this website: http://www.ericsdwyer.info/TUSC. When these lists are sifted through LexTutor, only 81.66% of TUSC K1 words correspond to West’s K1 list. In fact, 9.01% are considered off-list, and another 1.41% are represented by Coxhead’s AWL, thereby demonstrating distinct differences between adult vocabulary and that intended for schools. However, The TUSC academic word list, also made up of all non K1 and non K2 words but falling in all 4 subject domains, is made up completely (100%) of Coxhead’s AWL words, indicating that both her list and the shorter TUSC version thereof are indeed useful for kids’ classes.

References
A Comparison of Multiple-Choice and Yes/No Test Formats with a Meaning-Recall Knowledge Criterion

Stuart McLean  
Osaka Jogakuin University  
kansaiu.ac.jp/stuart93

Jeff Stewart  
Kyushu Sangyo University  
jeffrstewart@gmail.com

Brandon Kramer  
Osaka Jogakuin University &  
Junior College  
Brandon.L.Kramer@gmail.com

1 Introduction

The ability to receptively recall the meaning of individual words while reading has long been recognized as a pre-requisite for successful reading comprehension (Hu & Nation, 2000; Grabe, 2009; Qian, 1999). Meaning-recall tests are recognized as an accurate, reliable, and content valid method of determining students’ actual lexical knowledge. However, the marking of meaning-recall tests is laborious. Therefore, self-report Yes/No (YN) checklists and multiple-choice tests such as the Vocabulary Size Test (VST) (Beglar & Nation, 2007) are more commonly used as measures of L2 written receptive vocabulary knowledge due to their convenience and ease of marking. If these tests correlated highly to tests of receptive recall, they could serve as useful proxies for measurement on this construct. However, scores on such measures could vary for construct-irrelevant reasons, such as individual differences in self-reports and guessing effects in multiple-choice item formats. To what degree will scores differ on these tests relative to a more time-consuming alternative, such as a receptive test of L2 vocabulary knowledge which asks learners to write L1 definitions? The purpose of the study is to determine which more expedient test format most closely approximates a more laborious translational test of receptive word knowledge: a YN checklist, the original Vocabulary Size Test (VST) (Nation & Beglar, 2007), or a Japanese bilingual version of the VST (Mclean et al, 2016).

2 Methods

Participants (N = 397) were a mixture of first- and second-year Japanese university students from three private universities in western Japan. In the present study, because the meaning recall measure was to be used as the criterion measure, it was important to quantify this as accurately as possible, without the potential influence from having previously seen answer choices on the VST. Thus, participants completed a YN checklist test, then a meaning-recall task, and finally one of two VST forms. All tests were administered via the online survey website Survey Monkey (https://www.surveymonkey.com) with no time limit. Settings were used within Survey Monkey to ensure that learners were unable to revisit one modality of the test once they had completed it.

This study utilized four different instruments. First, a YN checklist test consisting of 100 items. The checklist test consisted of 70 words which appear on the 14,000-word version of the VST, with 5 words randomly chosen from each 1000-word level. The YN checklist test also consisted of 30 pseudowords. All pseudowords were randomly selected from EFL Vocabulary Tests (Meara, 2010). Both the real words and pseudowords were presented in a simple non-defining context. The pseudowords were added to the test from the 3,000-word level of the test, and were increasingly present in each 1,000-word level as the test progressed. This test was accompanied with the following instructions in Japanese:

Please read the sentences from 1 to 100 carefully. If you know the underlined words, mark YES. If you do not know the underlined word or are not sure, please mark NO.

The second instrument was a meaning-recall test testing participants’ knowledge of the real words from the YN checklist test. Participants were asked to translate the meaning of the target word. Students were instructed that when necessary an explanation of the target word was acceptable. The same simple non-defining context sentences from the YN test were used in the meaning-recall test. The test was accompanied with the following instructions in Japanese:

This is a test of your English vocabulary knowledge. For each question, there is a target word followed by an example sentence using the word. For your answer, use Japanese to write the meaning of the target word. Be careful because some English words have multiple meanings; you should write the meaning of the word as it appears in the example sentence. You may answer in several different ways. If you know the corresponding Japanese word, simply write the Japanese translation. If you know the meaning of the English word but are unsure of the best Japanese translation, you may write a short phrase in Japanese to explain the meaning.
Finally, if you have only partial knowledge of the English word, please write an answer showing this partial knowledge. For example, for "cake", you could write “cake” if you know the exact meaning, or perhaps “a kind of food” if you know only that it is a kind of food. If you have no idea of the word meaning, you do not need to give any answer.

The third instrument consisted of either a monolingual or bilingual version of the VST (both available from Paul Nation’s website), randomly assigned to students within each intact class within the study. Items from these tests similarly measured participants’ knowledge of the real words from the YN and meaning-recall checklist test, with distractors either written in simple English or the students’ L1 Japanese. The same simple non-defining context sentences from the YN test were used in the meaning-recall test. Both instruments were accompanied by the following instructions in Japanese:

This is a different vocabulary test. There is a total of 70 questions in the test. Please choose the best answer from the given alternatives below and answer all questions.

3 Results

With a mean of 14.53 (20.8%), scores on the L2-L1 translation test were markedly lower than on either the YN checklist or the two VST versions; scores were 52% higher on the YN checklist, 78% higher on the original VST items and 131% higher on the Japanese bilingual VST items. Correlations of these measures to the translation task were correspondingly low, at 0.64, 0.65 and 0.62 respectively. The results indicated that the tested learners knew relatively few low-frequency words used on the VST. Following Beglar’s (2010) and Elgort’s (2012) recommendation that learners take only level-appropriate levels of the VST, the analysis was re-run on only words from the first seven 1000-levels of the test. Scores on this half of the test had a mean of 12.61 (36%), with overestimation rates dropping to 26%, 36%, and 80% on the YN test, Monolingual VST, and Bilingual VST, respectively. Correlations to written receptive knowledge as measured by the translation test also improved markedly, rising to 0.80, 0.75 and 0.71, respectively.

References


Exploring the Effect of a Collaborative Task on Vocabulary Acquisition: The Role of Language-Related Episodes during a Dictogloss

Mutsumi Kawakami
Tokyo Denki University
mutsumik -at- mail.dendai.ac.jp

1 Introduction

Many English learners in Japan traditionally tend to memorize word meanings by using lists, written and oral drills, and flash cards with little opportunity to use the words. The Ministry of Education has put pressure on English teachers to seek better ways for their students to acquire English for production. Research over the past two decades has demonstrated that collaborative learning during tasks can facilitate second language vocabulary acquisition, contingent upon learner interaction (e.g., Kim, 2008; Swain & Lapkin, 2001). Negotiation and modified output are also reported to contribute to vocabulary acquisition (e.g. Ellis & He, 1999). These studies use output-based tasks that enable learners to focus on form to complete the task, such as text reconstruction tasks in groups. On the other hand, some studies showed that deliberate vocabulary learning, such as memorization with word cards enhanced vocabulary knowledge for production (e.g., Elgort, 2010). Recent studies provide evidence that multiple encounters in context with form-focused activities were more important for vocabulary acquisition (e.g., Folse, 2006; Laufer & Rozovski, 2011).

2 Method

This study explores the efficacy of a collaborative text reconstruction task, dictogloss. Participants of this study came from two intact classes for reading and writing. The experimental group was 23 students whose major was electric and mechanical engineering and the control group was 22 students whose major was information system design. Their proficiency level ranged from CEFR A1 to A2 based on TOEIC Bridge scores. The participants in the experimental group were randomly assigned to six groups of three or four. The transcripts from the learners’ interaction in a dictogloss task were analyzed for instances of language related episodes (LREs), based on the categories presented in Swain and Lapkin (1995). Vocabulary knowledge is measured by a pretest and two posttests, administered immediately and three weeks after the treatment. The research questions addressed in this study are:

1. What are the effects of a dictogloss task on the students’ vocabulary retention?
2. What is the main focus in LREs?
3. Will LREs lead to the vocabulary acquisition?

3 Results

The results showed that participants in the both groups gained vocabulary in the receptive vocabulary test immediately after the treatment and maintained their scores for three weeks. A slight decrease was found only in the control group in the receptive delayed posttest. Additionally, students in the experimental group outperformed in the two productive posttests. Investigating LREs, learners attended more to lexical features (76%) than syntactic features (24%). LREs of Group A and Group F indicate that there are many correct resolutions without negotiation of meaning among peers. Their scores in the productive delayed posttest decreased. Furthermore, The LREs of Group B and F tended to result in correct resolutions but the students had less opportunities to negotiate meaning. Their scores in the productive delayed posttest decreased significantly compared with other groups.

<table>
<thead>
<tr>
<th>Group</th>
<th>Control group</th>
<th>Experimental group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>M</td>
</tr>
<tr>
<td>Pretest</td>
<td>23</td>
<td>7.3</td>
</tr>
<tr>
<td>Immediate posttest</td>
<td>23</td>
<td>9.8</td>
</tr>
</tbody>
</table>

Table 1. Receptive Vocabulary Test.
<table>
<thead>
<tr>
<th>Group</th>
<th>Control group</th>
<th>Experimental group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>M</td>
</tr>
<tr>
<td>Immediate posttest</td>
<td>23</td>
<td>4.2</td>
</tr>
<tr>
<td>Delayed posttest</td>
<td>23</td>
<td>2.6</td>
</tr>
</tbody>
</table>

Table 2. Productive Vocabulary Test.

4 Discussion

The results of this study indicated the dictogloss task can help retain both receptive and productive vocabulary knowledge. In addition, the main focus in LREs was lexis. The students in this study often discuss meaning of words rather than syntactic features of words. Furthermore, correct resolutions of LREs through negotiation seemed to result in vocabulary acquisition.

References


The Effect of Data-Driven Learning on Vocabulary Learning and Retention on Thai EFL Learners

Nuchsara Choksuansup
Kasetsart University
nuchnuchchok-at-hotmail.com

1 Introduction

While vocabulary is the heart of language learning, there are many problems exist in learning and teaching English in Thailand. Thai learners struggled to learn English. Using attractive and appropriate methods will enable learners to enrich information about the target word and to integrate those words into long term memory. Data-Driven Learning coined by Tim Johns is characterized as “autonomic learning”, “authentic learning”, “self-discovery”, and “inductive learning”. Developed from corpus linguistics in order to give learners authentic and hands-one experiences with real data through concordance lines, DDL plays a new role in revolutionary vocabulary teaching in Thailand which greatly helps to develop vocabulary learning and teaching. Few studies have been based in Thailand indicating that not many teachers know of this method. Therefore, this study aimed to investigate the effect of integrating the Data-Driven Learning (DDL) with vocabulary learning and also the vocabulary retention of the learners.

2 Method

The study started in August-October 2015 and two groups with the same number of learners were voluntarily involved. One was experimental group, the other was control group. The printouts of concordance line from British National Corpus (BNC) were applied with the 40 learners from the former group. The latter group was controlled by using traditional vocabulary instruction and traditional vocabulary exercises. 60 words selected from the tourism book which was used as a supplementary for this course was assigned to the learners as target words. Before the class had started, learners’ chose 60 words out of 200 words which they did not know. Thus, 60 target words were originally selected from the high frequency words that learners thought they did not know or see before.

To measure the learners’ achievement, an achievement test which included 60 multiple choices items were used assess learners’ both group. The statistic used to compare learners’ performance of the two groups was an independent-sample t-test. Three weeks later, a list of 30 words randomly brought from the vocabulary which had taught in class during the experiment period in order to measure learners’ retention. Again, an independent sample was used to compare two groups on their vocabulary retention.

3 Findings and discussion

In this study, the researcher studied the effect of Data-Driven Learning (DDL) on vocabulary learning and vocabulary retention of EFL learners. The learners’ achievement test was analyzed by using t-test scores. The result showed that there was a significant difference between the performance of the experimental and control group.

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>24</td>
<td>5.17</td>
</tr>
<tr>
<td>Experimental</td>
<td>45</td>
<td>7.04</td>
</tr>
<tr>
<td>p = 0.03</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>13</td>
<td>6.24</td>
</tr>
<tr>
<td>Experimental</td>
<td>22</td>
<td>4.67</td>
</tr>
<tr>
<td>p = 0.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p< 0.05

Table 1: Achievement Test Scores

Table 2: Vocabulary Recall Test Scores

The t-test result indicates that the experimental group of learners who participated in Data Driven Learning technique performs much better than the control group. The achievement Mean score of the experimental group is 45 and the control group Mean score is 24. For the vocabulary recall test, the experimental group scores 22 and control group scores 13 with significant difference p. value at 0.00.

Findings support Schmitt (2000) who displayed the advantages of using concordance in language teaching. The findings are also consistent with Cobb (1997b) who claimed that Data Driven Learning promotes learners’ enthusiasm and memory. All findings in this study seem to ensure that Data Driven
Learning enriches vocabulary teaching and learning, besides it contributes learners’ vocabulary retention as mentioned that Data Driven Learning helps learners interact the authentic language which are normally used by native speakers which can convince the learners than the example in grammar books (Cobb, 1997; Mayer, 2004).

4 Conclusion and implication

The term “Data-driven learning” (DDL) has been coined by Tim Johns in 1991. He primarily used DDL as an inductive learning process, whereas the teacher did not even know the answer or the patterns that the student would discover (Johns, 1991a). Another benefit of DDL is that it increases learners’ autonomy potential in the classroom, with less reliance on the teacher. This autonomy, as well as inductive and deductive learning methods, gives opportunities for the development of cognitive skills (Boulton, 2009). With greater autonomy, students can also use DDL techniques to answer their own questions about language, as well as to become more independent language learners (Hunston, 2002). This study explored the effect of Data-Driven Learning on vocabulary learning and retention of Thai EFL learners. Findings suggest that supporting the vocabulary learning with concordance lines provides positive impact on learners and also helps them retain the meaning of those words more efficiently.

The corpora and corpus materials specifically have affected on vocabulary issues such as grammar, idioms and other phrases, choosing what words to teach, and knowing what a word means. Therefore, applying DDL to classroom can help both teachers and students explore some issue that might not be discovered before (Nation, 2001). Moreover, DDL in vocabulary teaching provides great opportunities for learning more about vocabulary, teachers who are aware of some aspect of vocabulary issues can use DDL to aid their students in comprehending the full meaning of words and practicing them in the classroom (Flowerdew, 2009).

5 Acknowledgements

This study was supported by Kasetsart University, Thailand. I would like to show my gratitude to the graduate school of Kasetsart University for funding me in presenting this paper. I am also immensely grateful to Asst. Prof. Dr. Montri Tangpijaikul, my thesis supervisor, for his insightful comment and also a huge encouragement.

References

The “I Don’t Know” Option and L1 Answer Choices: A Comparison of Four Variants of the Vocabulary Size Test

Tim Stoeckel  
University of Niigata Prefecture  
timstoeckel -at- gmail.com

Jeffrey Stewart  
Kyushu Sangyo University  
effjrstewart -at- gmail.com

1 Introduction

The Vocabulary Size Test (VST) (Nation & Beglar, 2007) is designed to assess the overall size of examinees’ written receptive vocabulary knowledge, the type of lexical knowledge needed for reading. The instrument’s multiple-choice (MC) format makes it a test of “meaning recognition,” the ability to recognize a word’s meaning from a list of choices; however, the type of lexical knowledge actually needed in reading more closely approximates “meaning recall,” or the ability to recall from memory the meaning of a tested word (Gyllstad et al., 2015). Despite its otherwise sound psychometric characteristics (Beglar, 2009), the VST has consistently overestimated the lexical knowledge needed for reading when compared to criterion meaning recall measures (e.g., Zhang, 2013). In recent years, researchers have added an ‘I don’t know’ (IDK) answer choice to the VST to reduce overestimation of meaning recall knowledge from the use of test strategies or guessing under the test’s MC format. IDK has resulted in lower scores (Zhang, 2013), but there has been concern that if used differentially, it may introduce a non-relevant construct affecting scores and weakening validity (Stoeckel et al., in press). Scholars have also rendered VST answer choices in examinees’ L1 to avoid conflating vocabulary knowledge with knowledge of language used in the test items. Bilingual formatting has yielded somewhat higher scores (Elgort, 2013), but the reasons for this have not been determined. The present study evaluated four variants of the VST (monolingual, monolingual+IDK, bilingual, and bilingual+IDK) and addressed two research questions: (1) Which variant produces scores that are most representative of a criterion measure of meaning recall knowledge? and (2) Does differential use of IDK lead to differences in VST scores for examinees with similar levels of meaning recall knowledge?

2 Methods

Two hundred Japanese university students completed (a) a meaning recall task in which they were asked to provide written translations of the words tested on the VST and (b) one of the four VST variants. Significance was set at .05 throughout.

3 Results

Reliability indices ranged from .84 to .92 on the VST and from .86 to .93 on the meaning recall task for the four study groups. Because there were no significant differences in mean scores on the meaning recall task, the four groups were considered equivalent in terms of the criterion measure for the purposes of the study.

To address RQ1, three analyses were performed. First, VST mean scores were compared. The left side of Table 1 shows the bilingual variant produced significantly higher scores than the monolingual and monolingual+IDK variants. Second, correlations between meaning recall and VST scores were compared. IDK corresponded with a lower r value in both the monolingual and bilingual formats (Table 1, right side).

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>M</th>
<th>(SD)</th>
<th>p</th>
<th>r</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>1. monolingual</td>
<td>50</td>
<td>44.8</td>
<td>(9.9)</td>
<td>*</td>
<td>.905</td>
<td>*</td>
</tr>
<tr>
<td>2. monolingual+IDK</td>
<td>50</td>
<td>41.8</td>
<td>(12.5)</td>
<td>*</td>
<td>.866</td>
<td>*</td>
</tr>
<tr>
<td>3. bilingual</td>
<td>50</td>
<td>51.3</td>
<td>(8.7)</td>
<td></td>
<td>.906</td>
<td>*</td>
</tr>
<tr>
<td>4. bilingual+IDK</td>
<td>50</td>
<td>47.0</td>
<td>(8.3)</td>
<td></td>
<td>.774</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Scores and Correlation with a Measure of Meaning Recall Knowledge for Four VST Variants

Third, we examined the percentage of correct VST responses at three levels of meaning recall knowledge (known, partially known, unknown). Table 2 shows that although IDK was associated with fewer correct responses on items testing unknown words (a desirable outcome), a large amount of inflation persisted. Regarding test language, the bilingual versions produced greater accuracy on items testing known words (also a desirable outcome) but also greater success on items testing unknown words, an unintended source of...
inflation with this format. For RQ1, due to higher correlations with the criterion measure, it appears that the two non-IDK variants are more reflective of the tested construct. The lower correlations for the IDK variants raise the possibility that the secondary construct of willingness to use IDK has affected scores.

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Known</th>
<th>% Correct</th>
<th>p</th>
<th>Partially Known</th>
<th>% Correct</th>
<th>p</th>
<th>Unknown</th>
<th>% Correct</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. monolingual</td>
<td>50</td>
<td></td>
<td>89.7</td>
<td>*</td>
<td>*</td>
<td>60.0</td>
<td></td>
<td>35.3</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>2. monolingual+IDK</td>
<td>50</td>
<td></td>
<td>90.5</td>
<td>*</td>
<td>*</td>
<td>65.1</td>
<td></td>
<td>29.6</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>3. bilingual</td>
<td>50</td>
<td></td>
<td>96.1</td>
<td></td>
<td></td>
<td>83.8</td>
<td></td>
<td>44.5</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>4. bilingual+IDK</td>
<td>50</td>
<td></td>
<td>96.3</td>
<td></td>
<td></td>
<td>74.4</td>
<td></td>
<td>37.7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Percentage of Correct VST Items at Three Levels of Meaning Recall Knowledge

For RQ2, we first confirmed differential use of IDK on both IDK variants. We then identified all pairs of learners on all four variants who had the same meaning recall score and calculated the difference between the two VST scores for each pair. The left side of Table 3 shows that the two IDK variants corresponded with larger score differences for learners with identical meaning recall scores. To see if this was associated with differential IDK use, we examined IDK use for the person with the higher VST score and for the person with the lower VST score in each pair. IDK use was significantly higher for examinees with the lower VST score (Table 3, right side). In other words, increased IDK use was associated with lower VST scores for learners with identical meaning recall ability. We thus conclude that differential use of IDK was associated with differences in VST scores for learners with similar levels of meaning recall knowledge.

<table>
<thead>
<tr>
<th>Group</th>
<th>n (pairs)</th>
<th>VST Score Difference</th>
<th>IDK Use for Persons in Each Pair</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. monolingual</td>
<td>24</td>
<td>4.25 (2.92)</td>
<td>Higher VST Score: M (SD)</td>
</tr>
<tr>
<td>2. monolingual+IDK</td>
<td>28</td>
<td>6.71 (4.50)</td>
<td>Lower VST Score: M (SD)</td>
</tr>
<tr>
<td>3. bilingual</td>
<td>22</td>
<td>3.27 (2.71)</td>
<td>p</td>
</tr>
<tr>
<td>4. bilingual+IDK</td>
<td>36</td>
<td>6.81 (4.90)</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Differences in VST Scores and IDK Use for Pairs of Learners with Identical Meaning Recall Scores

4 Conclusions

The findings suggest that regarding IDK, the convention is problematic as a means to address score inflation because its differential use results in different VST scores for learners of similar meaning recall ability, and this weakens the validity of score interpretations. Regarding the choice between monolingual and bilingual formatting, though, the latter is associated with more score inflation on items testing unknown words, its increased accuracy on items testing known words makes it an attractive format if a method to consistently account for score inflation can be found.

References


Knowledge of Words: Examining the Relationships and Order of Acquisition of Vocabulary Knowledge Dimensions

Beatriz Gonzalez-Fernandez
University of Nottingham
beatriz.gonzalez-fernandez -at- nottingham.ac.uk

1 Introduction

Vocabulary knowledge is widely recognized nowadays as a multidimensional construct. This multidimensional nature involves that vocabulary can no longer be understood as knowledge of only the form-meaning link of a word; instead, it requires the acquisition of various word knowledge components, such as spelling, form-meaning link, derivatives, polysemy and collocations (Nation, 2013). Many researchers have suggested different frameworks of what it comprises to know a word (e.g., Chapelle, 1998; Nation, 2013; Qian, 2002). Nevertheless, only a few studies have examined the acquisition of multiple word knowledge components concurrently (e.g., Chui, 2006; Schmitt, 1998; Webb, 2005). As a consequence, it is still unclear how these different components of vocabulary knowledge are acquired and relate to each other. There is, therefore, a need to study the nature of vocabulary knowledge in order to develop a general theory of vocabulary acquisition.

2 Vocabulary knowledge components

Researchers have tried to shed some light into the complexity of vocabulary knowledge by classifying it into different categories or components (e.g., Nation, 2013). Although there have been some studies that have investigated the acquisition of some of these components of word knowledge individually, they provide only limited insight into depth of word knowledge. As a consequence, research is needed that investigates multiple word knowledge components concurrently.

Only a few studies have looked at the acquisition of various word knowledge components simultaneously. For example, Schmitt (1998) measured the productive knowledge of four aspects of word knowledge: written form, derivative knowledge, associative knowledge and knowledge of polysemous meanings. Webb (2005) used a battery of tests to measure five aspects of word knowledge, both productively and receptively: orthography, form-meaning link, syntax (syntagmatic associations), grammatical functions (word class) and (paradigmatic) associations. Chui (2006) studied the knowledge of four lexical knowledge aspects: word class recognition, meaning recall, collocation recognition and derivative form production. In general, these studies suggest that the various aspects of word knowledge are interrelated somehow and develop more or less in a parallel manner. However, these studies have some limitations. They measure a limited amount of words or non-words, and their findings are based on correlations and mean scores. Therefore, there is a need to study various aspects of word knowledge concurrently, testing a reasonable amount of words by means of a battery of tests, in order to study the interplay between aspects of word knowledge empirically.

Thus, in order to attend to this gap in the literature, the present study aims to:

- Outline an order of acquisition of various word knowledge components by L2 learners
- Explore the relationships between the word knowledge components

In order to answer these questions, a battery of eight tests was designed to measure receptive and productive mastery of 20 target words across four word knowledge components: form-meaning link, derivatives, collocations and polysemy.

3 Methodology

Two different statistical analyses were used to answer the research questions: Implicational Scaling and Structural Equation Modeling.

Implicational scaling is a method of data analysis that enables the establishment of systematic relationships and patterns of acquisition between linguistic features. The main purpose of this technique is to ascertain a one-dimensional scale for a specific concept to be measured, so that if an implicational scale is found we can conclude that there exists a hierarchical relationship between the items so that \( X \) implies \( Y \) but not the reverse.

Structural Equation Modeling is a theory-driven statistical technique that uses a confirmatory approach to the analysis of a hypothesized structural model (Byrne, 2010). It requires the a priori design of a model of
relationships between different variables based on the theory, in order to check the validity of that theory. One of its main advantages is that it allows the measurement of unobservable constructs and can represent causal effects between variables.

4 Participants

A total of 144 Spanish speakers of EFL participated in this study, 42 males and 102 females. The mean age was 25.3 years old, and they had a range of proficiency levels, from beginners to advanced.

5 Results

The results for the implicational scaling analysis show that it is indeed possible to describe a systematic and reliable order of difficulty, and therefore a pattern of acquisition, of the different word knowledge components. This sequence of acquisition suggests that receptive knowledge is always acquired before productive knowledge in all aspects.

For its part, different structural equation models suggest that all the word knowledge components are very strongly intercorrelated, and that they are not acquired independently in a sequential manner, but in parallel and at different rates.

6 Discussion

This study shows that the different aspects of word knowledge are not acquired independent from each another. Rather they are acquired in parallel, but at different rates due to their different difficulty levels. Therefore, practitioners need to take into account this information to start introducing these aspects to the learner in the order suggested according to their difficulty, but from the early stages in the learning process.

References


Exploring the Role of Automatically-derived Text Complexity Features in L2 Reading Test Development

Sumi Han
Seoul National University, Korea
Sumihan20 -at- gmail.com

Jeong-Ah Shin
Dongguk University, Korea
jashin -at- dongguk.edu

1 Introduction

In reading comprehension assessments, texts are the major input that readers need to interpret while answering test questions and thus, the choice of texts as the very first step of test development requires careful decision-making to meet the test goal and test-takers’ reading ability (Alderson, 2000). To date, automatic text analyzers such as Lexile and their features have been widely used in text analysis in the field of corpus research, but little has been done in the field of reading assessments. As a variety of text features influence reading test performance, detailed profiles of texts can be provided; otherwise, test developers would solely rely on their own experiences and intuition. To address the lack of research, this study aimed to explore the relationship between a set of automatically-derived text complexity features and the difficulty of reading items of intensive English program (IEP) placement tests in the United States.

2 Literature Review

Though being defined variably, text complexity often means aspects measured in numbers such as word length or counts, sentence length, and text cohesion, which are typically measured by computer software programs. This automatic measurement of text complexity is most beneficial for predicting the difficulty of texts and selecting and revising long and various texts (Fisher & Frey, 2014). Four major automatic text analysis tools have been widely used in reading test developments: Lexile and FKGL are simply based on total word counts and sentence length in each text; The Compleat Lexical Tutor and Coh-Metrix are more sophisticated tools which quantify lexical, syntactic, or semantic aspects of texts in determining the degree of text complexity.

In early research with little technical help, text complexity features were manually identified and compared with reading comprehension (e.g., Gray & Leary, 1935). With the development of computer programs, more recent research has identified the role of various quantitative features of text complexity in accounting for reading item difficulty. However, different sets of features have been suggested as significant predictors of reading item difficulty: McNamara and Kintsch (1996) found that high text cohesion and coherence improved reading comprehension, especially for low proficient readers, but McNamara et al. (2012) pointed out that texts in high cohesion and coherence can be relatively difficult to read if they include many difficult words, mostly academic vocabulary. The influence of such significant features on text comprehension could vary across studies, and more empirical research is called for.

3 Research questions

This study aimed to meet a practical need of reliable and evidence-based text selection in the IEP test development by identifying automatic features of text complexity that predict the reading item difficulty. The following two research questions are addressed in this study:

1) To what degree do text complexity features correlate with the reading item difficulty of the IEP tests?
2) To what degree do text complexity features predict the reading item difficulty of the IEP tests?

4 Methods

The reading assessments data were from a total of 443 examinees who took one of the Fall 2010, 2011, and 2012 IEP placement tests at a Southwestern university in the United States. The reading tests were developed with a set of text selection guidelines such as use of academic vocabulary and text cohesion, but no empirical evidence had been sought for the use of the recommended text selection guidelines. Based on previous literature and the testing context, a total of eleven text complexity features as lexical, syntactic, and semantic variables were chosen, and their accountability for the IEP reading item difficulty was automatically measured by using three text analysis tools—Lexile, the Compleat Lexical Tutor, and Coh-Metrix: (a) five lexical variables (Lex1-Lex5) including word length, word counts, academic words, off-list word counts, and type-token ratio; (b) three syntactic variables (Syn1-Syn3) including sentence length, noun-phrase density,
and passive voice; and (c) three semantic variables (Sem1-Sem3) including co-reference cohesion, lexical semantic analysis, and connectives.

5 Results

Table 1 lists correlation coefficients between the reading item difficulty (ID; equated delta) and the 11 text variables, showing that the reading item difficulty moderately correlated with each independent variable (absolute coefficients ranging from .23 to .53). In particular, the reading item difficulty had a negative moderate relationship with Sem3 ($r = -.53$) but a positive negative relationship with Lex1 ($r = .52$). All the lexical variables but Lex5 significantly correlated with the item difficulty. Syn2, Sem1, and Sem2 did not demonstrate significant relationships with the item difficulty.

<table>
<thead>
<tr>
<th></th>
<th>Lex1</th>
<th>Lex2</th>
<th>Lex3</th>
<th>Lex4</th>
<th>Lex5</th>
<th>Syn1</th>
<th>Syn2</th>
<th>Syn3</th>
<th>Sem1</th>
<th>Sem2</th>
<th>Sem3</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>.52*</td>
<td>.46*</td>
<td>.34*</td>
<td>.23*</td>
<td>.23</td>
<td>.40*</td>
<td>.04</td>
<td>.37*</td>
<td>-.19</td>
<td>.15</td>
<td>-.53*</td>
</tr>
</tbody>
</table>

*Significant coefficients at the adjusted $p = .005$, 1-tailed

Table 1. Correlation Coefficient between Reading Item Difficulty (ID) and Each Feature

Regarding the results of stepwise multiple regression analysis with the reading item difficulty as the sole dependent variable and the eleven text variables as a set of independent variables, a total of eight regression models with different sets of predictors of the equated delta were found statistically significant at the $p$ level of .000. Sets of five text variables (Sem3, Lex3, Lex2, Lex1, Sem2) contributed to regression across the models, but Sem3 and Lex3 were not significant predictors, in turn, and each of them was deleted in the subsequent model. Thus, the eighth model had the most parsimonious set of significant predictors with the largest predictive power of the variance in the equated delta, with an $F (4, 105) = 23.329, p = .000$ ($R^2$ of .471; adjusted $R^2$ of .450).

6 Discussions

Regarding the first research question, the lexical features except Type-Token Ratio were significantly correlated with the reading item difficulty. Among the three syntactic features, not only sentence length but also the ratio of passive structures significantly correlated with the item difficulty. Moreover, the result shows that the more connectives are used, the more difficult they answered the questions of the text, as also found in McNamara et al.’s (2012) study. Regarding the second research question, a parsimonious set of significant predictors of the reading item difficulty, including word length, total word counts, LSA, and connectives, explained roughly half amount of the variance in the reading item difficulty.

7 Conclusion and future research

This study shows the importance of seeking empirical evidence to justify iterative decision-making steps in test development. Future research on employing text analyzers in various testing and instructional contexts will lead us to a better understanding of the relation between text characteristics and reading ability.

8 Note

This document is based on an article published in Modern English Education in May 2016.

References


An Empirical Investigation of the Relationship between L2 Vocabulary Size and L2 Listening Comprehension at the (Pre-)intermediate Level in English and French.

Britta Kestemont
KU Leuven
britta.kestemont -at- kuleuven.be

Ann-Sophie Noreillie*
KU Leuven
annsophie.noreillie -at- kuleuven.be

Kris Heylen
KU Leuven
kris.heylen -at- kuleuven.be

Piet Desmet
KU Leuven
piet.desmet -at- kuleuven.be

Elke Peters
KU Leuven
elke.peters -at- kuleuven.be

*joint first author

1 Introduction

Vocabulary knowledge is a key predictor of language proficiency (Schmitt, 2008). Although previous research has shown a strong relationship between vocabulary size and reading comprehension (Laufer & Ravenhorst-Kalovski, 2010), little research has looked into the relationship between vocabulary knowledge and listening comprehension. One study that investigated this relationship is Staehr’s (2009) study, which explored the relationship between advanced English-as-a-foreign-language learners’ listening scores on the Cambridge Certificate of Proficiency in English (CPE), a standardized C2-test aligned to the Common European Framework of Reference (CEFR) and their scores on the Vocabulary Levels Test (VLT) (Schmitt, Schmitt, & Clapham, 2001). His results indicated a strong relationship between vocabulary size and listening comprehension (r = .70). Further analyses suggested that 5,000 word families, which corresponded to 98% lexical coverage, seemed to suffice for advanced listening comprehension. Other research found similar results. Van Zeeland and Schmitt (2012) for example found that 95% lexical coverage seems necessary for relatively good listening comprehension, whereas 98% would lead to optimal comprehension. However, it is not clear whether these results can be transferred as such to other proficiency levels and other foreign languages. Therefore, an “approximate replication” (Porte, 2012, p. 8) of Staehr’s study was conducted in order to investigate the relationship between L2 vocabulary knowledge and L2 listening comprehension at the B1-level in two foreign languages, English and French. By comparing vocabulary knowledge and listening in two foreign languages, we hope to refine our understanding of the role of vocabulary in listening.

2 Research design

The two research questions we will address are: (1) Is there a correlation between vocabulary size and listening comprehension at the (pre-)intermediate level, as measured by a standardized B1 test? (2) What vocabulary size is necessary for listening comprehension measured by a standardized B1 listening test?

We recruited 199 English-as-a-foreign-language learners and 351 French-as-a-foreign-language learners from the second and fourth year of secondary education and first-year university in Flanders (Belgium).

Learners’ vocabulary knowledge was measured by means of a frequency-based test developed in the same way for English and French. The frequencies for the English test were sampled from the Contemporary Corpus of American English (COCA) (Davies, 2008) and for French from the Frequency dictionary of French core vocabulary for learners (Lonsdale & Le Bras, 2009). The format was a meaning recognition multiple choice test consisting of 120 items with 30 items per frequency band (2K, 3K, 4K and 5K). The 30 items for 2K were taken from the first 2,000 words (13 1K words for English and 12 1K words for French). Four definitions and an “I don’t know” option were given for every word. The latter was provided to minimize guessing and overestimation.

Listening comprehension was tested by a standardized listening comprehension test. For English the Preliminary English Test (PET), corresponding to the B1 CEFR-level (Cambridge), was chosen and for French the Diplôme d’Étude en Langue Française (DELF), also corresponding to the B1 CEFR-level (Centre International d’Études Pédagogiques) was selected.

For the participants in secondary education, two sessions were organized. In the first session, they were administered the vocabulary test, while in the second session they took the listening comprehension test. The
university students first took the listening comprehension test and then the vocabulary test in one single session. The difference in procedures was due to time constraints in secondary education.

3 Results

As can be seen in Table 1, the reliability of the test instruments was good.

<table>
<thead>
<tr>
<th></th>
<th>English</th>
<th>French</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vocabulary size test</td>
<td>.98</td>
<td>.96</td>
</tr>
<tr>
<td>Listening comprehension test</td>
<td>.75</td>
<td>.70</td>
</tr>
</tbody>
</table>

Table 1. Reliability (Cronbach’s Alpha) of Vocabulary Size Test and Listening Comprehension Test.

The first research question can be answered affirmatively. Our results indicated a strong correlation between vocabulary size and listening comprehension in both languages: a Spearman’s Rank Order Correlation of .64 for English and .70 for French at the 0.01 significance level. In order to answer the second research question, the participants were divided into groups according to their mastery of a frequency band. Afterwards, the average score on the listening comprehension test per group was computed, as shown in Table 2. A Kruskal-Wallis analysis with follow-up Mann Whitney U’s tests showed that there was a significant difference in comprehension scores between the groups, more specifically for English between 0K and 1K, 1K and 2K and for French between 0K and 1K. The results seemed to level off from 2K onwards for English and 1K onwards for French. An analysis of the lexical profile of the listening comprehension tests with Lextutor (www.lextutor.ca/vp) revealed that the first 1,000 words covered 91.5% of the English listening test, whereas they only covered 86.24% of the French listening test. Interestingly, 91.5% lexical coverage leads to a good comprehension score for English and to a relatively good score for French (Table 2).

<table>
<thead>
<tr>
<th></th>
<th>Listening comprehension</th>
<th>Lexical coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>English (N=65)</td>
<td>French (N=232)</td>
</tr>
<tr>
<td>0K</td>
<td>75.69%</td>
<td>67.86%</td>
</tr>
<tr>
<td>1K</td>
<td>86.83% (N=24)</td>
<td>75.74% (N=68)</td>
</tr>
<tr>
<td>2K</td>
<td>91.23% (N=44)</td>
<td>73.67% (N=24)</td>
</tr>
<tr>
<td>3K</td>
<td>91.20% (N=35)</td>
<td>75.20% (N=5)</td>
</tr>
<tr>
<td>4K</td>
<td>91.60% (N=10)</td>
<td>68% (N=1)</td>
</tr>
<tr>
<td>5K</td>
<td>92% (N=21)</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Average Score on Listening Comprehension and Lexical Coverage per Frequency Band.

4 Conclusion

Our results are in line with previous listening research, as we also found a strong correlation between vocabulary size and listening comprehension. Moreover, our findings seem to indicate that knowledge of the most frequent 1,000 words might be sufficient for a good score on these standardized listening tests at a (pre-) intermediate level. Interestingly, we seem to observe a threshold effect from 2K onwards for English and from 1K onwards for French. However, further research is necessary to strengthen this claim.

References


The Effects of Narrow Reading/Listening on Vocabulary Learning

Anna C-S Chang
Hsing Wu University
annachang -at- livemail.tw

1 Introduction

Narrow reading refers to readers focusing on the work of a single author or a single topic over the course of a number of texts for an extended period of time. The texts in narrow reading are more related to each other, which means that learners do not have to switch from one topic to another. There are several advantages in using narrow reading in L2 learning. From the point of view of background knowledge, topical familiarity has been found to be helpful to comprehension. If learners are exposed to the same topic, they will become familiar with the same topic and will have better background knowledge for the input. From the perspective of frequency of word occurrence, research has demonstrated that learners need multiple exposures to acquire a word (e.g. Webb, 2007); therefore, in narrow reading, key words and proper nouns in the relevant topics will recur, so learners will have multiple exposures to the same words and are more likely to become familiar with those words and thus ease the lexical burden (Hwang & Nation, 1989; Schmitt & Carter, 2000). From the point of view of processing input, if learners possess better background knowledge of a topic, it may allow them to have more attentional resources for processing other tasks, such as unknown words or phrases.

2 Review of literature

There have been some corpus-driven studies on narrow reading. The findings can be summarized as below:

- Running stories in the newspapers have more repetitions of words than random stories, especially for the words beyond the second 1,000 level (Hwang & Nation, 1989).
- Vocabulary load in the related academic texts contains a much smaller number of word types and word families than unrelated texts (Sutarsyah, Nation, and Kennedy, 1994).
- Related newspaper stories contain many fewer word types and the words are repeated more often than those from randomly selected texts (Schmitt & Carter, 2000).

Some empirical studies on narrow reading also show that L2 learners acquire more vocabulary knowledge from texts of the same series. For example, Cho and Krashen asked four L2 adults to read the Sweet Valley series for pleasure in their free time, and their vocabulary pick up rates were assessed through individual tests. The average correct percentage reported was 62%. Similar positive results were also demonstrated in their later study with young L2 Korean elementary pupils (Cho, Ahn, & Krashen, 2005), who were asked to read the Clifford books series. More recently, Kang (2015) asked 30 senior high school students to read four online texts — either thematically related articles on second-hand smoking or unrelated articles over one month. The students were tested on their gains of receptive and productive knowledge. The results show that narrow reading leads to better gains than random reading in both receptive and productive knowledge. Though the advantages of narrow reading are apparent, there has been a dearth of studies on this area.

3. The Study

The participants were 31 secondary students, aged between 18-19 years old. Study materials were two sets of graded readers, three on the same title: The Railway Children, published by Penguin, Compass, and Oxford Bookworm, the other three were on the same genre: Sherlock Holmes and the Duke’s Son, Sherlock Holmes Short Stories, and The Last Sherlock Holmes Story. They are Level-1, Level-2, and Level-3 respectively. The texts were analyzed by using BNC/COCA (Nation, 2012). The lexical profile of the study materials was presented in Table 1. Thirty target words and 10 easy words as motivators were chosen from each set of the study materials. Among the 30 words selected from The Railway Children, the frequency of word occurrence ranged from 1 to 56 and 18/30 of the target words distributed in one book, 8/30 in two books and 4/30 in three books. The frequency of word occurrence of the 30 words from the Sherlock Holmes stories ranged from 1 to 27 and 27/30 words distributed in one book, and 3/30 in two books. The dependent measures involves four aspects: spelling, aural meaning, written meaning and sentence-making.
Table 1: The Lexical Profile of the Two Sets of Study Materials.

<table>
<thead>
<tr>
<th></th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>The Railway Children</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,000</td>
<td>90.38</td>
<td>88.78</td>
<td>89.41</td>
</tr>
<tr>
<td>2,000</td>
<td>4.15</td>
<td>5.32</td>
<td>4.81</td>
</tr>
<tr>
<td>Proper Nouns</td>
<td>3.67</td>
<td>2.62</td>
<td>3.03</td>
</tr>
<tr>
<td>Word types</td>
<td>549</td>
<td>1029</td>
<td>1077</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1,507</td>
</tr>
<tr>
<td><strong>Sherlock Holmes stories</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,000</td>
<td>86.38</td>
<td>87.76</td>
<td>88.30</td>
</tr>
<tr>
<td>2,000</td>
<td>2.84</td>
<td>4.14</td>
<td>5.99</td>
</tr>
<tr>
<td>Proper Nouns</td>
<td>5.15</td>
<td>4.33</td>
<td>3.14</td>
</tr>
<tr>
<td>Word types</td>
<td>680</td>
<td>784</td>
<td>1214</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1,642</td>
</tr>
</tbody>
</table>

3 Results

The results show that most of the target words were unknown to the students. After reading three books either on the same genre or on the same title, apart from sentence-making, the students performed comparably (from 58% to 60%) in the post-tests across aspects of vocabulary knowledge gained. The correlations between frequency of word occurrence and learning gains showed that there were significant correlations between aural meaning ($r = .46$, $p < .01$)/written meaning ($r = .50$, $p < .01$) and frequency of word occurrence when reading *The Railway Children*. No significant correlations were found for the words selected from Sherlock Holmes stories.

4 Conclusion and limitations

Although the target words’ frequency of occurrence and distribution were higher in the *The Railway Children*, the students scored similarly across different aspects of vocabulary gained in the immediate post-tests. The reason might be that after studying the books, the students could remember most of the linguistic elements they just learned or came across. There might have been different results if there had been a delayed post-test. In the main study, it’s necessary to include a delayed post-test. There were some significant correlations between vocabulary gains and frequency of word occurrence for reading *The Railway Children* but there were no significant correlations found in reading stories about Sherlock Holmes. Whether the significant correlations would lead to a higher level of retention requires a delayed post-test to demonstrate it.

5 Acknowledgements

This research project is financially supported by a research grant gained by the author from the Ministry of Science and Technology (MOST 104-2410-H-266-004-MY2). The author would like to express thanks to abstract reviewers for spending their valuable time giving constructive comments.

References


Lexical Complexity in Japanese EFL Learners’ Spoken and Written Production

Mariko Nomura
Kanda University of International Studies
m.nomura311 -at- gmail.com

1 Introduction

This study investigates the lexical complexity of Japanese EFL learners’ spoken and written production. Lexical complexity is a multidimensional construct which can be divided into several dimensions: Bulté & Housen (2012) divide lexical complexity into density, diversity, compositionality, and sophistication. Previous studies on lexical complexity have focused exclusively on either spoken or written production. There has been little research on differences in lexical complexity between spoken and written discourse produced by the same learners (Yu, 2009; Nomura, 2013). The present study attempts to examine how different production modes (spoken vs. written) affect lexical complexity using Japanese secondary school students’ spoken and written discourse pairs.

2 Background

Nomura (2013) compared spoken and written texts produced by Japanese EFL learners using D (Meara & Miralpeix, 2007) as a measure of lexical diversity. The results showed that the learners’ written texts were more lexically diverse than their spoken texts and D proved to be valid for both modes of short texts produced by the same group of Japanese secondary school students. However, the study uses one lexical diversity measure and demonstrates only one aspect of lexical complexity. In order to capture the multidimensionality of lexical complexity and provide a more complete picture of the lexical complexity of the learners’ spoken and written texts, more measures of lexical complexity should be used.

The importance of using multiple measures has been stressed (Bulté & Housen, 2012) and recent studies related to lexical complexity have used multiple measures for three dimensions of lexical complexity: lexical density, diversity and sophistication (Lu, 2012; Lahmann et al., 2016; Thériault, 2015). For instance, Lu (2012) examined 26 different measures for the three dimensions to analyze large-scale data of ESL learners’ oral narratives, using a computational system, Lexical Complexity Analyzer, and the vocd utility of the Computerized Language Analysis program. The study showed that the three dimensions did not correlate strongly with each other, suggesting that they are indeed different constructs.

Based on the recent investigations of lexical complexity, five measures were selected in the present study. For lexical density (LD), the ratio of the number of lexical words to the total number of words was used. For lexical diversity, two measures in addition to D were selected: the number of types per family and the number of different families belonging to the 1000 most frequent word families (K1) in English. For lexical sophistication, the English Vocabulary Profile (EVP) (Cambridge University Press, 2015) was selected to measure students’ use of more advanced words based on the CEFR levels of words.

3 Research Questions

The following two research questions were addressed in the present study.
RQ1: Do different production modes (spoken vs. written) affect lexical complexity in Japanese secondary school students’ production?
RQ2: Are the lexical complexity measures examined valid for distinguishing spoken and written texts produced by Japanese secondary school students at different proficiency levels?

4 Method

The data used in the present study were selected from Japanese Secondary Learners’ Spoken and Written (JSSLSW) corpora (Nomura, 2012) comprising 324 discourse pairs on three topics. Forty two paired data samples with more than 50 words in each text were selected from 156 data sets which had information on students’ English proficiency levels, three EIKEN grades (Grades 3, Pre-2 and 2). EIKEN is Japan’s most widely used English language testing program which measures English proficiency. As a result, 14 paired data samples from each EIKEN grade on two different topics were used. The two topics were “Countries (or regions in Japan) you would like to visit in the future” and “Impressive school events.” The tokens of spoken and written data totaled 4,115 and 5,019 English words respectively.
For the paired data, the values of lexical density (LD), the number of types per family (T/F) and the number of families belonging to K1 (K1-F) were calculated with the vocabprofile program (Web VP Classic adapted by T. Cobb). As for D, the values computed with D_Tools (Meara & Miralpeix, 2007) for the same data in Nomura (2013) were used. Lexical sophistication was analyzed using Text Inspector offered by Cambridge University Press, which classifies each word in a text into the lists of the six CEFR levels (A1-C2) and Unlisted.

In order to answer RQ1, a paired t-test for each lexical complexity measure was conducted for overall spoken and written data sets to examine the effect of different production modes. In order to answer RQ2, spoken and written texts from three learner groups defined by EIKEN grades were analyzed regarding the mean values of the five lexical complexity measures by conducting a two-way ANOVA.

5 Results

Table 1 shows the descriptive statistics of lexical density and diversity measures examined. Statistical analysis showed that the mean values of each measure were significantly different between the two modes. The learners’ written texts were more lexically dense and diverse than their spoken texts.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Spoken</th>
<th>Written</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>M (SD)</td>
</tr>
<tr>
<td>LD</td>
<td>42</td>
<td>0.48 (0.05)</td>
</tr>
<tr>
<td>T/F</td>
<td>42</td>
<td>1.08 (0.05)</td>
</tr>
<tr>
<td>K1-F</td>
<td>42</td>
<td>33.02 (9.74)</td>
</tr>
<tr>
<td>D</td>
<td>42</td>
<td>23.51 (9.37)</td>
</tr>
</tbody>
</table>

Table 1. Descriptive Statistics.

With respect to lexical sophistication, a significant difference was found in the A2 level of vocabulary use between spoken and written texts. The measures used in this study did not significantly distinguish Japanese secondary school students’ performance by proficiency level except for K1-F and the B1 vocabulary level. No interaction between mode and EIKEN grade was found in any of the measures.

References


The Relationship between Vocabulary Size, Recognition Rate, Automaticity, and Overall Proficiency in an On-line L2 Vocabulary Test

Yutaka Yamauchi  
Tokyo International University  
yyama -at- tiu.ac.jp

Kay Husky  
Tokyo International University  
kkarolin -at- tiu.ac.jp

Aki Kunikoshi  
McRoberts B.V.  
yemaozi88 -at- gmail.com

1 Purpose of this study

This study attempts to develop a new on-line vocabulary test which can measure size, rate and automaticity in L2 vocabulary recognition, and to clarify the relationship between vocabulary size, recognition rate, processing automaticity, and overall proficiency of L2 learners.

2 An online test measuring vocabulary size, recognition rate, and processing automaticity

The test developed in this study contains seven levels of vocabulary in terms of frequency: 500, 1000, 1500, 2000, 3000, 5000 and 10,000 based on the frequency levels of BNC (British National Corpus). The test includes six different words in each level and so the total number of words in the test is 42. The test system automatically chooses a word out of the 42-word stock, presents the selected word on a PC monitor and records the option the test taker has selected and the time needed to choose it. Thus this test can measure L2 vocabulary size and recognition time. Automaticity in vocabulary recognition can be assessed by coefficient of variation (CV). CV, calculated by variance of reaction time (RT) divided by mean RT, can indicate how learners' language processing is automatized (Segalowitz & Segalowitz, 1993). CV has been reported to decrease as learners' proficiency levels become higher because their language processing changes from controlled to more automatized, and so variance decreases (Segalowitz, Segalowitz, & Wood, 1998; Harrington, 2007; Lim & Godfroid, 2015).

3 An experiment to examine how vocabulary size, recognition rate, processing automaticity and overall proficiency of L2 learners are related

Eighty-four Japanese learners of English with proficiency levels ranging from 200 to 900 as measured by TOEIC were required to take the online test developed in this study. They were requested to read a word presented on a PC monitor and choose the most suitable meaning of the word out of four options as quickly and accurately as possible. The option they selected and the reaction time (RT) in choosing it (from the time the four options were presented on the screen) were automatically recorded by the computer. Based on the RT, the CV was also automatically calculated by dividing the variance of RT by the mean RT. Thus, the test measures vocabulary size, recognition time, and CV. The statistical results of the experiment show that the four variables (vocabulary size, recognition rate, processing automaticity and L2 overall proficiency) had significantly high correlation, and that this test is valid for measuring vocabulary size, recognition rate, and processing automaticity.

References


Data-Driven Learning of Words and Phrases
by Low Proficiency EFL Learners through WebSCoRE

Kiyomi Chuo
Nihon University
chuujou.kiyomi -at- nihon-u.ac.jp

Laurence Anthony
Waseda University
anthony -at- waseda.jp

Kathryn Oghigian
Waseda University
k_oghigian -at- aoni.waseda.jp

1 Introduction

Research has shown data-driven learning (DDL) to be effective for learning EFL grammar and vocabulary items (Cobb & Boulton, 2015; Mizumoto & Chuo, 2015). However, DDL is usually used with intermediate or advanced level learners because the truncated monolingual Key Word in Context (KWIC) format that is generally used to present search results can be too difficult for lower level learners to interpret. The purpose of this study was to field test the effectiveness of WebSCoRE, a free, parallel concordancer with a specially developed bilingual sentence corpus (Chuo, Oghigian, & Akasegawa, 2015), in a low proficiency Japanese EFL university course aimed at improving their usage of words and phrases.

2 SCoRE and WebSCoRE

The Sentence Corpus of Remedial English (SCoRE) corpus currently contains 22 grammar categories with 10,113 level-specific semi-authentic sentences which are specially written to satisfy particular pedagogical considerations and to be free of copyright issues. SCoRE comprises complete sentences which are easily understood by low proficiency EFL learners (see Figures 1 & 2). Each English example sentence is accompanied by a Japanese translation which has been manually translated and corrected. In this case study, participants used a simple, user-friendly WebSCoRE (Anthony, 2016) concordancer to access the SCoRE corpus and check word and phrase usage, form hypotheses, and then produce language to consolidate learning.

As can be seen in Figure 1, ten concordance lines in the target language (English) are shown with the target word *furniture* aligned in the center, and corresponding translations given below as a reference corpus. From previous studies (Chuo, Anthony, Oghigian, & Yokota, 2013), it was found that learners are able to quickly scan the translation to confirm the specific meaning of English words or sentences, or to grasp the general meaning of the sentence. Looking at the concordance lines of *furniture*, learners can find that it is an uncountable noun which has only one form (not *furnitures*), that it can be used with *some*, and that it is followed by singular verb forms such as *is* and *was*. Figure 2 shows the search results of *歯* and its English equivalents *tooth* and *teeth*. Japanese learners sometimes find it difficult to remember which form is singular and which is plural. Looking at English phrases with indefinite articles such as *a shark’s tooth*, and others without such as *brushing her teeth* provides clear examples.
3 2016 Study and Results

Learning effectiveness was measured using pre- and post-tests. Statistically significant gains in test scores suggested improvement in learner proficiency. Learners also gave feedback using a 42-item questionnaire, with results indicating that the corpus tool was useful and helpful for learning the usage of words and phrases. Example comments included “I was surprised to find such a massive quantity of examples easily by searching only one word at one click”; “It’s helpful because we can confirm the usage and meanings of the words which I’d like to investigate”, and “Since it’s a discovery learning what I found was easily fixed in my memory.” Moreover, the learners reported the WebSCoRE tool was easy and convenient to use and they liked its layout and speed. Most learners used it with their smartphones to do their homework. The learners further reported that the sentences provided clear, observable language patterns that were appropriate in terms of vocabulary level, sentence length, and structure. This study demonstrates that lower proficiency learners can take advantage of DDL when using a user-friendly corpus tool such as WebSCoRE with an educationally modified bilingual corpus such as SCoRE.

4 Acknowledgements

This research was funded by a Grant-in-aid for Scientific Research (25284108) from the Japan Society for the Promotion of Science and the Ministry of Education, Science, Sports and Culture.

References


Flashcards – They’re Back, but with One Important Question Unanswered

Tom Cobb
Université du Québec à Montréal
cobb.tom -at- uqam.ca

After the denigration of stimulus-response learning models in the Chomskyan era and then the communicative language classroom that followed, vocabulary flashcards (any review format with word forms on one side and some form of gloss on the other) are regaining credibility as an efficient and effective tool for aspects of vocabulary acquisition (Nation, 2013, Ch. 11). Current research investigates technologies and methodologies for word review using flashcards, including these: Paper card or computer screen (Nakata, 2011)? L1 or L2 gloss (Elgort, 2011)? Long or little delay between encounters (Mondria, 1994)? Less evident on the research agenda is the success of flashcard learning as a function of where the content of the cards comes from, learners themselves or an outside source. Learner creation of cards (e.g. keyword links) has appeared to involve time and intensive work that divorces the activity from the flashcard concept, but this might not be true if automation were extended to the card creation process itself. This presentation will develop a technology and experimental design for comparing retention, automaticity, and transfer for pre-programmed flashcards vs cards assembled by learners themselves through a corpus activity.

The technology will be a development of the ‘learner as lexicographer’ software originally developed for my PhD experiments, which has been my most most-cited work ever since (e.g., Cobb, 1999) and has slowly been grown from toy to comprehensive scale and desktop to Web/smartphone delivery over several years. The experiment’s design will also borrow from this work for retention and transfer but Elgort’s (2011) work for automaticity.

References


From Corpora to Word Lists: Issues and Limitations

Brent Culligan
Aoyama Gakuin Women’s Junior College
B.Culligan.Ed.D –at– gmail.com

1 Introduction

While there had been many vocabulary lists developed for pedagogy, The General Service List (West, 1953) was significant as the first word list to use frequency counts from corpus analysis to support their choices of words to include on the list. The process was reasonable. Collect a number of texts into a corpus, and count the occurrence of the head word and all its inflected or derived forms. At that time, the process was not feasible for small scale research given that the 2.5 to 5 million word corpus used in the GSL was laboriously compiled by hand (Gilner, 2011). Since then, the advent of cheap personal computers and their many paraphernalia have allowed researchers with limited means to develop list using the same process. One of the most notable was the Academic Word List (Coxhead, 2000) who introduced a component approach whereby the word list was compiled from those high frequency words remaining after the GSL was extracted. My colleagues and I have followed this approach in the creation of a number of word lists, including the New General Service List (Browne, Culligan, & Phillips, 2013), the New Academic Word List (Browne, Culligan, & Phillips, 2013), the TOEIC Service List (Culligan & Browne, 2016), and the Business Service List (Culligan & Browne, 2016). We have had to deal with a number of issues and technical difficulties during the development process. Our responses to those issues are informed by pedagogical fields such as curriculum development and assessment.

2 Pedagogical objectives

From a pedagogical point of view, the primary objective of a corpus analysis for vocabulary is to provide a list of words that the student will most likely meet in the future. One outcome of reliance on a word list derived from a corpus is that the question of what words to teach is no longer the purview of the teacher’s intuition. Rather than selecting material based on ill-defined criteria, materials need to be selected based on the occurrence of the words of the list or rewritten to include them. The teacher’s goal is to deliver materials that best prepare the student for future vocabulary encounters. A second outcome is that it is not the result of the frequency analysis to determine what meanings to teach. Frequency analysis can determine the probability a letter string will be met in the future. These probabilities will be affected by homonyms, homographs, and variations in spelling. Subsequent Key Word In Context (KWIC) analysis is needed to determine which senses of the word to teach. For example, the word BOARD will be met with a predictable frequency which determines that it should be taught. However, only KWIC analysis can determine what senses to teach.

3 Operational definitions of a word

One of the main issues that need to be examined prior to the analysis of a corpus is the operational definitions of the concept of a word. While West (1953) used a modified lemma approach in the GSL, Bauer and Nation (1993) introduced the concept of a word family based on the assumption that “once the base word or even a derived word is known, the recognition of the other members of the family requires little or no extra effort” (p. 253). However, this assumption seems to have mixed support at best. Culligan (2008) found a statically significant difference between the item difficulties of base words and Bauer and Nation’s Level 3 affixes (t = -14.981, df = 303, p < .01) suggesting that for low intermediate to intermediate learners, derived forms require more than a little effort to acquire and may not be the best model of a word for these students.

4 Technical difficulties

Other important issues in the design, compilation, and analysis of corpora towards the objective of creating a word list for research and pedagogical purposes are a reflection of the tools we have at our disposal. One issue is with periphrastic words like SO-CALLED, which was included in the original GSL (West, 1953, p. 467) and in the original AWL (Coxhead, 2000, p. 235) but is not found on two of the more common programs for corpus frequency analysis due to programming limitations. These limitations also exclude
words spelled with non-ANSI characters like CAFÉ and RÉSUMÉ. For words like CAFÉ, this is not a major problem as it can also be spelled without the e-acute, but it does necessitate another step in pre-analysis corpus cleaning. For RÉSUMÉ, however, this is not sufficient as the plain e refer to a completely different word, RESUME.

5 Estimating corpus consistency

The final issue is that the interpretation of the results of an analysis is conditioned on the type of corpus, whether it is an open or closed corpus. A closed corpus is one in which, barring new revelations, is complete and will not have additional text added to it. For example, a complete book such as Anne of Green Gables is closed. We know with complete certainty all the words that the student will meet in the text. An example of an open corpus would be a newspaper or TOEIC corpus. With each new issue, more text will be added. In this case, the analysis of corpora gives us information about the frequency of a word’s past appearance and, as such, an estimate of the probability that the word will be met. Given that our predictions of future occurrence are based on an analysis of past occurrences, how can we make that prediction better? Drawing on parallels with the field of assessment, is the validity of our analysis limited by the consistency of the corpus in much the same way as the validity of our interpretation of test scores is limited in the reliability (i.e., consistency) of the testing condition?

There are various methods to determine the consistency of a corpus. One avenue has been to compare and correlate the rank frequency of the words in each corpus. Another method is based on the chi-squared statistic (Kilgarriff, 1997; Rayson & Garside, 2000) for frequency profiling to reveal keywords specific to a text when compared with another text. A third approach is stylometry which is the field of quantitative analysis of different texts to determine authorship (Eder, Kestemont, & Rybicki, 2016). The procedure to determine the relationship between consistency and predictability is to split the corpora into various parts, determine the consistency of the difference parts, and extract a word list from the split-half corpus. This word list will be used to calculate the coverage of the remaining split-half corpus. By comparing the consistency of a corpus to the coverage of its resultant wordlist when applied to its other part, this will determine if a consistency index is a worthwhile and useful statistic to report in a corpus analysis for vocabulary extraction.

References

How Much Vocabulary do Learners in Successful EMI Classroom Produce?  
A Case Study for Pedagogical Implications for Future EMI Classes

Masaki Eguchi
Graduate School of Education, Waseda University
egumasa-0101-at-fuji.waseda.jp

1 Introduction

This case study investigates lexical profiles of a group of learners during group discussions in an English Medium Instruction (EMI), defined as "non-language subjects taught through English" (Hellekjær, 2010, p. 11). Although attempts have been made to provide sets of receptive vocabulary coverage as realistic learning goals in various target situations (van Zeeland & Schmitt, 2012; Webb & Rodgers, 2009a, 2009b), few empirical studies have reported productive vocabulary that is indispensable for content-oriented instructions in the L2 (for a similar argument see Schmitt & Schmitt, 2012). Despite the fact that the amount of productive vocabulary is extremely difficult to measure in relation to the types or complexity of the tasks (Milton, 2009, p. 130), profiling lexical items used in an actual content-based classroom will benefit teachers and learners in EFL environment in that it will serve as a guide for the instruction in the L2 English. The current study therefore addresses two research questions (RQ) as follows. 1) What do the vocabulary profiles of a group of learners in successful discussions look like? 2) What kind of vocabulary do they have difficulty in producing?

2 Methodology

To address the RQs, an EMI class was selected because the students and the instructor did not resort to their shared L1 Japanese during the class even when the content of the materials was demanding. Each week, students were divided into four groups of four or five. Two students' presentations were given on assigned chapters from introductory textbooks on Second Language Acquisition (SLA) in L2 English, providing several discussion questions on the topics covered in the chapter (two or three questions in each presentation). The types of questions include: (1) definitions of technical terms (e.g., working memory, noticing hypothesis), (2) comparisons and contrasts of theories in SLA (e.g., differences between functionalism and innatist perspective in terms of language acquisition), (3) pedagogical implications and applications to English language teaching in Japan, for example.

In the current study, the group discussions were audio recorded (seven classroom times, 13 groups in total). A total of 18 upper-intermediate-to-advanced students were involved in the recordings, but the amount of recordings per participants varied (once to six times) due to a mixed constraint of forming discussion groups and availability of recording devices at a time. After audio recordings had been collected, transcriptions were performed by the author. Trivial errors in pronunciation and morphology were corrected to reduce the amount of Off-list words. For the same reason, unlexicalized pauses and back channellings were subsequently deleted for further analyses. Finally, the resulting transcriptions (17,842 tokens) were analyzed with AntWordProfiler (Anthony, 2014) based on BNC-COCA lists (Nation, 2012). To address RQ2, the current study follows Hilton (2008), which regards breakdowns before content words as lexical retrieval problems. The lexical items with relatively longer pauses (i.e., longer than one second) were marked with Praat (Boersma & Weenink, 2013), then coded into the transcriptions. Afterwards, those words were examined in terms of frequency and word class.

3 Result

The result of the vocabulary profiling (Table 1) shows that 90.49% of total words (tokens) is composed of the first 2000 words in BNC-COCA plus OFFLISTS. Moreover, the coverage reaches 94.64% when third 1000-word level is taken into account. Another important finding is that total number of different words produced in the discussions is 1548 words in type count.

As for RQ2, additional analysis (i.e., pause measurement) reveals the tendency of the learners to pause before (a) frequent adjectives and verbs such as "difficult," "remember," "say," etc. and (b) words in Academic Word Lists (AWL; Coxhead, 2000), such as "valid," "emphasize," etc.

4 Discussion

As indicated above, 94.64% of the current discourse consists of first 3000 words. It is noteworthy that the
current transcriptions were from the academic discussions in an EMI class rather than the daily conversation that previous studies targeted at (Adolphs & Schmitt, 2003). This confirms the significant roles of high frequent words even in the current academic discourse among non-native speakers. While few previous studies examined vocabulary profiling of academic discussion in actual classrooms, this study provides an empirical data that as for production learners do not necessarily have to produce a large amount of words across different frequency levels.

Regarding RQ2, relatively longer pauses before frequent adjectives and verbs might be due to the slow speed of retrieving less frequent, appropriate lexical items. This might entail problem-solving strategies that learners employ, where they were able to compensate for the insufficient knowledge with more general, familiar words to keep the discussion flowing at the expense of the infrequent, appropriate expressions (Segalowitz, 2010, p. 41).

Although the present study suggests important findings, there are limitations to overcome in future studies. First, it was suggested that a larger amount of transcriptions is necessary to more reliable results based on the notion of lexical saturation (Miller & Biber, 2015). Second, discourse analysis should be conducted to gauge the content of the discussion in more detailed manner. These issues should be addressed in further research.

5 Conclusion

The current study collected a total of 17,842 tokens of production from a series of academic discussions through an EMI class in a Japanese university, and found that the relatively small number of words (1547 different types) enabled the learners to function in the EMI discourse. Since the findings might be influenced by the small amount of production examined in the study, further studies should confirm these findings with more sophisticated methodology.

References


Classroom Fluency Activity for Vocabulary Development
Doreen Ewert
University of San Francisco
dewert -at- usfca.edu

1 Rationale

Many international students arrive abroad with limited productive ability in the language medium of the universities they are attending. In spite of achieving the standardized test scores required for admission, students often claim their problems are insufficient vocabulary and poor grammar even when they have previously engaged in long periods of classroom language study. What students actually seem to lack is fluency in spontaneous use of the target language. For this reason, more accuracy-focused classroom activity is not the best prescription for resolving the problem, and this may be particularly true in regard to vocabulary learning. Although there is definitely a role for explicit instruction and intentional learning in vocabulary development, that alone without fluency activity will be inadequate for the breadth and depth of vocabulary knowledge and speed of acquisition desired by most students and expected by faculty in university courses. To date, however, systematic attention to fluency-building of any of the skills is limited if at all evident in foreign or second language classrooms, textbooks, or in methods courses for language teachers.

Nation (2000, 2011) claims that fluency building activity should make up 25 percent of any language curriculum. Empirical research that directly links vocabulary acquisition to fluency activity is limited and fraught with the dilemma of how to control for all the variables inherent in measuring and establishing causal or strong correlative explanations for implicit learning. Nonetheless, because research and theory of language learning in general support attention to fluency development apart from accuracy-focused activity (Dapke, 2013; Ewert, 2011; Nation, 2000; Robb and Kano, 2014; Waring & Takaki, 2003; Wood, 2001), it behoves language instructors to include fluency activity for vocabulary development.

2 Classroom Practices

A course on vocabulary development for matriculated learners of academic English at a university in the US was developed based on the application of implicit and explicit second language learning principles (Ellis, 2005). For implicit learning, input frequency was achieved through extensive reading and content-based classroom activities such as 4-3-2 (Nation, 2000), and timed fluency writing and reading. For explicit learning, attention to form was achieved through self-made vocabulary cards, pair and group activity with the cards, content-related textbook vocabulary exercises (Schmitt, Schmitt & Mann, 2011), self-investigation of vocabulary depth and breadth using publically available word lists (NGSL, Browne, C. et al.; NAWL, Browne, C. et al.), and vocabulary measures (see http://lex.tutor.ca). In addition, students were given regular individualized quizzes based on their vocabulary cards. This was made possible by the use of a quiz template and student-submitted wordlists on spreadsheets. Throughout the course, students were introduced to research-based practices for both implicit and explicit learning so that they would be able to make informed choices about what and how to engage in further development of their vocabulary after the course ended.

3 Implications

Although only interesting correlations can be reported on the effectiveness of the activities themselves, students who applied themselves reported being motivated by the agency they had in determining what to study and in seeing how their vocabularies developed in the course of one semester.

References


Cobb, T. Complete Lexical Tutor v.8 [computer program]. Accessed 22 July 2016 at http://lex.tutor.ca/


Comparing Near-synonymous Attributive Verbs in Academic Written Texts

Hsueh-chao Marcella Hu
The Overseas Chinese University
hhcm -at- ocu.edu.tw

1 Introduction

In terms of collocation learning, near-synonyms particularly pose difficulties for most foreign language learners due to their similar meanings but un-interchangeable usage (e.g., strong tea vs. powerful car). In the context of an academic written discourse, in particular, the researcher has often to report, evaluate, and discuss (previous) research findings with attributive verbs. The language of attribution, according to Hunston (1995), can be used to “hedge a statement, to introduce information which corroborates the writer’s own view, to indicate gap in research, or to set up a point of view against which the writer wishes to argue” (p. 133-134). It is a great misinterpretation to say that an author ‘argues’ when what s/he is actually doing is ‘report’. The difficulty with using these verbs is that there are many of them, and each of them has a slightly different, and often subtle shade of meaning (e.g., point out vs. note). It has been suggested that the semantic/usage differences among synonyms are best identified by their typical distributional patterns (especially their typical collocational patterns) and that corpus analysis of such distribution information can often effectively differentiate synonyms (Liu, 2013). The present corpus-based study aims to compare and contrast the function and usage of three attributive verb pairs with similar meanings (i.e., accept vs. acknowledge, mention vs. point out, answer vs. respond) by examining the academic texts in the field of Applied Linguistics.

2 Method

The attributive verbs were chosen based on a list from the writing center at the University of Texas (http://wuc.utexas.edu/wp-content/handouts/Verbs-of-Attribution.pdf). They were also checked against in the Merriam-Webster Online Dictionary of Thesaurus (http://www.merriam-webster.com/thesaurus/) and Collins Online Dictionary (http://www.collinsdictionary.com/english-thesaurus/) to ensure they share similar meanings. The self-created corpus consists of approximate 507,510 words derived from 87 pieces of peer-reviewed academic papers in two well-renowned journals (i.e., Language Teaching Research, Language Learning). The journal articles were chosen on the basis of research with original data during the last seven years (2009-2015). While Biber and his colleagues (1999) decided to include combinations which occurred over 10 times in a million words and appeared in 5 or more texts, the researcher in this study used a more conservative cut-off point of 10 times per half million words and include those occurring in 5 of the texts. At least 5 examples of a specific verb within a co-text of no less than 20 words were used to analyze the data before a conclusion was drawn.

3 Results

This study compares three pairs of near-synonymous attributive verb pairs: accept vs. acknowledge, mention vs. point out, and answer vs. respond. The following sections discuss the three pairs respectively.

Accept vs. Acknowledge

Both verbs take human (first person pronoun included) and non-human entities as subject but only ACKNOWLEDGE collocates with nouns that are authorities, as shown in the following examples.

1. Despite this high chunk strength, learners correctly accepted them only 22% of the time.

2. As Taguchi et al. (2004) acknowledged, the pretest and posttest forms were not equivalent, so the pretest and posttest reading rates were not comparable.

From the examples above, it can also be found that when the subject position is human entities, the object position of ACCEPT is nouns or pronouns. On the other hand, ACKNOWLEDGE either functions as an intransitive verb in example 2 or takes that-clause as complement as in Example 3, when the subject refers to authorities.

3. The findings in the current study are encouraging; nevertheless, we acknowledge that there are limitations to our research design.

It should be noted that ACKNOWLEDGE does function as a transitive verb when its subject is not authorities.
Mention vs. Point out

Both can take human and non-human subjects and may occasionally have first person pronoun as subject. POINT OUT usually collocates with authorities with the subject position as humans, whereas MENTION does not, as shown in the following examples.

1. However, as Carrell (1998) points out, strategy instruction and its effects could be context-dependent.
2. researchers have mentioned the use of “I don’t know” as a formula (Butterworth & Hatch, 1978; Shapira, 1978).

MENTION is used to report general information, whereas POINT OUT is used to report something that is more specific or relevant to the study. Besides, both verbs exhibit a formulaic expression in common: As + V. Yet an examination of their co-text reveals that this shared structure functions differently, as shown in the following examples.

3. As mentioned above, Tinkham (1997) suggests that while semantic clustering has a negative impact on learning, thematic clustering has a facilitative effect.
4. As pointed out by Moerman (1988), though, this work relies, to a certain degree, on the analysts’ and readers’ membership knowledge.

Answer vs. Respond

ANSWER is mostly used as a transitive verb while RESPOND is usually used as an intransitive one. The subject position for the two verbs is similar, with human entities such as participants or students. Differences; however, were found in their object position. ANSWER almost always collocates with questions or items in the objection position while RESPOND (to) collocates with various forms of situations, as shown in the following examples.

1. For instance, learners would need to correctly answer at least 45-46 of the 48 items that make up the 1000 and 2000 word-frequency levels on the LVLT to comprehend aural texts made up of the first 2000 high….
2. It is difficult to know why these two top-down listeners responded to the course in such contrasting fashion.

It was also found that ANSWER collocates with adverbs either suggesting (in)correctness or degrees of fulfillment while RESPOND can be modified with various kinds of adverbs. This difference could be attributed to the types of nouns they collocate with.

RESPOND also seems to be used more frequently in reaction to previous literature than ANSWER does. Among the concordance lines being examined, there is only one instance for ANSWER but there are three for RESPOND. In the context where the writers used their own research in response to earlier literature, the two verbs seem interchangeable.

4 Conclusion

In sum, while the three verb pairs share similar denotative meanings they still function differently in the corpus. ACCEPT is used transitively while ACKNOWLEDGE can be used either transitively or intransitively and can take that-clause as a complement. MENTION is used only to refer to something aforementioned, whereas POINT OUT can be used to call attention to both old and new information. ANSWER and RESPOND exhibit both similarities and differences in terms of their object position and the adverbs that they collocate with. It appears that RESPOND is a generic verb while ANSWER is more specific. The above findings may offer some insights into the usage of near-synonymous attributive verbs in the context of Applied Linguistics.

5 Acknowledgements

This study is supported by a grant from the Ministry of Sciences and Technology in Taiwan (MOST 104-2410-H-240-006).

Selected References


Probes into Semantic Clustering

Tomoko Ishii
Meiji Gakuin University
ishii-at-gen.meijigakuin.ac.jp

1 Introduction

It has been argued by vocabulary researchers that semantically related words should not be taught simultaneously because they can interfere with each other (e.g. Tinkham 1993, Folse, 2004). However, the question as to what specific types of relatedness cause interference has rarely been examined. Moreover, a consensus has not yet formed on this issue, with some studies arguing that semantic clustering does not cause interference and confusion (e.g. Papathanasiou, 2009). In addition, studies reporting a negative impact of semantic clustering vary greatly in their target words and research design, with some studies conducted under strictly controlled experimental conditions and others situated in a more natural classroom setting. Given the equivocal state of the current research on this topic, this paper attempts to probe the issue of semantic clustering by focusing on the source of confusion when learning several vocabulary items simultaneously.

Turning to studies on working memory, Ishii (2015) points out that psychologists have long seen visual stimulus as an important component of information processing (e.g. Baddely and Hitch, 1974). Researchers of vocabulary learning have also witnessed some evidence that learners do resort to visual imagery when trying to remember new words (e.g. Laufer 1990; Boers et. al. 2009). Drawing upon this psychological and applied linguistic research, in this paper, the author adopted a working hypothesis that the alleged impeding effect of semantic clustering is at least partly due to the visual features sometimes shared among the referents of semantically grouped words. For instance, ‘fruit’ is one common semantic category employed in the literature, but many of the fruit selected have been round. Similarly, ‘animals’ are often four legged, and clothes such as jacket, shirt, and coat also share some physical features. Of course, being grouped under a semantic category does not necessarily yield physical commonality among the referents; however, a review of the literature reveals that many semantically grouped items happen to share physical features. It is therefore reasonable to propose that an observed negative impact of learning semantically grouped words claimed in the literature may at least partly be due to those studies having targeted words with physically similar referents. This paper presents two studies that examined this working hypothesis.

2 Studies

The first study compared the learning of (1) unrelated, (2) semantically related (but physically dissimilar), and (3) physically related (but semantically unrelated) sets of words in order to examine if a negative impact of semantic clustering is observed when the cluster was controlled for the physical features of their referents. Additionally, it was of interest to see if a semantically unrelated but physically related word set (e.g. globe, watermelon, and ball) would have any influence on the learning of those words. The results confirmed that physically related sets were harder to learn than the other two sets, both on immediate and delayed post tests (Ishii, 2015).

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unrelated</td>
<td>13.88</td>
<td>3.56</td>
<td>5.16</td>
<td>3.48</td>
</tr>
<tr>
<td>Semantically related</td>
<td>14.33</td>
<td>3.64</td>
<td>5.05</td>
<td>3.52</td>
</tr>
<tr>
<td>Physically related</td>
<td>12.68</td>
<td>3.89</td>
<td>4.06</td>
<td>3.37</td>
</tr>
</tbody>
</table>

Table 1: Results from the First Study

Following these findings, the second study examined the learning efficacy of unrelated and semantically related sets of abstract nouns, where no visual image is readily available to the learners. The observed scores of these two conditions were very close, and statistical analysis showed no significant difference in the learning of the two sets. This indicates that no negative impact of semantic clustering is observed when there is no readily available concrete image that can be attached to the words to be learned.
<table>
<thead>
<tr>
<th></th>
<th>Immediate post test</th>
<th>Delayed post test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unrelated</td>
<td>13.00</td>
<td>3.48</td>
</tr>
<tr>
<td>Semantically related</td>
<td>13.42</td>
<td>3.85</td>
</tr>
</tbody>
</table>

Table 2: Results from the Second Study (abstract nouns)

3 Conclusion

Both theories of working memory and studies on learning vocabulary suggest that learners make use of visual information when trying to memorize new words. When the repeated claim of a negative influence of semantic clustering is examined through this point of view, it raises the question as to the specific causes of any impeding effect and whether perhaps physical appearance and semantic clustering may have been conflated in the particular word sets used in the previous research. The two studies presented in this paper both supported the hypothesis that the claimed negative effect can at least partly be explained by shared visual features among the referents of semantically grouped words.

Despite the equivocal nature of the literature, the impact that Tinkham (1993, 1997) has left in the field has led to evidence supporting the use of semantic clustering often being dismissed. The interference among semantically clustered words, then, is sometimes discussed almost as if it were an established fact. However, the literature does include studies showing that semantic grouping can help vocabulary learning under certain conditions, and this paper brings to light a potential cause of the negative relationship observed in previous studies. Although more research is needed before a consensus can be reached as to the effect of semantic clustering on vocabulary learning, the two studies presented in this paper show that relatedness in visual images is one confounding factor that can produce a seemingly negative effect of semantic clustering and which must be accounted for in future research.

References


Laufer, B. (1990). Why are some words more difficult than others?- Some intralexical factors that affect the learning of words. International Review of Applied Linguistics in Language Teaching, 28(4), 293–308.


Indirect Spaced Repetition Software (ISRS)

Louis Lafleur
Department of Foreign Languages, Okayama University
louislafleur333 –at– gmail.com

1 Abstract

The main goal of this research was to conceptualize, design, develop and test a prototype of a spaced repetition software titled “Indirect Spaced Repetition Software” (ISRS) by the author. In comparison to other spaced repetition software (SRS), additional tasks/questions for a same word item are not incorporated as additional flashcards (which could possibly cause overlapping and subsequently reduce the “spacing effect”) but under the same original digital flashcard which cycles task/question focus between meaning, form and use according to the reached interval. For this study, ISRS was specifically customized to meet Japanese Junior high school (JHS) student curricular needs. Therefore, the integration of an approved MEXT textbook “New Horizon 1” corpus, which comprises 800 word items, was selected.

2 Introduction

One of the most important contributions to the science of memorization was enabled through the pioneering research in the field of cognitive psychology by Hermann Ebbinghaus in the late 1800’s. His many discoveries including the “spacing effect” which accounts for why learners have better memory retention when they engage in spaced learning (multiple but short study sessions) compared to when they engage in massed learning (a single or very few long study sessions) were ground-breaking. The discovery of the “spacing effect” has now been substantiated by more recent discoveries in neuroscience have confirmed that spacing repetitions at appropriate intervals allow for enough time for the neuro-chemical regeneration to occur which is necessary for establishing and strengthening brain connections (Baddeley, 1990). Ebbinghaus’ discoveries permitted Leitner (1972) to elaborate his box/compartment study system enable learners to take full advantage of the “spacing effect” should the learner have good study habits when utilizing flashcards. In more recent years, Leitner’s system would often become the inspiration or basis for current spaced repetition software according to Godwin-Jones (2010).

A great influx of interest and research efforts in the field of vocabulary acquisition, cognition and also mechanical engineering have provided leads which could potentially benefit SRS. These following ideas were used as the basis for the author to systemize “Indirect Spaced Repetition Software”:

- Nation (2001) has suggested an “in-depth” learning of vocabulary to assure the correct “use” of the words in terms of output and points to the importance of mastering all three elements/features of word knowledge: 1-Meaning 2-Form 3-Use.
- Nation (2001) has also recommended a balanced teaching approach between the “four strands of vocabulary teaching”: 1-Meaning-focused input, 2-Meaning-focused output, 3-Language-focused learning, and 4-Fluency development.
- Schmitt (2008) implied that different teaching approaches may be more beneficial for different stages of word (lexical feature) knowledge as some are mastered earlier than others; therefore focusing on the form-meaning link at first and later enhancing contextual (use) may prove to be effective.
- Integrating a dynamic systems approach based in the world of mechanical engineering, namely, CVT (continuously variable transmission) to allow the software itself to customize interval times on-the-fly through automated data analysis (success rates per item/interval/question types, and optionally response speed) to better meet individual learner’s needs.
- The National Reading Panel (2000) recognized (guided oral) reading activities as being an effective way to develop fluency. Such an activity could easily follow the completion of a flashcard set and display the specific dialog or text where the word items were taken from.

3 Method

At the early developmental stages of ISRS, the author enrolled seven highly digitally fluent volunteer university students majoring in Education or Psychology at Okayama University (students from these departments were judged as being the most qualified to understand either the English curriculum at JHS
and/or the science of memorization), most of whom were in-training to become JHS English teachers. A pre-project questionnaire, which assessed the participants’ previous experience with e-learning software and future teaching ambitions, was given before the start of the two-week trial. Then, participants were given login names and passwords with the directive to use the online software for a period of two weeks around five to ten minutes a day. A post-project questionnaire comprising Likert-scale questions paired with follow-up open-end qualitative questioning was completed to evaluate the software under three categories (enjoyment, usefulness, and usability) at the end of the two-week trial.

4 Results

Although no quantitative data of statistical significance was found, the participants gave lots of constructive feedback such as the need to incorporate (1) some gamification features, (2) more challenging word lists for proficient users, and (3) additional user performance feedback.

5 Discussion and Conclusions

Although the groundwork for ISRS has been established with this study, further research will be required to assess differences in learner acquisition between traditional SRS and ISRS. It is the author’s wish that by releasing ISRS to the public will increase interest and further research/debate about SRS concepts and algorithms. A current project for the author is the creation of a public domain standardized JHS corpus. Future projects include incorporating additional wordlists such as the “New General Service List” (NGSL) and further ISRS testing. See Figures 1 and 2 for an overview of ISRS.

Figures 1 and 2. Indirect Spaced Repetition System/Software (ISRS)

Indirect Spaced Repetition System/Software (ISRS)

Table 1: Example Variable Questionnaire (R- Studying, L-Term, and I-Verbal, with two variable terms: i) 2nd tier (mid-term memory) 3 days (M-Mayo) ii) 3rd tier (long-term memory) 8 days (R-14 days)

![Figures 1 and 2. Indirect Spaced Repetition System/Software (ISRS)](image)

**Table 2: Examples of possible questionnaire tasks for each word/element**

<table>
<thead>
<tr>
<th>Level</th>
<th>Flow</th>
<th>Task Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>90%</td>
<td>ENG-Eng to JPN-Eng</td>
<td>Multiple Choice</td>
</tr>
<tr>
<td>60%</td>
<td>JPN-word to Eng-word</td>
<td>Multiple Choice</td>
</tr>
<tr>
<td>60%</td>
<td>ENG-word to JPN-eng</td>
<td>Spelling/Writing</td>
</tr>
<tr>
<td>40%</td>
<td>JPN-Chinese to Eng-word</td>
<td>Multiple Choice</td>
</tr>
<tr>
<td>40%</td>
<td>JPN-English to JPN-English</td>
<td>Multiple Choice</td>
</tr>
<tr>
<td>40%</td>
<td>JPN-English to JPN-English</td>
<td>Writing</td>
</tr>
<tr>
<td>40%</td>
<td>JPN-Chinese to JPN-English</td>
<td>Writing</td>
</tr>
<tr>
<td>40%</td>
<td>1-2 Test</td>
<td>Hotel Reservation</td>
</tr>
</tbody>
</table>

**ENG** - English, **JPN** - Japanese.
* Given names and range of memory terms are used here for the sake of presentation. Memory spans of a continuous nature and naturally should not be artificially segmented. ** Specific intervals may also be also be customized on the fly automatically through the collection of reaction and data. ** Setting an answer time threshold within ISRS is optional. However, some developers may choose to set specific times for questions or use received data to automatically modify internal times. (c) For increased effectiveness, ISRS can be combined with a fluency building (vocabulary) reading activity. For example, studying a set of vocabulary cards can boost trigger such a task.

References


Browne, Charles, Brent Culligan, and Joseph Phillips. *New General Service List*. Licensed under a Creative Commons Attribution 3.0 Unported License. Permissions beyond the scope of this license may be available at [http://www.newgeneralservicelist.org](http://www.newgeneralservicelist.org).


How Much of the Vocabulary Size Test is Necessary?

Barrie Matte
Kansai University
mattebarrie -at- gmail.com

1 Introduction

The Vocabulary Size Test (VST) (Nation & Beglar, 2007) is a multiple-choice test used to measure written receptive English vocabulary size (Nation, 2012). Several versions of the VST exist, including a 140-item test measuring the first 14,000 frequent words (Nation & Beglar, 2007), and another with 100 items, measuring the first 20,000 frequent words (Coxhead, Nation, & Sim, 2014), including numerous bilingual versions.

The VST provides instructions that test takers are to take all items on the test (Nation, 2012), which has been supported in several studies (i.e., Nguyen & Nation, 2011; Karami, 2012; Nation, 2012; Coxhead et al., 2014).

In opposition to these instructions, several studies have suggested that taking all levels of the VST is inefficient and may result in increased anxiety amongst test takers (Beglar, 2010; Elgort, 2013; McLean et al., 2015). Rather, it has been proposed that test takers only take up to 2 levels of the VST above their estimated vocabulary size; however, no empirical validation of this has been done.

This study was done in order to provide such empirical validation, and to attempt to show a possible amount that may be appropriate for test takers instead.

The research questions of the study were as follows:

1) How much of the test is necessary in order for it to sufficiently measure test takers?
2) How much of a test taker’s vocabulary size is accurately measured by the VST?

2 Methodology

The dataset used in this study was from a larger study of Japanese university students (McLean et al., 2014), which contained data for only the first 8 frequency bands of the 14 band version of the VST. To be qualified for this study, test takers must have completed the entirety of the version of the VST that was administered. The resulting population (N = 2315) was separated into 3 groups, according to the guidelines of Schmitt et al. (2001). The groups were as follows: a vocabulary size less than 3000 words (N = 377), with 3000-4900 words (N = 1478), and those with a vocabulary size greater than or equal to 5000 words (N = 460).

3 Instrument

The 14,000 word family, 140-item version of the VST was administered to test takers in this dataset, however only the first 8 frequency bands (80 items) were given.

4 Analyses

The analyses used in this study included measures of person fit, item fit, person reliability, item reliability, and person ability (from Winsteps output), as well as a Pearson Correlation (using SPSS) of the person ability measures. Item accuracy was also analysed using mean scores within groups within each of the 8 frequency bands (using Excel).

5 Results

The results of the person and item fit analysis suggested that the population in the dataset were not behaving unexpectedly, and were thus worth pursuing further investigation. The results of the second analysis, measuring item accuracy, can be seen in Table 1. For the low group, accuracy dropped below 50% after the 1000 frequency band. For the mid group, accuracy dropped below 50% after the 4000 frequency band, and for the high group, accuracy was above 50% for all 8 frequency bands.
Table 1. Mean Item Accuracy Scores Across Frequency Bands 6000-8000.

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>1000</th>
<th>2000</th>
<th>3000</th>
<th>4000</th>
<th>5000</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>2315</td>
<td>75.76%</td>
<td>53.00%</td>
<td>60.06%</td>
<td>55.27%</td>
<td>43.68%</td>
</tr>
<tr>
<td>Low</td>
<td>377</td>
<td>51.59%</td>
<td>28.41%</td>
<td>31.46%</td>
<td>31.83%</td>
<td>24.43%</td>
</tr>
<tr>
<td>Mid</td>
<td>1478</td>
<td>78.48%</td>
<td>52.08%</td>
<td>61.32%</td>
<td>55.36%</td>
<td>41.43%</td>
</tr>
<tr>
<td>High</td>
<td>460</td>
<td>86.83%</td>
<td>76.13%</td>
<td>79.48%</td>
<td>74.17%</td>
<td>66.65%</td>
</tr>
<tr>
<td>Group</td>
<td>n</td>
<td>6000</td>
<td>7000</td>
<td>8000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-----</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>2315</td>
<td>35.21%</td>
<td>40.22%</td>
<td>41.75%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>377</td>
<td>21.80%</td>
<td>25.92%</td>
<td>23.45%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mid</td>
<td>1478</td>
<td>33.76%</td>
<td>39.13%</td>
<td>39.68%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>460</td>
<td>50.87%</td>
<td>55.46%</td>
<td>63.39%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The results of the third analysis, a Pearson Correlation of mean person ability estimates, suggested that the trend of scores declining would likely continue in each group were they to take all 14 frequency bands. An exception to this was found within the high group, whose similar scores on high frequency bands made it difficult to separate individuals, creating person reliability estimates that were not statistically significant.

The fourth analysis looked at the person and item reliability estimates for each group. The results suggested that the person reliability estimates of the low group remained low at all levels, but declined drastically after the 4000 word frequency band. The mid group began to decline after the 5000 word frequency band, while the high group continued to increase in person reliability along with the word frequency bands. Guessing could also be causing some of the inconsistencies in the measures in this study.

6 Discussion and Conclusion

The results of the study suggest that an empirical validation of the argument against using the entirety of the VST could be justified, and that provision of 1 or 2 levels above a test taker’s ability would be sufficient.

References


A Comparative Analysis of Body Idioms in Two Romance Languages Aiming at Vocabulary Teaching and Learning

Marilei Amadeu Sabino
Universidade Estadual Paulista
UNESP – SJRP/SP - Brazil
amadeusm -at- ibilce.unesp.br

Ariane Lodi
Universidade Estadual Paulista
UNESP – SJRP/SP - Brazil
ariane.lodi -at- outlook.com

1 Introduction

According to Lakoff and Johnson (1980), before the advent of Cognitive Linguistics, metaphor was considered a stylistic issue, but now it is viewed as a critical component of everyday language and a fundamental mechanism of human conceptualizations of the world. It means that “our ordinary conceptual system, in terms of which we both think and act, is fundamentally metaphorical in nature”. A conceptual domain is the representation of any coherent organization of human experience and when one domain is understood in terms of another, we have a conceptual metaphor.

2 Embodied cognition

Another interesting hypothesis in Cognitive Linguistics is that cognition is embodied (Lakoff and Johnson, 1999), that is, our cognition is influenced (if not determined) by our experiences in the physical world: the mind is connected to the body and the body influences the mind. Kovecses (2005) argues that many conceptual metaphors appear to be potentially universal or near-universal, because people across the world share certain bodily experiences. In these terms, it is believed that many metaphors may be identical or very similar in several languages.

3 Aims and methodology of research

In this study we analyzed some somatic (also called body) idioms of Italian and Portuguese languages, in order to investigate the proportion in which their metaphors are the same, similar or different in both languages. We selected hundreds of Italian idioms in dictionaries and looked for their corresponding idioms in Portuguese.

4 Analysis of selected data

Idioms containing same meanings and same metaphors (It was found a very large amount).

<table>
<thead>
<tr>
<th>Italian Idiom</th>
<th>Corresponding Idiom in Portuguese</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Essere sulla bocca di tutti.</td>
<td>Estar na boca de todos.</td>
<td>Ser assunto das fofocas.</td>
</tr>
<tr>
<td>Non chiudere bocca.</td>
<td>Não calar a boca (o bico).</td>
<td>Não parar de falar.</td>
</tr>
</tbody>
</table>

Table 1. Same Meanings and Same Metaphors

Idioms containing different meanings but same metaphors (It was found a good amount).

<table>
<thead>
<tr>
<th>Italian Idiom</th>
<th>Corresponding Idiom in Portuguese</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avere il cervello in vacanza.</td>
<td>Estar com a cabeça longe daqui (em outro mundo).</td>
<td>Fazer raciocínio estúpido, absurdo ou insensato.</td>
</tr>
<tr>
<td>Non avere né capo né coda.</td>
<td>Não ter pé(s) nem cabeça.</td>
<td>Não ter lógica, coerência.</td>
</tr>
</tbody>
</table>

Table 2. Different Meanings but Same Metaphors.
Idioms containing different meanings and different metaphors (It was found not so many).

<table>
<thead>
<tr>
<th>Italian Idiom</th>
<th>Literal translation</th>
<th>Corresponding Idiom in Portuguese and definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scegliere capo per capo.</td>
<td>Escolher cabeça por cabeça.</td>
<td>Escolher a dedo.</td>
</tr>
<tr>
<td></td>
<td>(Escolher detalhadamente)</td>
<td></td>
</tr>
<tr>
<td>Essere la testa di turco (di qualcuno).</td>
<td>Ser a cabeça de turco (de alguém).</td>
<td>Ser o bode expiatório.</td>
</tr>
<tr>
<td></td>
<td>(Ser o alvo, a vítima)</td>
<td></td>
</tr>
<tr>
<td>Avere un diavolo per capello.</td>
<td>Ter um diabo por cabelo.</td>
<td>Ter cabelo nas ventas; estar com a macaca.</td>
</tr>
<tr>
<td></td>
<td>(Estar mal-humorado)</td>
<td>[Ser o manda-chuva. [Ter grande poder e influência].]</td>
</tr>
<tr>
<td>Avere le braccia lunghe.</td>
<td>Ter os braços compridos.</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Different meanings and different metaphors.

5 Conclusions

Through this analysis, we conclude that much of the studied expressions are structurally, semantically and metaphorically identical or similar in both languages. Although this is mere speculation and needs more study, based on these results it seems reasonable to agree with Kövecses’s assertions (2000) since he considers that:

- Universality in conceptual metaphors may facilitate the learning of idioms in the classroom.
- The notion of conceptual metaphors should be taught in an explicit way in the classroom.
- An understanding of underlying conceptual metaphors can aid the retention of vocabulary for those who are learning a foreign language.
- The most common idioms based on metaphors are probably those that have to do with the human body and then should be primarily taught to learners of foreign languages.
- Teachers may improve learners’ awareness of conceptual metaphor, explore its relevance to the learning of idioms and provide strategies to the teaching of these combinations.

6 Acknowledgements

We thank FAPESP – São Paulo State Research Support Foundation – for the financial support.

References

The Effects of Error-Free and Trial-and-Error Techniques for Learning Phrasal Verbs

Brian Strong
Victoria University of Wellington
Brian.strong@vuw.ac.nz

1 Introduction

Phrasal verbs have a notorious reputation as one of the most challenging lexical items of English for second language learners to acquire. Several factors likely account for this apparent difficulty. One is the large number of meaning senses associated with phrasal verbs, many of which are semantically non-transparent. A second factor is that second language learners whose mother tongue has few phrasal verbs usually avoid using them. There has been little research on the effects of techniques to promote the acquisition of phrasal verbs by L2 learners, with the exception of some cognitive linguistics studies (e.g., Boers, 2000; Kovecses & Szabo, 1996; Yasuda, 2010).

One of the aims of the present study was to explore techniques to promote the acquisition of phrasal verbs. One of the most powerful methods of facilitating the learning of lexical items is through retrieval practice, in which a participant retrieves a previously met lexical item. A similar method is generation, in which a participant generates, or guesses, a response when presented with a cue and then is presented with feedback. Cognitive psychology studies have examined techniques to make retrieval practice as error free as possible in order to observe the benefits of later recall, and to make generation techniques as errorful as possible before feedback is given in order to observe the benefits of feedback on subsequent recall. Evidence supports the effectiveness of both techniques.

2 Present study

The present study examined the effects of error-free and trial-and-error on the learning of English phrasal verbs. In the experiment, 110 Japanese university EFL students were divided into two treatment groups (error-free and trial-and-error) and learned 24 phrasal verbs on computers. In the error-free group, participants studied a form-meaning sense correspondence of a phrasal verb (e.g., act up – when a child behaves badly) before it was replaced by a retrieval practice exercise in which the meaning sense cued participants to recall its form (e.g., __ __ -- when a child behaves badly). In the trial-and-error group, participants first were presented with a meaning sense and asked to generate a response, which was nearly always wrong. After submitting their response, feedback was instantly provided (e.g., act up – when a child behaves badly). A cue recall post-test and delayed post-test were administered following the treatments.

It was predicted that (a) the error-free technique would promote the creation of well-specified representations in memory that are necessary for retrieval and production (e.g., Warmington & Hitch, 2014), and (b) the trial-and-error technique would play a powerful role in potentiating encoding of corrective feedback needed for retrieval and production (e.g., Potts & Shanks, 2013). However, it was suspected that the effects of the error-free technique would ultimately be superior to the effects of the trial-and-error technique. Moreover, it was predicted that (c) the participants would recall more verbs than particles, as verbs are more salient than particles.

3 Results

As predicted, the results showed that participants in the error-free group recalled a larger number of correct responses than participants in the trial-and-error group in both post-tests and made more errors with particles than with verbs. A similar finding was reported in a study by Stengers and Boers (2015) with respect to learning collocations. The findings of the present study suggest that, despite corrective feedback after producing an error, making an error is detrimental to subsequent recall and should be avoided during the learning of phrasal verbs. The implication of these findings is that when learning phrasal verbs, learners should make as few errors as possible before any retrieval attempt. One way to achieve this is to provide exemplar material to assist learners before they attempt to recall lexical information in a later task.

References


Common Errors on an L2 to L1 Translation Test

Raymond Stubbe
Kyushu Sangyo University
raymondstubbe -at- gmail.com

Yumiko Cochrane
Kyushu Sangyo University
cochraneyumiko -at- gmail.com

1 Introduction

Translation ability has been identified as a strong indicator of which words students can actually understand while reading (Waring & Takaki, 2003). Furthermore, “asking participants to provide mother-tongue equivalents of the target language words was the most univocal way of verifying (accurate) recognition” (on a Yes-No test; Eyckmans, 2004, p. 77; “accurate” added by author). Schmitt (2008) reviews a number of studies supporting the use of first-language translation. This study investigates and categorizes the common errors that Japanese university students make when translating individual non-contextualized English words into Japanese (L2 to L1).

2 Methodology

An English to Japanese translation test of 80 individual, high-frequency items was prepared. Twenty loanwords and 20 non-loanwords were randomly selected from each of the two compulsory vocabulary wordlists that are used at Kyushu Sangyo University (KSU). The 40 items selected from the first year wordlist were all 1K words, and the 40 items selected from the second year wordlist were all 2K words, according to the JACET 8000 frequency levels (JACET Basic Word Revision Committee, 2003). For each of the 80 tested items, participants (n = 204) were requested to provide two different meanings in Japanese, if possible. Testing took place during December - January 2015-16. In addition to testing the words studied during the school year 2015-16, the second year-students (n = 93) were tested on the words studied in their first year to measure retention. First-year students’ (n = 111) knowledge of words they will encounter next year was also measured. Thus, roughly 54% of students were tested on 40 words that they had yet to study at this university. Each translation for every tested item was translated back into English by one of the authors, then pasted in a spreadsheet in a column under that tested word for analysis. As this study is interested in common errors made on L2-to-L1 translation tests, only repeated mistakes made by at least 20 test-takers were analyzed.

3 Results

Overall, the mean score on this translation test was 43.78 of the 80 tested items (54.7%). The standard deviation was 17.00 (21.3%), with scores ranging from 3 to 72. Of the possible 16,320 translations (204 X 80), 4,507 were left blank (27.6%), 8,932 had at least one correct answer (54.7%), leaving 2,467 (15.11%). However, as test-takers were asked to provide two translations, if possible, the total number of identified errors was 2,837, thus 370 errors must have been accompanied by a correct translation.

<table>
<thead>
<tr>
<th>Error category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synform errors</td>
<td>461</td>
</tr>
<tr>
<td>Part of Speech errors</td>
<td>390</td>
</tr>
<tr>
<td>LW errors</td>
<td>213</td>
</tr>
<tr>
<td>Opposite meaning errors</td>
<td>123</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1187</strong></td>
</tr>
</tbody>
</table>

Table 1. Categories and Counts of Translation Errors Repeated by 20 or More Students.

As mentioned, only the repeated errors made by 20 or more students were analyzed 1,187 in total (see Table 1). Analysis of these repeated errors revealed four specific categories of mistranslations: LW errors, synform errors, part of speech errors, and opposite meaning errors. The greatest number of re-occurring errors was “mail” translated as “email” by 157 students (77% of the total 204 students). This error was deemed a loanword (LW) error, as “mail” is one of the tested LWs. Perhaps surprisingly, only one other LW error was repeated by 20 or more students: “phone,” mistranslated as “mobile phone” by 59 students (29%).

Synform errors made up the largest category of errors, accounting for 461 (33%) of the 1187 repeated errors. According to Laufer (2005), synform errors “are word pairs or groups of words with similar (though not identical) sound, script, or morphology, which learners tend to confuse.” (A detailed explanation of
synforms is available in Laufer, 1988). The word “row,” for example, was mistranslated as “low, law (and) raw” by 34, 26, and 25 students, respectively, suggesting that the l/r distinction remains a problem for these low-level students. Also students’ understanding of vowel pronunciation appears to be weak. In the following, the tested item appears before the mistranslation: pour/poor; fund/find; further/father; duty/dirty (40, 27, 21, and 67 occurrences, respectively). The final two pairs exemplify the difficulty that students may have with the r-controlled vowels “ur” and “ir.”

Part-of-speech errors accounted for 390 (28%) of the repeated errors. The most common of these was “nearly” translated as “near” by 97 students (48%). Similar errors were: true/truth; medical/medicine; heat/hot; defense/defend/protect; mistake/make a mistake; marriage/get married (77, 62, 48, 40, 39, 27, respectively). The final three errors are examples of nouns being mistranslated as verbs.

The final category of common errors was opposite meaning errors. Three word pairs were found in this category (again with the tested item preceding the mistranslation): national/international; east/west; by/from (79, 29, 20, instances respectively). Each of these mistranslations also appears in the KSU wordlists. However, the individual words in these opposite pairs were not taught together, but in separate lessons, as suggested by Laufer (2005; and others). For example, “national” appears in list 2 of the 1st year wordlist, while “international” appears three weeks later in list 5. The words “east” and “west” are found in the 1st and 2nd year wordlists, respectively, separated by roughly 11 months.

4 Conclusion

The present study has focused on common errors made by low-proficiency Japanese university students (n = 204) when translating high-frequency English words into their native language. Four categories of errors were revealed: synform, loanword, part of speech, and opposite meaning errors. By far, the single most common error was the loanword “mail,” being mistaken for “email” by 77% of test-takers. Kyushu Sangyo University English teachers should be advised of this potential confusion when teaching “mail” to second-year students. These teachers should also be advised of the possible confusion with “phone/mobile phone,” when teaching the former to first-year students. Other teachers may also benefit from knowledge of these two potential loanword problems. The largest category of errors was synform errors, which accounted for 39% of the 1,187 errors repeated by at least 10% of students. Many of these synform errors seem to result from vowel confusion; especially r-controlled vowels, as well as l/r confusion. Future research could investigate whether or not phonetic vowel instruction and/or l/r distinction instruction lessens the occurrence of such synform errors. To lessen the number of part-of-speech errors, teachers can inform students of the usual use of the “ly” suffix, and remind students that words such as “mistake” and “marriage” are in fact nouns, not verbs. As for opposite-meaning errors, shortly after teaching the second word in the opposite pair, teachers might consider checking and reinforcing student correct understanding of such word pairs. Analysis of these common translation errors has helped to identify weaknesses in these students’ vocabulary knowledge, which can assist teachers in planning remedial instruction.

References


A Crucial Role of Basic Vocabulary in Second Language Development Of Speech Production: A Longitudinal Study

Shungo Suzuki
Waseda University
s2.kyudo.nut -at- asagi.waseda.jp

1 Introduction

Due to the limited use of English inside and outside of class, Japanese learners of English are less likely to attain a professional or even functional level of English. In other words, their linguistic systems, including vocabulary and grammar, largely comprise declarative knowledge rather than procedural knowledge for lack of meaningful use of English (Sato & Lyster, 2012). Even in English as a foreign language (EFL) settings, English-medium instruction (EMI) offers plenty of opportunities for learners to use English meaningfully and academically with both receptive and productive skills. As vocabulary plays a crucial role in second language (L2) speaking, especially for beginner-to-intermediate level of learners (e.g., Saito et al., 2015), it is highly beneficial to investigate the developmental relationships of vocabulary with a broader complexity, accuracy, and fluency (CAF) framework through such meaningful English use (i.e., EMI). These issues concerning speaking and vocabulary result in the following research questions.

RQ 1: How do lexical characteristics of L2 speaking change through EMI over one academic year?
RQ 2: If there are changes, how are they related to complexity, accuracy, and fluency in speaking?

2 Method

The research was conducted in a quasi-experimental design with the pre-, mid-, and post-tests. The participants in this study were 8 undergraduate Japanese learners of English as a foreign language (4 male, 4 female) recruited from the target EMI course. They took the two-hour EMI session every week over one academic year in Japan (in total, 56 hours of in-class exposure). Their proficiency level was regarded as intermediate according to the TOEFL ITP score (MTOEFL ITP = 523.7). All the pre-, mid-, and post-tests adopted two monologic tasks (narrative and argumentative speech) and were examined using quantifiable measures. The linguistic measures covered major domains of performance assessment: 1) lexical and syntactical complexity, 2) accuracy, and 3) speed, breakdown, and repair fluency (see Bulté & Housen, 2014).

3 Results

A set of non-parametric Friedman tests were performed for the performance in narrative and argumentative speech with all the linguistic measure scores as dependent variables, with an alpha level set at p < .05 (two-tailed test). Consequently, Bonferroni post-hoc tests were carried out. The descriptive statistics of pre-, mid-, and post-test lexical measure scores only in the argumentative task are summarized in Table 1 due to limited space.

While no lexical measures in the narrative task reach statistical significances, K1+K2 coverage in the argumentative task ($\chi^2 = 10.75, p = 0.005$) yields the statistical difference among pre-, mid-, post-tests. In addition, the log count index ($\chi^2 = 5.25, p = 0.072$) produces a relatively low $p$-value ($p > 0.1$), subsequently suggesting an ecologically statistical difference. Due to the limited sample size, however, only the log count index ($z = 2.38, p = 0.017, r = 0.60$) yields a statistical difference between the pre- and mid-tests, whereas large effects ($r > 0.5$) are detected in the pre-to-mid K1 + K2 coverage ($z = 2.10, p = 0.036, r = 0.53$) and the mid-to-post log count index ($z = 2.10, p = 0.036, r = 0.53$).

As for relationships between lexical changes and other measure scores, their scores were initially transformed into percentage change scores $^1$, following Lennon (1990). To detect either supportive or competitive relationships, a series of non-parametric Spearman’s rank-order correlations were performed. Two correlational relationships between pre- and mid-tests and four between mid- and post-tests were obtained across two different CAF domains, as summarized in Table 2.

---

$^1$ Percentage change scores were calculated as follows: Pre-to-Mid % change = (Mid-test score – Pre-test score)/Pre-test score × 100; Mid-to-Post % change = (Post-test score – Mid-test score)/Mid-test score × 100.
4 Discussion & conclusion

There was the statistical increase in K1+2 coverage in the argumentative task. In EMI, the communicative needs, which might have exceeded their language ability at the beginning of the year, may have forced them to prefer using relatively frequent and easy vocabulary in order to function with their limited L2 oral skills (Polat & Kim, 2014). In other words, if learners are required to pay attention to the content of speech (e.g., difficult topics), their attention to lexical sophistication can be distracted (Skehan, 2009). The EFL learners in the study could not develop lexical complexity without explicit instruction. This finding goes against those of studies of EAP programs in ESL settings (e.g., Mazgutova & Kormos, 2015). Thus, in order to express cognitively demanding or academic topics, learners might have to modify their L2 system, such as the connection between concepts and corresponding lemmas, which requires time for consolidation (Housen et al., 2012).

In EFL settings, when explicit language instruction is not available, L2 learners tend to prefer frequent or relatively easy vocabulary to accomplish demanding communicative needs. Dealing with consistent topics in multiple modalities in EMI can be a meaningful and repetitive practice for learners, promoting the proceduralization of frequent vocabulary at the temporal expense of speed fluency.

References


Analyzing Spoken texts of Under 300 Tokens: The Vocabulary Profiles and Lexical Richness of Japanese EFL Students.

Andy Vajirasarn
Toyo University
vajirasarn -at- toyo.jp

1 Introduction

Measuring lexical richness is useful for a variety of situations, including evaluating the difficulty of a text, such as an authentic novel, for use in class, checking the range of L2 learners’ productive vocabulary, and even for clinicians who screen patients, looking for language or developmental deficiencies.

The present study explores the relationship of certain measures of lexical richness to the proficiency levels of ESL students’ oral interview data. This study allowed the researcher to identify which of the lexical richness measures correlate with raters’ judgement of the proficiency level of EFL students.

2 Lexical richness

Over the years, different components of lexical richness have been proposed, as well as different ways to measure these components. Read’s (2000) conceptualization focuses on four interrelated features in particular: lexical density, lexical sophistication, lexical diversity, and number of errors in productive use of vocabulary. Nation and Webb (2011) state that the two aspects of lexical variation and lexical sophistication have received the most attention in studies of productive vocabulary. The present study focuses on measuring lexical richness in terms of lexical variation and sophistication.

The most often researched aspect of lexical richness is lexical variation, also called lexical diversity. It refers to the range of vocabulary used by the learner in a spoken or written text. The classic metric of lexical variation is the Type-Token Ratio (TTR), also known as the number of different words (NDW). As Nation and Webb (2011), among others, point out, TTR is sensitive to text length such that the longer the text, the smaller the ratio of new different words.

Other ways using computational methods have been developed to adjust for TTR’s sensitivity to text length. These approaches include the D measure (Malvern & Richards, 1997), the measure of textual lexical diversity (MTLD) (McCarthy & Jarvis, 2010), and the moving-average type-token ratio (MATTR) (Covington & McFall, 2012). In the present study, D, MTLD, and MATTR were used to explore lexical diversity of short, spoken texts.

Laufer and Nation (1995) argue that, by taking into account the levels of vocabulary sophistication, we can get a better picture of the vocabulary in use. The concept of lexical sophistication refers to the proportion of “advanced” words used by the learner. Lists of word frequency can help the researcher to define what counts as “advanced” by setting a certain word frequency level. There are software packages and websites which offer vocabulary profiling features to identify the frequency of words used in a text.

3 Methodology

Speech data was collected from 20 EFL university students in Japan, 14 females and 6 males. Two students were from China and the rest were from Japan. They were grouped in three levels based on their scores on an in-house speaking test, High-Intermediate (n = 5), Intermediate (n = 10), and Low (n = 5). Recordings were transcribed and cleaned before conducting final analyses. In the present study, D, MTLD, and MATTR were used to explore lexical diversity, and vocabulary profiling was used to examine lexical sophistication in short spoken texts of around 300 tokens.

4 Findings

As shown in Table 1, the lexical diversity measures for MATTR and MTLD shown in Table 1 display expected results; as proficiency levels increase, so does the measure of lexical diversity. However, VOCD-D measure indicates that the Intermediate group seems to have performed better than the High group.
To investigate this pattern further, Pearson correlational coefficients and significance tests were calculated. Table 2 shows the values for the correlations. All measures of lexical diversity correlated significantly and positively with the proficiency scores (logits), \( r_{\text{MATTR}}(18) = .57, p < .01 \), \( r_{\text{MTLD}}(18) = .71, p < .01 \), \( r_{\text{VOCD}-D}(18) = .72, p < .01 \). Those with higher proficiency tended to use a greater variety of words.

<table>
<thead>
<tr>
<th>Proficiency</th>
<th>MATTR(10)</th>
<th>MTLD</th>
<th>VOCD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proficiency</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MATTR(10)</td>
<td>0.57</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>MTLD</td>
<td>0.71</td>
<td>0.90</td>
<td>1.00</td>
</tr>
<tr>
<td>VOCD</td>
<td>0.72</td>
<td>0.59</td>
<td>0.69</td>
</tr>
</tbody>
</table>

Table 2. Pearson's Correlation among the LD Measures and Proficiency Scores.

Table 3 shows the lexical sophistication measures. The intermediate group performed better than the High group in terms of the number of word families and also had a higher proportion of AWL word families (3.15%) than the High group (2.53%). This is a counterintuitive finding which needs to be clarified, perhaps by examining participant factors and the topics of the prompts used in the interview test.

<table>
<thead>
<tr>
<th>Level</th>
<th>GSL 1k</th>
<th>%</th>
<th>GSL 2k</th>
<th>%</th>
<th>AWL</th>
<th>%</th>
<th>Offlist</th>
<th>%</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-Int.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Families</td>
<td>207</td>
<td>87.34</td>
<td>24</td>
<td>10.13</td>
<td>6</td>
<td>2.53</td>
<td>?</td>
<td>?</td>
<td>237</td>
</tr>
<tr>
<td>Types</td>
<td>255</td>
<td>86.15</td>
<td>25</td>
<td>8.45</td>
<td>7</td>
<td>2.36</td>
<td>9</td>
<td>3.04</td>
<td>296</td>
</tr>
<tr>
<td>Tokens</td>
<td>937</td>
<td>92.59</td>
<td>40</td>
<td>3.95</td>
<td>12</td>
<td>30.00</td>
<td>23</td>
<td>2.27</td>
<td>1012</td>
</tr>
<tr>
<td>Intermediate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Families</td>
<td>248</td>
<td>86.71</td>
<td>29</td>
<td>10.14</td>
<td>9</td>
<td>3.15</td>
<td>?</td>
<td>?</td>
<td>286</td>
</tr>
<tr>
<td>Types</td>
<td>304</td>
<td>86.86</td>
<td>32</td>
<td>9.14</td>
<td>10</td>
<td>2.86</td>
<td>4</td>
<td>1.14</td>
<td>350</td>
</tr>
<tr>
<td>Tokens</td>
<td>1213</td>
<td>94.47</td>
<td>44</td>
<td>3.43</td>
<td>17</td>
<td>1.32</td>
<td>10</td>
<td>0.78</td>
<td>1284</td>
</tr>
<tr>
<td>Low</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Families</td>
<td>153</td>
<td>88.44</td>
<td>16</td>
<td>9.25</td>
<td>4</td>
<td>2.31</td>
<td>?</td>
<td>?</td>
<td>173</td>
</tr>
<tr>
<td>Types</td>
<td>180</td>
<td>87.80</td>
<td>16</td>
<td>7.80</td>
<td>4</td>
<td>1.95</td>
<td>5</td>
<td>2.44</td>
<td>205</td>
</tr>
<tr>
<td>Tokens</td>
<td>585</td>
<td>93.75</td>
<td>27</td>
<td>4.33</td>
<td>4</td>
<td>0.64</td>
<td>8</td>
<td>1.28</td>
<td>624</td>
</tr>
</tbody>
</table>

Table 3. Lexical Sophistication of EFL Learner by Word List Type and Unit of Vocabulary.

*Note.* GSL = General Service List; 1k = first 1000-word list; 2k=second 1000-word list; AWL = Academic Word List

References


Constructing a ‘Sound’ Vocabulary Course where Theory and Practicality Converge

Joseph P. Vitta  
Tokyo International University  
vittajp -at- gmail.com

Simon W. Albright  
Nottingham University  
uroshnor -at- hotmail.com

1 Introduction: An innovative lexical intervention in the Middle East EFL setting

Beginning in 2013, Saudi Arabia started a technical college program where the medium of instruction would be English. These institutions were designed to have two concurrent ELT goals: 1 – general EFL proficiency that was a mandated prerequisite to begin technical training (A2/B1 on the CEFR as measured by the Cambridge KET/PET); and 2 – Technical ESP development with an aim to develop the content language within a technical field while identifying common carrier language. As managers of the ELT provision at one of these college’s from 2014 to 2016 (two academic years), we designed, operated, and managed and a stand-alone vocabulary course to help us attain these goals. Herein, we report on the theory underpinning the course, a brief presentation of the course and the process driving and emerging for it, and a consideration of its effectiveness is presented.

2 Underpinning theory

Lewis’s (1993) Lexical Approach and related viewpoints, such as Thornbury, 2002, that called for a handling of vocabulary teaching that is cognizant of the complexity of lexis was the first underpinning theoretical trend. In addition, the following studies informed our work:

1 - Lafer and Nation (1999) – productive vocabulary size (by frequency bands) were strongly associated with overall proficiency.

2 - Matthews and Cheng (2015) -- receptive vocabulary size (up to 3k) significantly accounted for 52% (r^2) of the variance in Chinese ELL performance on the ILETS listening exam.

3 - Hulstijn, Schoonen, de Jong, Steinel and Florijn (2011) - productive vocabulary size was the strongest factor in models predicting (88% success) observed Dutch SL speaking proficiency levels (A2-B2).

3 The course

Our College had two ‘tracks’ of instruction, a general EFL-focused Foundation Year (FY), 90% general EFL + 10% basic IT, that preceded technical training years. Both had ELT curricula that included the stand-alone vocabulary course governing this project. For both versions of the course, lessons were designed to progress through three sequential stages from a focus on single words to a consideration of multiword constructions since most students entered with such low proficiency. Higher proficiency students were streamed into sections where the polyword focus began almost immediately, especially in general EFL. Both iterations of the course met for five hours per week and followed a modified PPP approach where purposefully chosen words/lexical items were presented, practiced with teacher guidance, and then produced/processed in less-controlled student-centric tasks. Lexical notebooks were kept where students transcended more L2-L1 translation and recorded words and later polywords with the detailed information as per Lewis’s (1993) model. Regarding assessment, weekly quizzes were given which served both formative and summative purposes. At trimester end, there was a final exam.

The operational process we developed and suggest was/is cyclical in nature with the end of the cycle driving into the beginning. First, the type/goal (general EFL vs. technical ESP, in our case) and level (single vs. polyword) of the lesson are identified. Second, the words/lexical items are purposefully chosen. In our case, this might have been our confirming that a selected word was in both our course book and PET/KET vocabulary list or that a collocated phrase such as ‘low/high voltage’ was in our technical materials and had a sufficient number of hits in a well-known corpus. Third, the lesson is constructed according to the modified PPP approach already mentioned with activities whose effectiveness has been established by the literature – e.g. cloze activities (Folse, 2000). Fourth comes the establishment of assessment and feedback loops. Regarding the latter, we would survey the students every trimester and have the bi-lingual teachers conduct informal interviews in the L1 periodically while also seeking our teachers’ opinions of the course consistently during our staff meetings. What is essential to the steps in this process is managerial control that also takes teacher and student voices into account.
4 Considerations of effectiveness

Because of limitations on the ground and the singular focus of the College’s stakeholders on training, a true empirical approach (i.e. experimental or quasi-experimental design) to assess our course’s effectiveness was not possible. This reflection, therefore, relies on indirect yet credible evidence.

Our college was qualitatively and independently reviewed twice during the 2014-2015 school year (reports for 2015-2016 school year are still unpublished). These were positive and pointed indirectly to the effectiveness of the course when it was run to serve primarily general EFL purposes. In the report from the first observation, the English curriculum was described as “well planned and thought out,” and the department’s management “thinking creatively about how best to deliver quality.” While this assessment was holistic, the vocabulary workshop’s contribution to this positive conclusion could be reasonably assumed since it was the only component of our EFL program that was truly in-house and internally developed. In a later institutional review of the College, the vocabulary workshop and its associated PD sessions and materials led in part to the reviewers to state that “management at curriculum level is effective” with their overt referencing of this innovation.

There were also quantitative data that pointed to the overall effectiveness of the general EFL instruction and given our innovation’s prominent place in it, we offer these as indirect support of its possible effectiveness without claiming true causality:

- Cambridge English Placement Test acted as pre- and post-term assessment for the first trimester of the 2014-5 academic year. It was observed that students had significantly and weakly to moderately, ($p = .03; Z = -2.926, r = .24; r^2 = .06$) improved when analyzing the valid cases (N = 141) via Wilcoxon Signed Rank testing, a nonparametric equivalent of dependent sample t-testing.
- At the end of the 20 our remaining (N = 71) students took the KET exam and our observed A2 or higher (pass) rate was 10% greater than the CoE average for the 2013-2014 academic year (22% vs. 20%).
- In 2015-2016, our student numbers were quite low so empirical data was lacking. Nevertheless, 28% (4 of 13 A2 students) AD students who needed to attain B1 in the second year did so.
- Student performance on General EFL vocabulary workshop quizzes and finals consistently had a pass rate (score>70%) of above 80% for both years.

Data on the effectiveness of the course vis-à-vis our College’s ESP goals were divergent. Regarding the ability to pre-teach basic terminology for technical training in the Foundation Year, there was positive evidence as 97 out 141 students passed (pass > 70%, $M = 80.5; SD = 15.67$) the IT vocabulary section of their final exam (1st trimester – 2014-5 academic year). Informal dialogue with the IT trainers revealed that most students indeed recognized the terms in subsequent instruction. Despite this positive finding, it is important to remember how small a component of the course this was. In the technical training year, we observed that while students did at least grasp the collocations being taught (e.g. low voltage, voltage source) but formative and summative assessments saw students performing poorly. Technical trainers also commented that they had to reteach terms that were supposed to be covered in the ESP vocabulary lessons. What was more was that performance on mock AD final exit exams saw students unable to pass.

References


The Lexical Research Foundation

Rob Waring
Notre Dame Seishin University
waring.rob -at- gmail.com

1 Introduction

Interest and activity in second and foreign language vocabulary acquisition, corpus linguistics and many other areas of lexical acquisition and research have quite rightly been booming in recent years. This has lead to a vast increased in knowledge in, and about, our field, which has to some degree become fragmented, disparate, and rather difficult to access. The Lexical Research Foundation (www.lexicalresearchfoundation.org) has been set up to address this and other issues.

2 The remit of the Lexical Research Foundation

The Lexical Research Foundation (LRF) is a not-for-profit organization that has been set up to attempt to bring together this disparate and fragmented field by creating an umbrella organization to oversee vocabulary conferences, promote good research and pedagogy in second language vocabulary / lexical acquisition; become a resource center; and act as a focal point for developing a community of good practice and research.

The LRF is modelled on the Extensive Reading Foundation. As a foundation, rather than an association which has members, 8-10 eminent individuals will be appointed to the LRF Board to run and operate it on a volunteer basis. Other individuals will be asked to contribute to the work of the LRF. There will be no members, no memberships, no dues. All funding will come from sponsorship, donations and the profit from running conferences. Any profits from the Vocab@Tokyo conference will be used by the LRF to assist with the funding of future vocab@ conferences and LRF initiatives.

The following tasks are expected to be part of the LRF remit:

1. To run vocabulary conferences
2. To maintain a searchable bibliography on SLVA with links to papers online
3. To conduct or oversee projects to develop non-proprietary open-source vocabulary resources, vocabulary tests, software, assessment tools, wordlists and databases, vocabulary learning apps etc.
4. To act as a central go-to resource (software links, files, links, tests, worksheets etc.) for the fragmented resources currently available online
5. To publish or store materials promoting good practice in vocabulary pedagogy
6. To host PowerPoint files and handouts from conferences and show videos (hosted by YouTube and others) taken of talks.
7. To nurture a vibrant community of vocabulary researchers and practitioners
8. To offer grants

Other tasks will emerge in time and in response to needs and the capabilities of the Board.

3 Setting up

Interested parties are welcome to discuss the setting up and running of the LRF to create its by-laws and structure.

Individuals are also invited to volunteer their services to help the LRF and the field.
Can Japanese University Students Learn Spelling Through Extensive Reading?

1 Introduction

Learning spelling is an important but neglected area of vocabulary acquisition. For Japanese learners, one of the most difficult aspects of learning spelling is the correct usage of L/R (Cook, 1997). In this study, we investigate the possibility that learners can incidentally acquire knowledge of correct L/R usage, primarily for proper nouns, through extensive reading.

2 Background studies

A number of researchers have investigated whether learners acquire word meanings through reading (cf., Webb & Chang, 2015). While a number of studies have examined the acquisition of word meanings, the few studies that have examined the acquisition of word forms (Webb, 2005; Webb, 2007) have done so as a secondary aspect of the acquisition of word meaning. To the authors’ knowledge, the present study is the first attempt to see the effect of a large amount of input solely on the learning of forms through extensive reading. The study examines whether or not Japanese learners can pick up one aspect of word forms, that of spelling. With L1 spelling that is markedly different from L2, the acquisition of spelling knowledge is a daunting task for Japanese learners (Cook, 1997; Figueredo, 2006). This study focuses on the acquisition of a distinction that is one of the hardest among Japanese learners, that of L/R.

3 Goals of the study and research questions

The goal of the study is to investigate whether or not Japanese university students can learn spelling, focusing on learning the distinction of L/R in words that appear repeatedly in extensive reading. The study examined the following research questions.

- RQ1: Can learners pick up recognition knowledge of L/R through extensive reading?
- RQ2: Can learners pick up production knowledge of L/R through extensive reading?
- RQ3: Is there a relationship between the frequency of the target words and learning the L/R distinction?

4 Methods

The participants were 42 Japanese first-year university students majoring in English language and literature. The study was conducted as part of class curriculum in a 15-week course during one semester. As required homework, the students read 13 graded readers from the 15 available Level 2 Cambridge English Readers, reading an average of one book a week. Each book contained around 10,000 words, so each student had read approximately 130,000 words by the end of the semester.

The target words were chosen from the corpus of the 13 books and consisted of 19 proper nouns and one noun (“flat”) based on frequency in the corpus, which ranged from 30 to 200 times.

Two types of spelling tests were given: One was a recall test (productive knowledge) and the other was a recognition test (receptive knowledge). In the recall test, the learners had to spell out L2 forms for each word written in the L1 (Katakana). In the recognition test, the learners had to choose a correct form, with either L or R in each word, such as “Alex” or “Arex.” The students took a pretest with the two types of spelling test before beginning the reading assignments. After finishing the course, they took a post-test with the same two types of spelling tests.

5 Results

RQ 1 & 2: Effects of extensive reading on spelling, results of recognition and production tests

As Table 1 shows, the results of the recognition and production tests revealed significant differences between the pre- and post-tests, indicating that learners became better at recognizing and producing the correct forms, in this case, spelling, with a particular focus on the distinction of L and R, through exposure to a large amount of input via extensive reading.
A number of researchers have investigated whether learners acquire the acquisition of word meanings through reading (Cook, 1997; Scrivener, 2004; Figueredo, 2006; van Pernis, 2007). In this study, we examined whether learners can acquire the distinction of L/R in words that appear repeatedly in extensive reading on spelling, which is primarily focused on the distinction of L/R in words (Figueredo, 2006; van Pernis, 2007). To the authors' knowledge, the present study is the first attempt to examine whether learners can learn the L/R distinction in spelling in proper nouns through exposure to large amounts of input through extensive reading. The frequency factor of the words seemed to be related with the learning but could not reliably predict the level of increased acquisition, suggesting that other factors need to be considered.

### 6 Conclusions and needs for further studies

The results indicate that learners can learn the L/R distinction in spelling in proper nouns through exposure to large amounts of input through extensive reading. The frequency factor of the words seemed to be related with the learning but could not reliably predict the level of increased acquisition, suggesting that other factors need to be considered.

### 7 Acknowledgement

This study was supported by JSPS Grant-in-Aid for Scientific Research (C), Grant Number 15K02688.

### References


L2 Vocabulary Acquisition: Processing Techniques to Help Turn Input into Intake

Muhammad Sarwar Bajwa  
Professor (ELT) and Dean Social Sciences  
NCBA&E, Bahawalpur, Pakistan  
sarwar_bajwa-at-yahoo.com

Nargis Rashid  
Lecturer (English)  
Govt. College for Technology (W), Bahawalpur  
nargis rashid124-at-gmail.com

1 Introduction

Vocabulary has a pivotal position in language learning and use (Zhavoronkova, 2014; Shoari & Aidinlou, 2015, p. 208; Li & Zhao, 2015, p. 39). However, its acquisition is not an easy task (Yu-rong 2015, p. 253). Vocabulary items remain mostly unassimilated and unavailable to the users of language at production stage. Many researchers have tried to find out solution to this problem, however, they do not seem fully confident about their findings. Along with many other researchers Schmitt (2008) says that “the best means of achieving good vocabulary learning is still unclear” (p. 329). In this scenario of uncertainty, vocabulary acquisition researchers can be divided into two categories. Broadly putting, the researchers like Schmitt, Nations, etc., abide by information model and they believe that the vocabulary learning is incremental, developmental, and accumulative. The other category of researchers, like Barcroft and Vela are the followers of cognitive model of vocabulary learning. These researchers believe in the deep processed input and long-term retention and emphasize the deep processing i.e., analysis, understanding and organization of input. A step ahead, this study focuses the generation of mental representations of input lexical items (encoding) that can consolidate with the pre-existing information in long-term memory meaningfully according to the framework of Atkinson Shiffrin’s Theory of Memory.

2 Research Questions

1. What are encoding techniques for generating mental representations of input lexical items?
2. What is the impact of encoding techniques on the assimilation of input lexical items?

3 Methods

The design of the study was a pretest and posttest experiment with two groups i.e., control group and experiment group. The independent variable was processing of learning materials whereas dependent variable was learners’ assimilation of target vocabulary knowledge Intervention materials were developed on the basis of comprehensive input and in the light of principles given by Tomlinson (1998) and the criteria suggested by Dick and Carey (1990). The researchers selected eight encoding techniques for materials development on the basis of the opinions of various researchers (Azin, N., et al., 2015, and others). The intervention went on for ten weeks. Vocabulary acquisition was assessed with the help of pre-test and posttest. Each test carried ten-items. These items aimed to assess learners’ acquired knowledge at receptive and productive levels. Data received as pre-test and posttest awards were analyzed by employing t-test.

4 Results

The aspects like ‘multi-word units’ and ‘conventionalized metaphors’ were found at means level 7.0 each and from pre-test to posttest mean difference 4.82 and 4.73, respectively. The means of aspects like ‘derivations’, ‘sense selection’, ‘reference specification’, and ‘level of formality’ were in between 6.8 and 6.5 with mean differences in between 4.8 and 4.3. The means of the ‘form recall’, ‘long composition’ and ‘fluency’ were 5.8, 6.2, and 6.2 respectively with mean differences 4.1, 4.2, 4.1 respectively. These results show that the acquisition in case of “multi-word units” and “conventionalized metaphors” was the highest of all, while the acquisition in case of ‘derivations’, ‘sense selection’, ‘reference specification’ and ‘level of formality’ were medium whereas the acquisition in the area of ‘form recall’, ‘long composition’, and ‘fluency’ were the lowest of all. When
the acquisition is seen on the whole mean difference is significant.

5 Discussion and Conclusion

The aspects of acquisition that were at the highest or medium level were comparatively receptive because the learners frequently deal with them and further the lexical items at receptive level are presented more specifically. On the other hand, the aspects like ‘form recall’, ‘long composition’ and ‘fluency’ were comparatively low in scale. The reason that can be assigned to the backwardness of these aspects is that these are purely productive in nature in respect of which learners need to depend upon forms that are to be recalled and retrieved from the long-term memory. The learners, though have sense or meanings of the items that are to be used productively (Schmitt, 2014, p. 919), find it difficult to recall or retrieve the required forms because the forms which are not mostly in active use go submerged in the long-term memory and are retrieved with difficulty. That is why it is imperative that the mental representations of the target vocabulary items should be enacted soundly. This can be done through making the learners understand comprehensive input optimally and meaning encoded systematically so that the representations of these items duly find their proper station in the long-term memory. If this becomes possible, contrary to the apprehension of (Moras 2001), the issue of forgetting forms is almost resolved and concepts and forms remain available for productive use.

References


Written Production of Metaphorical Vocabulary in a CLIL Course: Effects of L1 Priming and Explicit Instruction

Phil Bennett
Miyazaki International College
philbennett59 -at- gmail.com

1 Background

Research has shown that metaphor awareness is important for learners in academic contexts, and studies are now beginning to investigate the explicit teaching of metaphorical vocabulary and how learners add this to their written production (Boers, 2004; Littlemore et al., 2013). Word frequency is also known to exhibit a powerful influence on vocabulary learning, and some studies are beginning to consider the frequency of figurative language as a potential influence on learning (Macis & Schmitt, 2016). However, the frequency effects of the L1 on metaphor production have yet to receive much attention. The present exploratory study seeks to make a small contribution in this regard by examining the priming influences due to frequency in both the L1 and L2 as well as classroom exposure on learners’ production of metaphorical vocabulary.

2 Purpose

The study will consider the following research questions:

1. Can evidence be found of facilitative or debilitative influences from the L1 on learners’ production of specific metaphors?
2. How do lexical frequency in the L1/L2 as determined by corpora and frequency of classroom input relate to learner production of discipline-specific metaphor?

3 Method

The control (n = 23; TOEIC = 475) and experimental (n = 23; TOEIC = 498) groups were Japanese university students enrolled in a CLIL anthropology course on Japanese Popular Culture. While content was identical in both conditions, the treatment of metaphor differed. The experimental group completed explicit awareness-raising activities on how lexical items can take on extended, figurative meanings to express abstract concepts. The control group completed activities to improve knowledge of high-frequency and academic vocabulary present in the course input, but metaphor awareness was not explicitly taught. Both groups submitted weekly reflective writing in response to the same prompts. This writing was used to create learner output corpora so that the production of metaphorical vocabulary in the two conditions could be analysed.

In order to reduce the complexity of the data set, only metaphorical verbs (e.g., evolve, absorb, shape) in a collocational relationship with the noun culture(s) were analysed in this study. All instances of these verbs being used in metaphorical senses as collocates of culture(s) were recorded. COCA was used as a reference corpus to provide data on the frequency of these collocations in general English, and the frequencies of their translated forms in Japanese were estimated with the jpTenTen11 corpus. Corpora based on transcripts of teacher talk and classroom materials provided data on collocational frequency in classroom input. All data were converted to normalized frequencies and correlations were used to examine the relationships between learners’ lexical choices and frequency in both English, the L1, and in classroom exposure.

4 Results

In total, 14 metaphorical verbs were used by learners with the noun culture(s), and the corpus data for a selection of these verbs is provided in Table 1. Absorb and take in appear to be cases where the L1 facilitates production in English; the Japanese translations of these words, kyuushuu and toriireru, are both more strongly associated and more frequent collocates of bunka, the Japanese for culture, than their English equivalents. The cases of borrow and mix, meanwhile, illustrate varying degrees of responsiveness by learners. In neither language are these words particularly strong collocates of culture, but while experimental group learners appeared to respond very positively to the explicit focus on mix as a metaphor (as evidenced by high output frequency), borrow was used only twice as a metaphor and never as a collocate of culture. The case of shape is interesting. In English, it is among the more common verbs to collocate with culture(s) (both shaped and shaping are in the top 100 verb form collocates by MI score in COCA), and its equivalent
in Japanese is more frequent than mix, yet learners barely used this word productively. In fact, the output frequency of 16.72 per million words in Table 1 represents just a single use of this word in collocation with culture(s). Further investigation may be needed to discover why learners avoid some words. Finally, an example of how differing conceptual patterns in the L1 can sometimes mislead learners can be seen in the frequent use of grow/bring up as a direct subject or object of culture(s) (e.g., grow up a culture). In Japanese, the verb hagukumu is a highly frequent collocate of bunka that collocates with other abstract notions such as love, friendship, and motherhood, as well as concrete terms such as children in its more literal sense. However, this appears to be a source of negative transfer that instructors could usefully make learners aware of.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>absorb</td>
<td>kyuushuu</td>
<td>2.75</td>
<td>2.00</td>
<td>7.75</td>
<td>16.30</td>
</tr>
<tr>
<td>take in</td>
<td>torireru</td>
<td>4.09</td>
<td>-0.35</td>
<td>3.67</td>
<td>0.00</td>
</tr>
<tr>
<td>borrow</td>
<td>kariru</td>
<td>-1.11</td>
<td>1.26</td>
<td>5.77</td>
<td>8.15</td>
</tr>
<tr>
<td>mix</td>
<td>mazeru</td>
<td>1.17</td>
<td>0.88</td>
<td>7.46</td>
<td>0.00</td>
</tr>
<tr>
<td>shape</td>
<td>katachidzukuru</td>
<td>3.92</td>
<td>12.40</td>
<td>14.67</td>
<td>0.00</td>
</tr>
<tr>
<td>grow up</td>
<td>hagukumu</td>
<td>6.07</td>
<td>43.84</td>
<td>/</td>
<td>/</td>
</tr>
</tbody>
</table>

* Frequency values per million words
b Values calculated for collocation with the Japanese bunka (culture)

Table 1. Corpus Data for Selected Metaphorical Verb Collocates of culture(s).

To investigate overall frequency effects for the L1, L2, and classroom input, correlations were calculated for learner output of the 14 metaphorical culture(s) collocations and the frequencies of both the individual verbs and the culture(s)/bunka collocations. As Table 2 shows, there are both similarities and differences between the conditions. In neither condition is there any evidence of a relationship between learner output and L2 frequency. Nor is there much suggestion that word (rather than collocation) frequency relates to learner output. The only correlations that reach moderate strength are both for collocations: from frequency in the L1 for the control group and classroom input in the experimental condition. This may suggest that learners in incidental learning conditions are more influenced by the norms of their own language, while explicit instruction can help to raise awareness or noticing of target features in classroom input.

<table>
<thead>
<tr>
<th>Condition</th>
<th>L1 (jpTenTen11)</th>
<th>L2 (COCA)</th>
<th>Input corpora</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Word frequency</td>
<td>Collocation frequency</td>
<td>Word frequency</td>
</tr>
<tr>
<td>Control</td>
<td>.187</td>
<td>.418</td>
<td>-.157</td>
</tr>
<tr>
<td>Experimental</td>
<td>.219</td>
<td>.156</td>
<td>.157</td>
</tr>
</tbody>
</table>

* Significant at the 0.05 level

Table 2. Spearman's Rho Correlations between Learner Output and L1, L2, and Classroom Input Frequency

5 Discussion

These findings imply that learners in incidental learning conditions draw on the conceptual patterns of their L1 when writing on abstract concepts in the L2. This should come as no surprise, but the study also suggests that by identifying potential pitfalls and by providing explicit instruction in common metaphorical patterns, instructors can guide learners toward greater lexical variety. Further studies will be carried out on a broader range of metaphors using mixed effects models to account for non-independent observations in the data.

References


The Relationship between Knowledge of Collocations and General Proficiency

Dale Brown
Kanazawa University
dbrown -at- staff.kanazawa-u.ac.jp

1 Introduction
The relationship between L2 learners’ knowledge of collocations and their general L2 proficiency is unclear, with some studies, such as Al-Zahrani (cited in Nesselhauf, 2005), Gonzalez Fernandez and Schmitt (2015) and Barfield (2009) finding evidence of a relationship, while others, for example, Howarth (1998) and Nesselhauf (2005) do not.

Barfield’s study cited above introduced a new instrument for eliciting learners’ productive knowledge of collocations. The instrument, named LexCombi, asks learners to provide three collocates for each of 30 noun cues. In the course of further developing LexCombi, I have conducted a series of four studies in each of which the correlation between LexCombi scores and a measure of proficiency was calculated. As Table 1 shows, one study found a correlation similar to Barfield’s, while the other three found almost no correlation.

This paper explores three factors that may explain these contrasting results: (1) Differences between the participants in the studies; (2) Differences in the measures of proficiency; and (3) Differences in the measure of knowledge of collocations, i.e. differences between the successive versions of LexCombi as the instrument underwent development.

<table>
<thead>
<tr>
<th>Study</th>
<th>N</th>
<th>Correlation (Pearson’s)</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barfield (2009)</td>
<td>89</td>
<td>.57</td>
<td>$p &lt; .001$</td>
</tr>
<tr>
<td>Study A</td>
<td>36</td>
<td>.12</td>
<td>$p &gt; .05$</td>
</tr>
<tr>
<td>Study B</td>
<td>125</td>
<td>.52</td>
<td>$p &lt; .001$</td>
</tr>
<tr>
<td>Study C</td>
<td>223</td>
<td>.13</td>
<td>$p &gt; .05$</td>
</tr>
<tr>
<td>Study D</td>
<td>146</td>
<td>.05</td>
<td>$p &gt; .05$</td>
</tr>
</tbody>
</table>

Table 1. Correlations between LexCombi Scores and General Proficiency in Five Studies.

2 Differences between the participants
The participants in all five studies were Japanese L1 university students. In Barfield’s study and Study A all were non-English majors, while Studies B, C and D included both English majors and non-English majors. A wide range of proficiency levels were included in four out of the five studies; in Study A, however, participants were drawn from two classes of similar proficiency as judged by an internal university placement test. This narrower proficiency range may then account for the low correlation found in Study A.

3 Differences between the measures of proficiency
Barfield’s study used TOEIC scores as a proficiency measure, while Studies A to D used Yes/No vocabulary size tests. There is therefore a clear difference between the studies. Nevertheless, vocabulary size in general and Yes/No tests in particular have been found to correlate strongly with other measures of proficiency: see, for example, Meara and Jones (1988), Stæhr (2008) and Milton, Wade and Hopkins (2010). In addition, participants in Study D were asked for any language proficiency test results obtained within twelve months, with 29 participants providing a TOEIC score and 35 a TOEFL score. The correlations between these scores and the LexCombi scores are also, however, very weak: .14 for the TOEIC scores and -.02 for the TOEFL scores. Finally, Studies B, C and D all made use of Yes/No tests and yet contrasting results were found for the correlation with LexCombi scores. It seems unlikely therefore that the differences in the measures of proficiency can account for the contrasts in the correlations between LexCombi scores and proficiency.

4 Differences between the versions of LexCombi
As mentioned in the introduction, Studies A to D were carried out in the course of further developing LexCombi. There were therefore different versions of LexCombi used. These differences are with respect to the format of LexCombi, its scoring and its cues.

Format: Study A followed Barfield’s original format for LexCombi, in which the cues were presented in a list and three spaces given after each cue for responses. Studies B to D used a revised format, in which each
cues. In the course of further developing Nesselhauf (2005) do

Finally, Studies B, C and D all made use of Yes/No tests and yet contrasting results were found for ‘


to the contrast in correlations. The one difference that does coincide with the contrasting correlations is with respect to the cues. It therefore seems likely that the difference in the cues primarily accounts for the contrast in correlations and it may be that Barfield’s cue selection procedure was biased towards selecting cues that differentiate between learners of different proficiency levels.

Two points may be said to arise. First, the paper shows the necessity of carefully reviewing the details of studies that provide contrasting findings. Some of the contrasts between the findings of the studies mentioned in the introduction may be a consequence of such differences. Second, the paper suggests that for the learners concerned here, Japanese university students lying broadly between low intermediate and high intermediate levels of proficiency, there is no discernible relationship between their general proficiency in English and their ability to produce collocations.

5 Conclusion

This paper has reviewed five studies which show contrasting correlations between a measure of knowledge of collocations, LexCombi, and proficiency. There were a number of differences between the studies, but in most cases these do not correspond with the contrast in correlations. The one difference that does coincide with the contrasting correlations is with respect to the cues. It therefore seems likely that the difference in the cues primarily accounts for the contrast in correlations and it may be that Barfield’s cue selection procedure was biased towards selecting cues that differentiate between learners of different proficiency levels.

Two points may be said to arise. First, the paper shows the necessity of carefully reviewing the details of studies that provide contrasting findings. Some of the contrasts between the findings of the studies mentioned in the introduction may be a consequence of such differences. Second, the paper suggests that for the learners concerned here, Japanese university students lying broadly between low intermediate and high intermediate levels of proficiency, there is no discernible relationship between their general proficiency in English and their ability to produce collocations.

References


Investigating Learner Performance on Productive Vocabulary Knowledge Tests

Jon Clenton
Hiroshima University
jclenton -at- hiroshima-u.ac.jp

Tess Fitzpatrick
Cardiff University
FitzpatrickT -at- cardiff.ac.uk

1 Introduction

This paper explores what tests of productive vocabulary knowledge capture. The paper compares learner performance on four tests that elicit productive vocabulary knowledge using: (i) a word association task (Meara and Fitzpatrick, 2000); (ii) a composition task (Laufer & Nation, 1995); (iii) a brainstorm task; and, (iv) a gap fill task. All tests assess knowledge according to word frequency and by sampling/ not eliciting pre-determined items. We conduct three studies in which we investigate whether the tests produce comparable measures of learners’ productive vocabulary knowledge.

2 Study 1

We begin by comparing learner performance on two tasks: The Lexical Frequency Profile (LFP) (Laufer and Nation, 19995), and Lex30 (Meara and Fitzpatrick, 2000). Incomplete examples of both tasks are shown in Figure 1.

Table 1 shows the difference in learner performance on the two tasks. There is a lack of a significant correlation between performance on the LFP and Lex30: $r = 0.108, p = 0.339$. Potential reasons for this lack of correlation might relate to the discursive nature of LFP, such as the demands of register, topic, and cohesion. We explore this claim in our second study.

3 Study 2

For our second study we introduce the ‘BFP’ (the Brainstorm Frequency Profile). The BFP was designed to explore whether performance on a non-discursive equivalent of the LFP task correlates more strongly with Lex30 scores than the LFP composition task. A completed example of the BFP is shown in Figure 2.

Table 1. Lex30 and LFP Scores ($n = 80$).

<table>
<thead>
<tr>
<th>Test</th>
<th>Mean score</th>
<th>SD</th>
<th>Min score</th>
<th>Max score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lex30(%)</td>
<td>43.63</td>
<td>5.89</td>
<td>29.41</td>
<td>57.83</td>
</tr>
<tr>
<td>LFP (%)</td>
<td>10</td>
<td>3.14</td>
<td>0.69</td>
<td>11.22</td>
</tr>
</tbody>
</table>

Example of a completed Brainstorm Frequency Profile (BFP) task:

"Should a government be allowed to limit the number of children a family can have? Discuss this idea considering basic human rights and the danger of population explosion."

Figure 1. Example Excerpts from Completed LFP & Lex30 Tests.

"Write down the first four (English) words you think of when you read each word in the list. Incomplete example:"

1. attack - fight, punch, kick, body
2. board - ski, snow, kick...

Figure 2. A Completed Brainstorm Frequency Profile Task
Table 2. Lex30 and BFP Scores \((n = 80)\).

Table 2 shows the difference in learner performance on the two tasks. There is no significant correlation between performances on BFP and Lex30: \(r = 0.211, p = .061\). Potential reasons for this lack of correlation might relate to differences in elicitation techniques (single vs. multiple activation cues; pool of lexical items exhausted during BFP task).

4 Study 3

Our third study introduced G_Lex a sentence-based gap fill task. The G_Lex task was designed to explore whether performance on a task using multiple activation events, using syntactic & semantic cues, correlate more strongly with Lex30 scores than either the LFP or the BFP (both use a single prompt). Figure 3 shows an extract from a completed G_Lex task.

Figure 3. Example Excerpt from Completed G_Lex Task.

Table 3 shows the Lex30 and G_Lex task scores. There is a significant correlation between performances on G_Lex and Lex30: \(r = 0.645, p < .01\). The significant correlation suggests that performance on one task is a moderate predictor of performance on the other.

5 Conclusion

These findings enable us to consider systematic differences and similarities in the knowledge captured by different tests of productive vocabulary knowledge.

6 Acknowledgements

We would like to thank the students at Osaka University for taking part in the studies presented in this paper.

References


Using a Word Knowledge Framework for the Analysis of L2 Written Errors

Andrew Gallacher
Kyushu Sangyo University
gallacher -at- ip.kyusan-u.ac.jp

1 Background

Nation’s (1990) Word Knowledge Framework has provided the foundation for a number of studies concerning vocabulary acquisition and testing. However, it has seldom been used to analyze student-produced written work. Addressing this gap in the literature, Gallacher (2015) used word knowledge types as a framework for analyzing students’ productive knowledge of high-frequency vocabulary (1,000 ~ 3,000 band on the BNC). Specifically, Gallacher examined students’ ability to demonstrate the Meaning and Usage (the remaining word knowledge types combined) of target words within written compositions called Mastery Sentences; which are essentially sentences that indicate target word meaning through the utilization of multiple aspects of word knowledge. Results showed that, for each level of proficiency, students were better able to demonstrate meaning than usage of target vocabulary. However, at the mid and high levels of proficiency meaning errors continued to differ while usage errors began to plateau (see Figure 1).

![Figure 1. Usage and Meaning Errors by Proficiency.](image)

2 Present Study

The present study builds on Gallacher’s (2015) previous work by unpacking the 'Usage' category of productive word knowledge into its’ individual components. Thus, Meaning, Written Form, Grammatical Behavior, Collocation/Association and Frequency were treated as separate categories of error in the analysis of students’ Mastery Sentences. Quantitative analysis revealed a hierarchy of difficulty between error types that was consistent among varying levels of English proficiency, while qualitative analysis uncovered patterns of mistake underlying each word knowledge-based error type.

3 Results

Table 1 shows the total number of errors coded for each word knowledge type for both participants. Overall, the high-proficiency student had fewer errors than the low-proficiency student in all but one category: grammatical behavior, where they had an equal number of errors.

Pearson’s Chi Square and Likelihood Ratio tests of all four variables were statistically significant (p<.0001) for both students, indicating that the observed differences in error rates were due to differences in difficulty between word knowledge types. Among the variables, a hierarchy of difficulty was observed for the high-level student with the number of errors increasing from least-to-most accordingly: written form > meaning > grammatical behavior > collocation. This pattern was not the same for the low-level student, where meaning proved slightly more difficult for them than grammatical behavior. Subsequent Pearson’s Chi Square and Likelihood Ratio tests were conducted between adjacent categories to confirm this hierarchy by showing that neighboring knowledge types differed in difficulty from one another as well (see Table 2).
by showing that neighboring knowledge types differed in difficulty between word knowledge types. Among the variables, a hierarchy of grammatical behavior, where they had an equal number of errors.

Table 1 shows the total number of errors coded for each word knowledge type for both participants. The difference between written form and meaning, the easiest and second easiest categories, was significant or very near significant (Chi Square of .0539) at the P >/= .05 level for both students. The difference between grammatical behavior and collocation, the hardest and second hardest categories, was also statistically significant for both. However, neither student demonstrated statistical significance for the middle variables, grammatical behavior and meaning. This suggests that these two types of word knowledge are close in terms of difficulty for these particular students and lie somewhere between the outliers, written form and collocation. Therefore, the overall hierarchy of difficulty between types of word knowledge observed for both students, is as follows: written form > meaning/grammatical behavior > collocation.

Qualitative assessment of the participant’s written samples yielded patterns of mistake unique to each level of proficiency. For the low-level student, ‘r/l’ distinction was an issue, as was basic syntax, obscurity in writing and dependency on high frequency vocabulary. For the high level student, complex syntactical patterns (continuous tense, passive constructions, etc.) and morphology of high level vocabulary (adjectival use in particular) proved difficult. Collocation and the use of formulaic expressions in English were challenging for both, and for the most part avoided altogether.

4 Conclusion

Using a word knowledge framework to analyze students’ written productions proved to be a useful means of identifying the difficulties participants had in demonstrating their productive knowledge of meaning, the written form, grammatical behavior, collocation, and frequency of lexical items. Unfortunately, the limited sample size used in this study restricts the ability to generalize these results onto a greater population. Therefore, it is hoped that future research will help substantiate these preliminary findings with a larger sample size.

References


Using Overlapping Checklists to Calculate the Vocabulary Size of Japanese Learners

Aaron Gibson
Kyushu Sangyo University
aaronlgibson -at- gmail.com

1 Introduction

Vocabulary is an important component to any language-teaching environment. It is necessary for teachers to know the vocabulary size of students so as to be able to recommend a course of study. The aim of this research is to calculate the vocabulary size of students using vocabulary checklists.

One of the most basic ways of obtaining difficulty measures of vocabulary is through yes/no checklists (see Anderson & Freebody, 1983). Checklists can be analyzed with Item Response Theory models such as the Rasch model (see Bond & Fox, 2007), which allows for item difficulty measures and person ability measures for a group of participants to be taken which are independent of the items being tested.

2 Method

A sample was taken of over 1000 students from a large private university in Japan. Students were in low-level compulsory English courses with average TOEIC bridge scores between 80 and 150.

Students were given a 120-item Yes/No checklist test. This was distributed in their regular English class at the beginning of the semester. Multiple versions of the test were distributed and linked together with common items so as to test over 720 words. The words used were selected from the 2000 most frequent words in English. Thirty anchor items were used to link the test together with previous years.

3 Results

Difficulty estimates of the checklist items were calculated using Rasch with the program WINSTEPS (Linacre, 2015). Person estimates were then used to calculate how many words examinees were expected to know of the entire set. In total the combined data collection sampled over 3000 students and 1600 items with 30 anchor items. The data had both high item and person reliability (> .97). The results show students’ vocabulary levels and also gaps in students’ knowledge of high frequency words.

4 Discussion

Teachers can use a similar format to conduct their own vocabulary size estimates for their students. This is part of a larger project to find estimates of the entire set of the 3000 most frequent words in English for Japanese learners.

5 Acknowledgements

This research was partially funded through JSPS Kakenhi Grant number 15K16810 (Young Scientist B).

References


The Mnemonic Effect of Assonance in Collocations: A Replication Study

Michael Green
Cardiff University / Kansai Gaidai University
GreenMC1 @ cardiff.ac.uk

1 Introduction

Becoming proficient in a second language (L2) involves acquiring a significant number of multi-word expressions such as binomials, collocations, idioms and sentence stems, here subsumed under the umbrella term formulaic sequences (FS) (Wray, 2002). Such FS are ubiquitous and their omnipresence makes them an essential component of mature linguistic competence, but target language idiomaticity is problematic even for highly proficient learners. One potential avenue for helping learners remember FS is to exploit the phonological patterning found in language. Previous classroom-based experiments with L2 learners have indicated a mnemonic advantage for collocations showing alliteration and assonance (e.g. Boers & Lindstromberg, 2005; Lindstromberg & Boers, 2008). An approximate replication of the second experiment in Boers et al. (2014) was undertaken and provides further empirical evidence of a mnemonic effect of assonance in English phraseology. This replication assesses the validity and generalisability of the previous findings and changes the parameters of the original study in terms of L1 and level of proficiency.

2 Participants

In the original study (Boers et al., 2014), the participants consisted of 44 language major undergraduates in their first year at a university in Belgium. Their first language (L1) was Dutch and their English proficiency was at the B2 level of the Common European Framework of Reference for Languages (CEFR), that is, at an ‘upper intermediate’ level.

In contrast, the participants in the replication study were all Japanese L1 undergraduate students in Osaka, Japan. In the initial data gathering there were 125 participants (81 female and 44 male) and their ages ranged from 18 to 19. Their mean TOEFL ITP score was 459 (N = 115, SD = 40.7), which approximately correlates with the B1 level of the CEFR. Thus, these participants were at a lower level of proficiency than those in the original study. The participants formed a convenience sample of five intact classes and, as in the original study, the experiment was conducted in their scheduled class time with their regular class teacher.

3 Materials

A set of twenty two-word phrases was compiled for the original study and the same phrases were used in the replication. The phrases consisted of either adjective+noun collocates or noun+noun compounds. Ten of the phrases showed assonance (e.g. town house) and ten were non-assonating (e.g. town square). The phrases were matched as far as possible to control for confounding variables such as syntactic structure, length, frequency and concreteness of meaning. To control for novelty effects, all the word types were high frequency; in the top 2000 word families of the British National Corpus with 31 tokens in the first 1000 (K1) and 9 tokens in the second 1000 (K2) list.

4 Procedure

The experimental phase consisted of a dictation activity followed by a series of unannounced recall tests. In both the original study and the replication, the classroom teacher began by explaining what assonance was and giving examples on the board. There was then a dictation exercise in which the randomised target phrases were read aloud by the teacher, the students repeated the phrase aloud, wrote it down and indicated by means of a tick or plus sign if the phrase displayed assonance, or a cross or minus sign if it did not. The full set of expressions was read out again, in a different order, and the participants checked their judgements. The dictation papers were then collected and a free recall test was immediately administered; on a new piece of paper the participants wrote down as many of the dictated phrases they could remember. The papers were collected and normal class resumed. Approximately one hour later, a second free recall test followed, in which the students again wrote down as many of the complete phrases as they could.

Because of timetable constraints, the replication study differed slightly from the original study for the final stage of the procedure. In the original study, there was a one-week delayed cued recall test: the
following week, the teacher read out each of the ten shared words of the twenty phrases and the students wrote down the remainder of the phrase, or both words if they could. For example, if the teacher said the cue town, it was hoped this would elicit house and/or square as the written response. Due to differing class schedules during the replication study, some of the participants (N = 45) took a two-day delayed cued recall test, while the remainder took a one-week delayed cued recall test (N = 68). In both cases, the test format was exactly the same as the original study.

5 Results

In the original study, participants recalled more assonating items than non-assonating items at a greater than chance likelihood across all three post-tests. A paired-samples t-test showed these differences in recall to be statistical: in both the Immediate Recall Test and the 1-hour Delayed Recall Test, a medium effect size (d = .45) was observed. In the 1-week Delayed Cued Recall Test, a large effect size was observed (d = 1.15).

In the replication study, boxplots indicated the presence of five outliers and skewed distributions. The skewness ratio indicated a significant departure from a normal distribution. As this violates the assumptions for t-tests, a non-parametric test was used (Wilcoxon Signed Ranks test). On average, participants recalled more assonating phrases than non-assonating control items across all the post-tests. In the Immediate Recall Test, participants recalled more assonating phrases (M = 3.58, SD = 1.76) than non-assonating (M = 3.12, SD = 1.86, z = 3.12, p = .002) and a medium effect size was observed (r = .32). Similarly, in the 1-hour Delayed Recall Test more assonating phrases (M = 3.34, SD = 1.87) than non-assonating phrases (M = 2.72, SD = 1.72, z = 3.92, p < .001) were recalled and a medium-large effect size was observed (r = .41). In the 2-day Delayed Cued Recall Test more assonating phrases (M = 2.96, SD = 1.59) compared to the non-assonating controls (M = 2.14, SD = 1.56, z = .61, p = .54) the difference was not statistical and the magnitude of the impact of assonance on recall was very small (r = .08).

6 Discussion and Conclusion

The results partly support the original findings and thus there is greater confidence that they are not artefacts of the methodology or sampling variation. The collocations with assonance were recalled to a greater degree than those with no phonological pattern. The difference was statistical for the immediate, 1-hour delayed and the 2-day delayed tests, and the effect sizes ranged from medium to large. The disconfirming results of the one-week delayed test should not be dismissed out of hand. The p-value is an unreliable measure of evidence and there was an effect size, albeit very small. The difference between the large effect size found in the original study (d = 1.15) and the smaller one here (r = .08) could be due to affective factors such as motivation, in that there was insufficient interest aroused by participating in the experiment.

The aim of this approximate replication study was to increase the explanatory power and generalisability of the findings from Boers et al. (2014). The experimental outcome, particularly for the immediate, one-hour and two-day delayed tests, suggests that assonance in L2 phrases does facilitate recall if the phonological pattern has been made explicit. Thus, converging lines of evidence do seem to indicate a mnemonic effect for phonological patterning and this appears to be a useful tool to add to the classroom teacher’s toolkit.

References


Updating and Validating the Vocabulary Learning Questionnaire

Peter Yongqi Gu
Victoria University of Wellington
peter.gu -at- vuw.ac.nz

1 Introduction

The Vocabulary Learning Questionnaire (VLQ), first developed and published in Gu and Johnson (1996), has been one of the most adapted instruments on vocabulary learning strategies. The latest version was VLQ5 (Gu & Hu, 2003) which is now 13 years old. In addition, despite the existence of an English version, validation of the VLQ was done only with the Chinese version. This paper reports on the validation of a recent ESL version of the VLQ.

The VLQ makes use of a 7-point Likert scale, ranging from Extremely untrue of me (1) to Extremely true of me (7). Likert scales that specify a bipolar continuum of a few points (e.g., 3, 5, or 7 points) have become standard practice in survey research. However, the ordinal nature of the data thus obtained has received ongoing debate in terms of statistical analysis. With the increasing ease of online research tools such as Qualtrics, a slider bar between bipolar ends (0-100) may provide continuous data with finer grains than traditional Likert scales, which, as a result, may bring an alternative, if not a solution to the problem. Hence a second aim of the project is to create a slider bar version of the VLQ and see how it performs against the Likert scale version.

2 Procedures

<table>
<thead>
<tr>
<th>Version</th>
<th>Purpose</th>
<th>Procedures</th>
</tr>
</thead>
<tbody>
<tr>
<td>VLQ6.0</td>
<td>Create English version of VLQ5</td>
<td>Find all English statements in previous versions and match them with the 90 statements in VLQ5.</td>
</tr>
<tr>
<td></td>
<td>Make an ESL version</td>
<td>Go over each statement and simplify them. Aims:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Most words should fall within the first 2000 most frequent words</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Include as few as possible academic words and off-list words.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Include as few as possible meta-language jargons.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Where jargon can’t be avoided, give simple examples.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Use simple sentence structures</td>
</tr>
<tr>
<td>VLQ6.1</td>
<td>Trial and fine-tune for clarity</td>
<td>• Get ESL teacher feedback re clarity of statements</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Trial 1, one advanced level ESL user, to ensure clarity of statements</td>
</tr>
<tr>
<td>VLQ6.2</td>
<td>Create online slider bar version; Trial among a small group of students</td>
<td>• Trial 2: four ESL students, think-aloud for both paper version and online versions. To</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• catch understanding issues and simplify statements further, and</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• catch potential similarities and differences for Likert scale and slider bar versions.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Re-write problematic statements after trial 2.</td>
</tr>
<tr>
<td>VLQ6.3</td>
<td>Formal piloting</td>
<td>• Trial 3: Among a larger group of ESL learners</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Statistically compare Likert scale version with slider bar version</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Catch potentially problematic items</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Catch problems with online administration</td>
</tr>
<tr>
<td>VLQ6.4</td>
<td>Construct validation</td>
<td>• Administer questionnaire among a large group of participants</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Exploratory factor analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Do the data support the existing structure of the questionnaire?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Can the questionnaire be shortened?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Should certain categories be deleted?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Can some categories be combined?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Can some items be deleted?</td>
</tr>
</tbody>
</table>
One hundred and five (105) students in the English Language Proficiency (EPP) programme at Victoria University of Wellington have answered the Likert scale version of VLQ6.3. Twenty five of these students answered both the Likert scale version and the slider bar version. Four students completed the think-aloud process while they tried both versions. Data collection is still going on to obtain enough data for construct validation (VLQ6.4). This paper reports on preliminary findings to the following two questions:

1. What are the psychometric properties of the ESL version (VLQ6.3) based on the 105 answered questionnaires?
2. Does the online slider bar version produce continuous data?

3 Preliminary findings

The small sample size (n=105) prevented the use of exploratory factor analysis (EFA) for the whole questionnaire (Costello & Osborne, 2005). However, EFA was performed on each subcomponent of the questionnaire (e.g., rehearsal strategies) to examine the underlying structure of that component. This resulted in the reduction of the questionnaire from 90 to 76 items. Two repetition strategies (oral and visual repetition) were collapsed into one (Repetition). Two encoding strategies (visual and auditory encoding) were combined into a new strategy (Perceptual encoding). Another two (meaning-oriented vs use-oriented note-taking) were collapsed into one (Note-taking). Next, Cronbach’s alpha was obtained to examine the internal consistency of each strategy. The overwhelming majority of the alphas obtained were satisfactory beyond .70, with the exception of Semantic encoding (alpha=.531). Since the sample size so far is too small for confirmatory factory analysis (CFA) (Gagne & Hancock, 2006), CFA results will be reported at a later stage after the sample size reaches an acceptable level.

Despite the subjects’ preference for the slider bar version as offering more accurate and more refined answers (think-aloud data), the 0-100 slider bar scale was shown to be practically more ordinal than interval, with most data clustering around tens and fives (e.g., 0,5,10,15,20,25,30 …).

4 Summary

An ESL version of the Vocabulary Learning Questionnaire has been created. Preliminary findings suggest that the questionnaire is performing satisfactorily in terms of construct validity and internal consistency. Despite a wider range of answers and the possibility of fine-grained choices, the slider bar version of the VLQ does not result in continuous data.

References


1 Introduction

Knowledge of vocabulary is a crucial component to both communication and comprehension for L1, L2, or foreign language learners (Hunt & Belgar, 2005). L2 vocabulary instruction and learning are often emphasized because vocabulary knowledge is considered predictive and reflective of reading achievement (Godev, 2009) and lack of vocabulary knowledge is identified as a serious obstacle to ELLs in content-area classes (Nam, 2010). Vocabulary instruction involves the careful selection, sequencing, and presentation of words to promote the acquisition of words (Nation & Newton, 1997). One of the most powerful tools to facilitate vocabulary selection is the use of corpus data to find information about the frequency and lexical patterning of words (Schmitt, 2010). Despite the existence of free online corpora tools, Keck and Kim (2014) argue that the majority of L2 teachers and L2 textbooks have not made use of corpora or corpus-based research. In this classroom action research, we describe a novel approach to vocabulary selection and word list creation through the use of the Corpus of Contemporary American English (COCA) with specific attention to tracking vocabulary exposure in two adult ELL classrooms at an IEP: one low-beginner listening and speaking (LS) class and one high-beginner LS class. This study aims to elucidate the following questions:

- Do modified words lists promote vocabulary acquisition?
- Is there a relationship between frequency of exposure via modified word lists and LS students’ vocabulary accuracy and production?

2 Background

Researchers and practitioners alike agree that vocabulary acquisition is shaped by many factors such as how word knowledge is measured, level of word knowledge, and the selection of words to teach. Nation (2007) notes that frequency and range, as well as the ability to combine, define, and replace other words are important factors of vocabulary selection. The use of a corpus is an effective tool to acquire this information, however, Römer (2009) found that fewer than 2% of language teachers in her study consulted corpora for vocabulary instruction. Little has been done to close the gap between corpus research and pedagogical practice and word lists based on the needs of L2 or L3 speakers continue to be ‘scarce’ (August et al., 2005; Smart, 2014). Sibanda and Baxen argue that the formation of corpus-informed word lists seems to have stagnated, “compelling teachers to rely on the lists that are already available…without due regard to the corpus from which they were created, and the unit of word used for their generation” (2016, p. 59). In spite of these observations over the last decade, it is still unclear how teachers can effectively use their available resources (such as corpora and corpus-informed texts) to choose lexical items. Consequently, classroom action research intervention is needed to incorporate these findings into practice.

3 Methods

Modified vocabulary instruction was carried out in a low-beginner LS class (N=6) and a high-beginner LS class (N=8) over a 9-week period. Both classes used the Touchstone™ textbook series, which were developed using the multi-billion word Cambridge International Corpus (Cambridge University Press, 2016). Because this corpus is not publically available, the instructors further modified the textbook word lists using COCA, a 520-million-word collection gathered from written and spoken texts that is free and available to the public online (Davies, 2016). The word list creation process is detailed below.

- Using COCA, the instructors created modified word lists derived from the vocabulary lists in the classroom textbooks. The parameters of word frequency, register, and applicability to classroom activities were used to guide the selection of words.
• Word lists were organized according to meaning, function, and relation to task/assessment and given to the students in the form of a paper vocabulary list. The list was visually organized so that each group or section of words would correspond to a particular kind of exercise or activity (e.g. the words in a square box required a preposition, the words in the circle are hobbies).

• The students interacted with the list throughout the duration of a unit. The activities (individual, pair, and group) were carefully planned with specific attention to explicit (e.g. dictionary definition) and implicit (e.g finding patterns, using pictures or gestures, use in a story) vocabulary exposure using a variety of traditional materials and technology-enhanced media.

This research is currently in progress and the data will be collected and analyzed using the free online software, Vocabprofile (Cobb, 2016). In line with Horst and Collins (2006), data from LS student production in speaking and writing tasks and unit quizzes will be compared with data from course syllabi, such as implicit and explicit activities and vocabulary lists measuring exposure frequency and task type.

4 Results and Discussion

Vocabulary acquisition research indicates that both instruction and instruction materials should be informed by corpus data (Nation, 2007; Schmitt, 2010; Keck and Kim, 2014). While corpus-informed textbook word lists are convenient to use, this research provides evidence for promoting the process of generating modified vocabulary lists from textbook word lists to facilitate vocabulary acquisition. Our results tentatively indicate that needs-based, modified word lists benefit students in terms of both written and spoken production and accuracy.

References


Academic Vocabulary in English for Chinese Medical Purposes

Cailing Lu  
Victoria University of Wellington  
Cailing.Lu -at- vuw.ac.nz

Philip Durrant  
University of Exeter  
P.L.Durrant -at- exeter.ac.uk

1 Introduction

In order to address the vocabulary needs of students who are learning English for academic purposes, different sets of academic word lists have been created. Currently, two of the most widely used academic word lists are Coxhead’s (2000) Academic Word List (AWL) and Gardner and Davies’ (2014) Academic Vocabulary List (AVL). A growing body of research is now investigating the validity of these lists, and in particular how relevant they are to the needs of learners in particular disciplines (Chen & Ge, 2007; Durrant, 2016; Martínez, Beck, & Panza, 2009; Valipouril & Nassaji, 2013). To date, most discussion has centered around the utility of the AWL. However, there are reasons to believe that problems of this sort may exist for any list of general academic vocabulary. Traditional Chinese medicine (TCM), as a significant and growing field which was not included in the corpora on which these lists were made, offers an interesting test case in this context.

Another reason to conduct this research is that English for Chinese medical purposes (ECMP) academic courses are compulsory courses for both undergraduates and postgraduates in Chinese medicine in China due to the internationalization of TCM. These interdisciplinary courses aim to bridge the gap between TCM learners’ content knowledge and their English-language knowledge so that TCM learners can keep up with cutting-edge TCM research reported in English or report their own research in English. A key challenge facing students of ECMP is that of mastering sufficient vocabulary to meet their study needs. However, few attempts have been made to address the vocabulary needs of these TCM learners. Therefore, the present study makes a preliminary attempt to fill this research gap.

2 Method

This study investigated the utility of the AWL and AVL for learners of ECMP. A corpus-based approach was adopted to examine the usefulness of Coxhead’s (2000) AWL, and Gardner and Davies’ (2014) AVL in the field of traditional Chinese medicine. To determine the usefulness of the academic word lists, the coverage and frequency of AWL and AVL items in a custom-built TCM corpus (also called the TCM Corpus) were calculated for the corpus as a whole. To understand the nature and importance of other vocabulary in this corpus, a list of high-frequency words which were not found on the GSL, AWL or AVL was also created and its coverage evaluated.

3 Results

The AWL was found to cover 10.64% of tokens in the corpus, while the AVL was found to cover 21.17% overall. In both cases, the majority of the coverage was achieved by a relatively small subset of the items on the lists. This suggests that both of these academic word lists can provide valuable information for course planning and individual lexical learning.

<table>
<thead>
<tr>
<th>1</th>
<th>patient</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>acupuncture</td>
</tr>
<tr>
<td>3</td>
<td>cell</td>
</tr>
<tr>
<td>4</td>
<td>Chinese</td>
</tr>
<tr>
<td>5</td>
<td>clinical</td>
</tr>
<tr>
<td>6</td>
<td>symptom</td>
</tr>
<tr>
<td>7</td>
<td>score</td>
</tr>
<tr>
<td>8</td>
<td>acupoint</td>
</tr>
<tr>
<td>9</td>
<td>therapy</td>
</tr>
<tr>
<td>10</td>
<td>liver</td>
</tr>
</tbody>
</table>

Table 1. Top 10 Off-list Words in the TCM Corpus
In addition to academic vocabulary, we also took a close look at vocabulary which is prominent in Chinese medicine but does not appear on any of the word lists used in this study. Analysis of these most frequently used off-list words in the TCM Corpus shows that a small number of such words achieve a high level of coverage, suggesting that they should be given a great deal of attention by learners in this area. All together, the findings indicate that a discipline-specific listing would be of great benefit to learners in this area. Table 1 provides a list of the top 10 off-list words.

References
Vocabulary Learning Through Cooperatively Structured Art-Based Tasks

Steve McGuire
Nagoya University of Arts
spm -at- nua.ac.jp

1 Introduction

Bush (2007) noted that visual images have been used in language learning and since antiquity, and Barnes-Karol and Broner (2010) offer an idea for use of visuals for the learning of culture, but while he has used images to elicit particular forms, there is little specific research extant regarding language learning specifically through the use of art. Additionally, there has been little research on cooperative learning carried out in the field of foreign language learning and even less in the Japanese context (see Fushino, 2008) with none of that research reporting specifically on the degree to which language was learned. This study is the first to examine vocabulary acquisition through artwork based on a pedagogical approach called Visual Thinking Strategies (VTS). VTS has been used to help students acquire visual literacy and critical thinking skills through the use of three questions, “What’s going on in this picture?” “What do you see that makes you say that?” and “What more can we find?” (Housen, 2001; Housen, 2002; Desantis & Housen, 2004).

2 Method

Using an exploratory mixed methods approach, data was collected as 80 Japanese college students in 20 cooperative groups in two classes over one semester chose, used, shared, and learned why carrying out cooperatively structured think-pair-share activities adapted from VTS. Students' written interactions, including the VTS worksheets they used as they carried out the art-based tasks, were also all input using Moodle and Google Forms, and oral interactions were recorded using CHIeru, a computer classroom management system. Students also filled out a course-final questionnaire regarding their perceptions of their learning experiences. Vocabulary data was analyzed using AntwordProfiler.

Vocabulary learning was assessed through a pretest and posttest on a Vocabulary Usage Inventory (VUI) in which students assessed their level of knowledge regarding a list of vocabulary at the 3,000 baseword level and above, derived from their written answers to the three VTS questions regarding the artworks. Part 1 of the VUI was adapted from the Vocabulary Knowledge Scale; the categories ranged from a complete lack of knowledge of a word, having seen a word before, being able to translate into Japanese, and being able to write a sentence. In Part 2 students wrote translations in Japanese and in Part 3 they wrote English sentences using the words in reference to the artwork.

3 Results

The analysis showed that while there was variation in the numbers of tokens, types, and word families for the artworks, the percentage of vocabulary elicited at the 3,000-word frequency level and above between the four groups was extremely similar, varying by only .4%, as was the data within each group for each of the artworks. Native speakers of Japanese and English raters were in agreement at 98.1% regarding students’ self-assessments in the respective parts of the VUI of their ability to write translations and sentences. The comparison between an art drawing task carried out individually and collaboratively showed that the students who carried out the task individually did as well or better than those who did the task collaboratively.

Analysis of the data showed significant improvement in students’ vocabulary knowledge between the pretest and posttest. Unfortunately, a Latin Square design comparing the vocabulary learning of students who carried out a drawing task found that students who worked with the textual descriptions of an artwork individually learned vocabulary as well or better than students who carried out the tasks cooperatively. However, the course-final survey found students’ perceptions to be positive regarding the cooperatively structured art-based tasks, regarding working together in cooperative groups, and regarding how frequently and how well they employed the skills the teacher felt were necessary to carry the tasks.

Students’ use of vocabulary regarding the artworks was extremely idiosyncratic, even for artworks that were similar in composition and content except for a few words that were clearly connected to a particular artwork (cactus for an image depicting a desert, for example). This means that it is not possible for teachers to predict what vocabulary students might use for an artwork. However, as mentioned above, the kind of...
The vocabulary used was consistent and predictable, although this needs to be explored further. The course-final survey, although filled out only by a small percentage of students, showed positive opinions regarding working in groups, especially the chance to share and hear other students’ opinions regarding the artworks, but students also liked the chance to interact with students who they hadn’t previously worked with, even with students who were in the same course of study.

4 Conclusions

First, the data suggests that students using these tasks will express, be exposed to, and process vocabulary at high levels with perceivable patterns in a rich environment that Nation’s Four Skills Approach (2007), Technique Feature Analysis and the Vocabulary Knowledge Framework (among others) (Nation & Webb, 2011) tell us are beneficial for language learning.

Second, we know from the VTS worksheets and from analysis of students’ interactions that they shared much of the vocabulary they used in their answers to the VTS questions regarding the artworks, leading to sharing of a rich, broad range of vocabulary.

Third, this study showed that cooperative learning can help students to feel comfortable sharing their ideas and, perhaps more importantly, interact cognitively with the ideas of others. The study didn’t show an advantage in use of cooperative learning as it was measured in the drawing tasks, but it did show that the level of interaction with ideas and vocabulary in the cooperative learning groups was very high. The focus on the cooperative skills raised students’ awareness of their degree of success in working together.

Finally, the students’ high level of participation in all of the tasks and their positive responses to the survey questions show that this approach shows great promise for language learning. Even students’ negative responses offer insight into areas that need to be attended to improve these tasks so they work even better.

The poster session will present resulting data, share the artwork used in this study (and the criteria for selection), will show the Visual Thinking Strategy Worksheet, a sample transcript of students’ interactions as they carried out the tasks, and the course-final survey and resulting data. Visitors will learn how to use art in a foreign language class and will see the rich degree to which even lower proficiency foreign language learners use language in regards to the artwork and the degree to which learners acquire language from these interactions.

References


The Picture Vocabulary Size Test

Paul Nation
Victoria University of Wellington
Paul.Nation -at- vuw.ac.nz

Laurence Anthony
Waseda University
anthony -at- waseda.jp

1 What is the Picture Vocabulary Size Test?

The Picture Vocabulary Size Test (PVST) is a test of receptive vocabulary size. The test measures whether the test-taker can find a suitable meaning (a picture) for a given partly contextualised word form. It is a recognition test primarily intended for young pre-literate native speakers up to eight years old and young non-native speakers of English. Two versions of the test have been designed each with 96 test items and following the same design procedures. The test was designed by Paul Nation and implemented as a software application by Laurence Anthony. Jannie van Hees played an important role in trialling the test. Figure 1 shows the latest design of the application.

2 What were the reasons for making the PVST?

Although there has been a lot of research on the vocabulary knowledge of young native-speakers from different socioeconomic groupings (Farkas & Beron, 2004; Rowe, Raudenbush & Goldin-Meadow, 2012), there has been little research on the total number of words that young learners know. Do five year olds know only a few hundred words when they begin school or do they know several thousand? How quickly and at what rate do the vocabulary sizes of young learners grow? Do young girls race ahead of young boys in their vocabulary growth? Can teachers have a significant impact on learners’ vocabulary size by deliberately targeting vocabulary knowledge? These are important questions that can directly affect what teachers do. The PVST provides an indication of how many words a learner knows.

3 What can the PVST be used for?

The test is intended to be used for formative assessment, particularly diagnostic assessment, where performance on a task such as reading or participating in content-focused oral discussion may be affected by vocabulary size. Because the PVST is a largely decontextualised measure of receptive vocabulary knowledge to some degree isolating vocabulary knowledge from other skills, learners' performance on the test can be used as an indication of whether vocabulary knowledge is likely to be a major factor accounting for poor performance on the task. Thus, the test could be used along with diagnostic procedures for examining learners’ reading, listening, speaking, and writing skills.

The test is based on the most frequent 6000 word families of English for young native-speaking children. As a result, a ceiling effect is likely to be observed if the test is used with native-speakers of English who are more than eight years old as they are likely to have a larger vocabulary size than that shown by the test.

4 How do you administer the PVST?

The PVST works on an ordinary computer, a touch-screen computer, or a tablet. The learner touches the "Listen" button for each question and then hears a sentence which also appears in its written form. The learner then has to choose one picture from four pictures which best matches the spoken sentence. The learner touches the picture to make the choice. The learner then touches an arrow to move to the next screen, or the program can be set to automatically move to the next screen.

The learner can listen to a sentence again by re-touching the "Listen" button or can touch a different picture to change the choice. If the learner is not sure of an answer, they can touch the "Don't know" button instead of a picture. There is no time limit on the test. It takes around 15 minutes to sit the whole test, but experience shows that some learners may take as long as 25 minutes, agonizing over items and giving it their best shot. One-on-one administration of the test makes sure that learners remain engaged and do not employ random guessing. When administering the test, it is useful to explain the following points to learners.
They should look at all four pictures when considering choices.
They can push the "listen" button again to hear the question item again.
They can change their mind about a choice, by touching a different picture.

The administrator should be watching each learner to ensure they perform each procedure where needed.

5 How should you interpret the results of the PVST?

Each test word represents 62.5 words in the source lists. So, a learner’s score on the test needs to be multiplied by 62.5 to get their total vocabulary size. Thus, a learner with a score of 54 has a receptive vocabulary size of 3,375 word families. The program makes this calculation in the results it produces. The rough rule of thumb for native-speakers’ vocabulary size is to take 2 away from the learner’s age and multiply by 1000 to find their likely vocabulary size. So, a seven year old is likely to have a receptive vocabulary size of around 5000 word families. This rule of thumb fits with Biemiller’s (2005) data from young Canadian children. There can be wide variation of vocabulary size at any particular age level.

Table 1 roughly summarises and rounds off Biemiller’s (2005: 233) findings for young native speakers at school. It is worth noting that even the lowest scoring native speakers at most levels knew several thousand words. Because the test is a sensitive receptive recognition test, it gives credit for partial knowledge. This means that getting a word correct on the test does not necessarily mean that the word is well known.

Table 1. Average Native-speaker Total Vocabulary Sizes for Various Age Levels

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Size</th>
<th>Range</th>
<th>Type of Text</th>
<th>Font Size</th>
<th>Style</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 year olds</td>
<td>3,000 wf</td>
<td>1,000-5,000</td>
<td>9 year olds</td>
<td>7,000 wf</td>
<td>4,000-11,000</td>
</tr>
<tr>
<td>6 year olds</td>
<td>4,000 wf</td>
<td>1,500-5,000</td>
<td>10 year olds</td>
<td>8,000 wf</td>
<td>5,500-12,000</td>
</tr>
<tr>
<td>7 year olds</td>
<td>5,000 wf</td>
<td>2,000-8,500</td>
<td>11 years olds</td>
<td>9,000 wf</td>
<td>7,000-13,000</td>
</tr>
<tr>
<td>8 year olds</td>
<td>6,000 wf</td>
<td>3,000-10,000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

wf = word families

6 What are the strengths and weaknesses of the PVST?

Strengths
- The PVST provides an estimate of a learners total vocabulary size, not just knowledge of a particular set of words.
- The PVST can be used with learners who cannot read.
- The PVST has good face validity.
- The PVST is quick to administer and scores itself
- The PVST is easy to interpret.
- Children enjoy sitting the test.
- The test shows that it is possible to test most words using pictures.

Weaknesses of the test
- The test measures up to the most frequent 6000 words leading to a ceiling effect with older learners.
- As with all multiple-choice tests of vocabulary knowledge, answering an item correctly does not mean that the word is necessarily well known or can be used.
- The test needs to be administered one-to-one to ensure learners are doing the best they can. This makes administration to large numbers of learners very time-consuming.
- The use of multiple-choice items allows guessing, which can inflate scores. The "Don't know" option only partially deals with this problem.
- Some of the items require more interpretation of the pictures than others. For example, the item testing flit requires the learner to work out that a butterfly moves in this way. Items like table or grass, require no such inferencing. Surprisingly even the five year olds seemed to have no great difficulty performing this inferencing.

References


Understanding Vocabulary Demands in Factual Textbook Texts

Kimberly Skjelde  
Knarvik Upper Secondary School  
kimskj3 -at- hfk.no

Torill Hestetræet  
University of Stavanger  
torill.hestetreet -at- uis.no

1 Introduction

The purpose of this study was to conduct an in-depth investigation of general academic vocabulary usage in English Language Teaching (ELT) course material (Skjelde, 2015). Previous studies conducted in Norway show that students struggle with comprehension when reading academic texts (Hellekjær, 2008, 2012). Textbooks and their accompanying websites remain a major source of reading material in Norwegian English as a Foreign Language (EFL) classrooms (Hestetræet, 2012; Juuhl, Hontvedt, & Skjelbred, 2010). However, there is a lack of corpus studies conducted on textbooks written for advanced EFL learners in Norway. Several international studies investigate overall use of general academic vocabulary in textbooks (Matsuoka & Hirsh, 2010; Ruegg & Brown, 2014). Through the use of a small corpus of factual textbook texts, the current study aimed to examine the extent to which general academic vocabulary use may be expected to promote the implicit acquisition of these word families during unassisted reading (Skjelde, 2015).

2 Research Methods and Data Collection

The study was conducted as a mixed methods corpus study in which general academic vocabulary was operationalized through Coxhead’s (2000) Academic Word List (AWL). The AWL word families were charted and also subdivided into British National Corpus (BNC) and Corpus of Contemporary American English (COCA) frequency levels. The underlying assumption was that students could be expected to struggle with form-meaning knowledge of AWL word families used less frequently in the BNC and COCA corpora.

The corpus was comprised of 21 factual texts from three textbooks widely used in Norwegian high school EFL classrooms. The distribution of texts from each textbook included five tailored texts i.e., written for the textbooks, and two authentic texts found on the textbooks websites, but not written for ELT. In total, the corpus contained 28,734 tokens. The small size of the corpus allowed for an in-depth study of each text, the manual analysis of glossary use in tailored texts, and a comparison of textbooks and texts related by topic.

Data collection was based on the use of three computer programs and manual analyses of the corpus. The three computer programs were: VP-Classic, VP-Compleat and Range (Cobb, n.d.). The VP-Classic analyses provided AWL coverage and frequency rates for each text. Range analyses were used to provide further insight into frequency rates, but now across topic-related texts. A minimum frequency rate of six repetitions defined the study’s cut-off rate for implicit vocabulary acquisition through unassisted reading. The VP-Compleat analyses determined BNC/COCA frequency levels for each text as a whole and were also used to divide the AWL vocabulary into BNC/COCA frequency levels. Placing AWL word families into BNC/COCA frequency levels provided a deeper understanding of glossary use in the tailored texts. Text analyses with the VP-Compleat program also provided data regarding lexical coverage. Lexical coverage was defined as “what percentage of the vocabulary in a stretch of spoken or written discourse needs to be known by a learner in order for him or her to understand the discourse” (N. Schmitt et al., 2015, p. 2). Again, the assumption was that reading comprehension would be difficult for texts using a larger percent of vocabulary at higher BNC/COCA frequency levels.

3 Findings

Findings generally support previous research in the field (Cobb, 2010; Matsuoka, & Hirsh, 2010; Nation 2006). The corpus contains an average AWL coverage rate comparable to that expected in English language newspapers, at 5.3%. However, a majority of the tailored texts (67%) have rates below this 5% level, and only 3% of the texts use AWL rates close to the 8% AWL coverage expected in academic texts (Coxhead, 2000; Nation, 2013). Differences between tailored and authentic texts were also found. Half of the authentic texts have AWL coverage rates at an academic level of 8%, but none of the tailored texts reached this level.

Just under 3% of the in-text AWL word families were recycled six or more times. Even over three and four topic-related texts, a majority (60%) of the AWL vocabulary was only used once.

Glossary coverage rates i.e., the number of all glossed terms in relation to the total number of tokens in the
text, were below Nation’s (2013) recommended 3% - 5% in 60% of the texts. The extent of glossed AWL words in the 15 tailored texts varied from 0 to 27%, showing that a majority of the AWL word families were not used in a glossary. The authentic texts did not contain glossaries. A clear majority (around 80%) of the AWL word families not found in the glossaries were also used only once in the text and many of these (60% - 80%) were above the 2000 BNC/COCA frequency level.

To achieve the desired 98% lexical coverage rates for adequate reading comprehension (Laufer, 2010), students would need a vocabulary size up to and including the BNC/COCA 9000 level in a majority of the analyzed texts. However, a larger number of authentic texts reached 95% lexical coverage with vocabulary up to and including the BNC/COCA 3000 level, despite having a higher average use of AWL vocabulary.

4 Conclusion

By conducting an in-depth study of AWL vocabulary usage in a small corpus, it has been possible to examine whether implicit acquisition of general academic vocabulary can be expected during unassisted reading. The findings clearly indicate that Norwegian EFL students exposed to these textbook texts cannot be expected to learn a majority of the AWL vocabulary implicitly, but will need support in this learning process. It may be expected that students will receive help with salient terms during reading in the EFL classroom. However, this study suggests that students will also need assistance to acquire AWL word families as they are often used only once in a text and are commonly left out of glossaries. The findings also indicate that the use of authentic texts in the EFL classroom may provide better exposure to AWL vocabulary without compromising student reading comprehension.

References


Using the Reflective Portfolio to Teach Writing and Vocabulary Acquisition

Peter Tall
Zayed University
Peter.tall -at- zu.ac.ae

1 Introduction

The reflective portfolio is seen as an essential classroom tool within many academic disciplines, and has gained popularity within the ESL world over recent years.

What is it, and why should we use it? “The learning portfolio is a flexible evidence-based process that combines reflection and documentation. It engages students in ongoing reflective and collaborative analysis of learning.” (Zubizarreta, 2004) When applied to writing and vocabulary acquisition, the reflective portfolio comprises a collection of graded and drafted papers, kept either physically or electronically.

2 Using the reflective portfolio with writing and vocabulary acquisition

With traditional grading methods, marked work is often given a cursory glance by the student, then placed in quickly forgotten piles of papers or the wastepaper basket. Spelling and meaning errors notated by the teacher are often perused by the student and then subsequently forgotten.

The reflective portfolio method requires the student to actively acknowledge mistakes (spelling, meaning, grammar, format, and style) and provide corrections. In effect, the reflective portfolio is “creating significant learning experiences” (Zubizarreta, 2004). Solutions are not simply provided by the teacher, but must be discovered and provided by the student. In this way, the student takes possession of the work. When effort has been applied to find corrections to mistakes, the student is far more likely to retain the knowledge.

During the reflective stage of the process, the student is asked to write a reflection upon what has been learned during the exercise, and to acknowledge mistakes, and explain solutions which he/she has devised.

The student is also required to file and retain the work (either electronically, or physically), and is graded on the portfolio as a whole. Completed work may therefore not simply be discarded, but is kept as a permanent resource, able to be referenced at any time.

3 Results

It has not been possible to obtain quantitative data to show that, for instance, “class A,” which used the reflective portfolio, gained better results than “class B,” which did not. In any case, such studies would not be valid, since students and classes are not uniform. Such a comparison also would be difficult to justify ethically. On a qualitative level, the students seem to enjoy the process and appreciate that it helps them to progress. The teachers (many of whom showed initial suspicion), generally agree with their students, and see the reflective portfolio as not just another classroom task, but a positive and helpful learning tool.

4 Conclusion

“Typically, techniques are chosen because they represent ways of presenting language material which the teacher feels are going to do the most good for the largest number of students” (Brown, 1995).

Teaching methods are often a matter of taste (where choice is given). Where we have that choice or a portfolio is required, then the re-drafting and reflection method of learning (for writing and vocabulary acquisition) can be an effective and interesting alternative to traditional “mark and forget” methods.

The metacognitive nature of learning with the reflective portfolio can be seen as a “step away from a narrow cognitivist view of education” (Quicke, 1994). Students take possession of the work and, thereby, are more likely to see value in it. “The greater the value that individuals attach to the accomplishment of or involvement in an activity, the more highly motivated they will be both to engage in it initially and later to put sustained effort into succeeding” (Williams & Burden, 1997).

The reflective portfolio can be an effective method for reinforcing general writing study and vocabulary acquisition, rather than simply another classroom chore imposed upon teachers and students.
References


Using Type-token Area Indices to Assess Lexical Diversity

Dax Thomas
Keio University
daxmthomas -at- gmail.com

1 Introduction

Lexical diversity (LD) is generally recognized as the relationship between the total number of words (tokens) and the total number of distinct words (types) in a text. The traditional measure of this relationship is the type-token ratio (TTR). TTR, however, is known to be problematic (Malvern et al., 2004). It is highly dependent on text length and does not take into account fluctuations in the rate of new vocabulary introduction within a text. While numerous alternatives to TTR, such as voc-D, HD-D, and MTLD (McCarthy & Jarvis, 2010), to name only a few, have made progress towards a more accurate measure of LD, these problems still remain.

Type-token curves (TTCs; Youmans, 1990) are a graphic representation of the running type-count plotted at each token. These are a useful means of studying LD in that they deal with the problem of text length and vocabulary introduction rate fluctuations by giving the researcher a bird’s-eye view of the text’s LD, but limit the researcher to a visual (non-numeric) analysis. Additionally, the number of texts that can be compared at one time in a single line-graph is limited by graph crowding (Thomas, 2005).

This paper explores the potential for using the type-token area index (TTAI) to assess LD and attempts to address two main concerns: 1) the rate of new type introduction in a text; and 2) text length.

2 The type-token area index

The two key features of TTAI are: 1) the use of area-under-the-curve calculations; and 2) the use of a “normative” reference curve. Area-under-the-curve is useful as it takes into account the full shape of the curve and, therefore, any fluctuations in the rate of new type introduction throughout the text.

The problem of text length is addressed with the use of a reference curve. In this study, a 300,000-token reference curve was generated from the average running type counts for ten long public domain books. This reference curve was then used as a constant against which to compare the curves of other texts.

A program, written in the Python language, counts the text’s types and tokens, creates a TTC, and calculates the area under the curve. It then calculates the area under the curve for a segment of the reference curve that is the same length as the text being examined. The area value of the text being examined is divided by the area value of the reference curve segment, thereby obtaining the TTAI.

3 Testing TTAI on novels

As an initial test of the metric, ten public domain books were analysed. The books ranged in length from 50,000 to 75,000 words and differed from those used to create the reference curve. The results of the TTAI analysis are shown in Table 1.

<table>
<thead>
<tr>
<th>Filename</th>
<th>TTA Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>willows.txt</td>
<td>1.097</td>
</tr>
<tr>
<td>worlds.txt</td>
<td>1.093</td>
</tr>
<tr>
<td>room.txt</td>
<td>1.083</td>
</tr>
<tr>
<td>sawyer.txt</td>
<td>1.076</td>
</tr>
<tr>
<td>sailing.txt</td>
<td>1.061</td>
</tr>
<tr>
<td>frank.txt</td>
<td>1.040</td>
</tr>
<tr>
<td>mars.txt</td>
<td>0.999</td>
</tr>
<tr>
<td>prisoner.txt</td>
<td>0.948</td>
</tr>
<tr>
<td>jungle.txt</td>
<td>0.930</td>
</tr>
<tr>
<td>treasure.txt</td>
<td>0.920</td>
</tr>
</tbody>
</table>

Table 1. Type-Token Area Indices (TTAI) of Ten Novels.

A TTC comparison graph (not pictured here) was then generated and compared with the TTAI data as an informal check of the validity of the indices. The indices seem to match reasonably well with the general curve positions for each text in the graph.

To explore this in more detail, the TTCs for four of these books were calculated: the book with the highest
TTAI (Wind in the Willows at 1.097, included for scale), and the three books with the lowest TTAIs (Prisoner of Zenda, 0.948; The Jungle Book, 0.930; and Treasure Island, 0.920). These four TTAI’s are shown in Figure 1. While traditional TTR would assign Prisoner of Zenda the lowest LD, below the LD’s of both Treasure Island and The Jungle Book, visual analysis of their curves indicates that for much of its length, the curve for Prisoner of Zenda is steeper than the other two, dipping and crossing below the other lines only near the end of the text. This would seem to indicate a higher overall LD for this novel than traditional TTR would assign. The TTAIs for these three books corroborate this, assigning Prisoner of Zenda a higher LD than the other two (Table 1).

![Type-Token Curve Comparison](image.png)

Figure 1. TTCs of Four Novels.

4 Discussion

At this stage, TTAI seems to work well in overcoming some of the problems associated with lexical diversity analysis. Some of the other advantages of TTAI include 1) freedom from a reliance on sample segments drawn randomly from the text (as with vocd-D; McCarthy, 2007), and 2) versatility in use, either for direct comparison with other texts, or as a stand-alone metric, with the figure representing the text’s divergence from the norm (over 1 indicating a higher LD than the norm and under 1 indicating a lower LD).

Much work remains to be done with regard to formal construct validity tests of TTAI. Moreover, comparison studies with current LD metrics, such as vocd-D, HD-D, MTLD, still need to be conducted. More experimenting is also warranted in the selection of texts that are used for constructing the reference curve.

5 Conclusion

TTAI has potential application in the analysis and comparison of LD in classroom textbooks and graded readers, where vocabulary recycling is an important concern, as well as in the study of written and transcribed spoken work by students, and teacher talking time in the classroom.

References


Teaching Form and Usage of Discipline-Specific Vocabulary in an EAP Program

Marie-Claude Toriida
University of Calgary in Qatar
mcmtorii -at- ucalgary.edu.qa

1 Introduction

Corpus linguistics has had a big impact on the way that English programs approach vocabulary teaching and learning in the last few decades. Thanks to its development, it is now understood that not all words are created equal, with some words being more frequent than others in any given language. This knowledge helps in efficiently teaching and acquiring a new language. Many frequency lists have been compiled and extensively used by English language teachers since West’s (1953) seminal General Service List. However, some researchers, such as Hyland and Tse (2007), have pointed out that learning general and academic vocabulary from lists does not address the fact that many words have different meanings and usages in different fields, and that teachers must “recognize that students in different fields will need different ways of using language” (Hyland and Tse, 2007, p. 249). While teachers in many English for Academic Purposes (EAP) programs do not have the luxury of preparing students for a single major, those who do are in a position to address this point for the benefit of their learners.

2 Context and rational

This vocabulary project (Shimoda, Toriida, & Kay, 2016) was initiated at the University of Calgary in Qatar to address the vocabulary gaps of students enrolled in a Nursing Foundation EAP program leading to acceptance into a four-year Bachelors’ of Nursing, or a two-year Post Diploma Bachelor of Nursing Program. An IELTS score between 4.5 and 6 is needed for acceptance in the three-tiered EAP program. A further placement test determines each student’s starting EAP level. Students are mandated to complete the EAP program in 3 semesters (one year) or less.

Linguistics gaps were initially identified by both nursing and EAP instructors as being the greatest barrier to student success in nursing courses. Shortcomings in vocabulary knowledge hindered the students’ ability to read and comprehend nursing textbooks and class content. Another observation was that students in both streams had fairly different vocabulary needs. Given this context, an efficient way to select and teach context specific nursing vocabulary was deemed necessary to best prepare students in the limited time available.

3 Corpus making

It was decided that identifying and consistently introducing vocabulary from nursing textbooks used at our institution would be a best practice approach. A corpus was developed from Fundamentals of Nursing: The Nature of Nursing Practice in Canada (Kozier et al., 2003), an introductory nursing textbook extensively used across first year courses and in some second year courses at our institution.

After obtaining an e-book copy and converting it to a Microsoft Word document, a first round of word elimination was performed. The purpose of this step was to remove words that were not considered essential to understanding nursing content. References, repetitive textbook headings and vocabulary, chapter sections (e.g., suggested readings), and names of agencies are some examples of words that were eliminated. This data was saved as text files for later analysis. Using Laurence Anthony’s AntConc (2012), a freeware concordance program, a frequency word list was then generated. It was examined to identify non-content words that were not properly eliminated. A second round of word elimination was then performed to address problem entries and a final frequency list was produced.

Using Microsoft Access, the Fundamentals of Nursing (FON) Corpus was compared to the GSL (West, 1953); the New General Service List (NGSL) (Browne, Culligan, & Phillips, 2013); and the AWL (Coxhead, 2000). The FON corpus was composed of 152,642 word tokens (total word count) and 6,820 word types (the number of different words). It included 82% of the NGSL words, 71.9% of the GSL, and 88.9% of the AWL. The first 2500 high frequency words included 54% of the NGSL, 45% of the GSL, 58.4% of the AWL words, and covered 94.6% of the contents of the textbook.
The development of a vocabulary list from this corpus, with a focus on meaning and use in the field of nursing, was agreed upon to be the best study tool for students. After compiling the FON corpus, an annotation of the first 2500 words was made in a table format (25 lists of 100 words). Using a standardized process for listing, each word entry included: the frequency in the corpus, the word form, a definition, collocation(s), and a sample sentence. The first step of the process was completed using AntConc’s concordance function. The highest frequency word form for each entry, as used in the textbook, was identified by looking at all sentences in which a word and its corresponding lemma word forms occurred. When two word forms were often used, both were listed along with their definition and sample sentences.

The next step included listing a definition for each word entry. Context specific definitions were taken from the Cambridge Learner’s Dictionary Online or the web-based Medical Dictionary (http://medical-dictionary.thefreedictionary.com). A definition was also given when a high frequency collocation of a word was found to have its own meaning. Following the definition, two- to four-word high frequency collocations were identified and listed for each word entry using AntConc’s Cluster/Ngram function.

Providing sample sentences from the text completed the final stage of the annotation process. Sentences were shortened and simplified for clarity and easier understanding. As much as possible, an attempt was made to use lower frequency words in sample sentences to facilitate comprehension and to build in a repetitive review of previous words.

A vocabulary acquisition program was put in place using spaced repetition in the form of the hand computer flashcard method (Mondria & Mondria-De Vries, 1994), and the computer-based Anki program. It was considered important to have a method that would be individualized, continuous, and independent of EAP level. In our program, words are self-selected by students at a rate of 10-15 per week. Flashcards include the key word on one side, and the part of speech, definition, collocation(s) and sample sentence on the other. Arabic can also be used on this side. Testing is individual and done once a week, or once every other week on approximately five randomly selected words. Students are asked orally to identify the part of speech and to explain the meaning. They are also asked to produce a collocation and/or give an example of how the word is used in a sentence. This time is used to clarify meaning of words and how they are used in the field of nursing. Testing is cumulative during the term, and continues across all EAP levels. Students are also asked to keep a written log of words selected at each level, and this is photocopied at the end of term and passed on to the next level teacher. 5% to 10% of the Academic Reading course grades are for direct vocabulary learning.

Acknowledgements

We are greatly indebted to Dr. Charles Browne for his teachings and his continued guidance and support throughout this project. We would also like to thank Professor Akano at Kyoto University of Foreign Studies for introducing us to AntConc and for his individual guidance on how to use it.

References

Building Learner Autonomy and Success Through Vocabulary: A Case Study from a Japanese High School

Lori Zenuk-Nishide  
Kobe City University of Foreign Studies  
zenuk-nishide -at- inst.kobe-cufs.ac.jp

Jody Shimoda  
University of Calgary in Qatar  
jshimoda -at- ucalgary.edu.qa

1 Background

English language learners (ELL) often struggle with comprehension due to untargeted vocabulary skills instruction. Additionally, government authorized texts are difficult for learners to read as they do not provide enough coverage of the most frequent words and contain short reading passages that do not allow learners to meet words often enough to facilitate vocabulary acquisition (Browne, 1996; 1998). The results of a vocabulary levels test at a Japanese high school revealed that 55 students lacked knowledge of many high frequency words. In order to help students overcome barriers in written and oral communication, increase comprehension, and enhance skills development, a research-driven three-year individualized vocabulary and extensive reading program was developed. The goal was to increase students’ willingness to study vocabulary and assist them in acquiring it more quickly and systematically, enabling them to more fluently and accurately express themselves in English.

2 Methodology

As part of the program, students were taught how to choose words, make and study vocabulary cards, and select graded readers with 95-98% word coverage. Word cards were created (Nation, 2001) using West’s General Service List (GSL) (West, 1953), and Coxhead’s Academic Word List (AWL) (Coxhead, 2000). Cards were then studied using Leitner (1972) and Mondria and Mondria-DeVries (1993) “Hand-Computer Method” based on Landauer and Bjork’s (1978) “Expanded Rehearsal Strategy.” As there is a complementary relationship between learning from decontextualized word cards and the contextualized exposure received through reading graded readers (Nation, 2008), the direct study method was run synchronously with an extensive reading program to enhance opportunities for incidental vocabulary learning (Waring and Nation, 2004). Assessment was done using the Vocabulary Levels Test (VLT) (Nation, 1993; Schmitt, Schmitt, and Clapham, 2001) to measure word knowledge at the 1,000 to 10,000 levels in addition to the AWL. Students were also given individual productive word card tests every six weeks on the self-selected words they were studying.

3 Results

Non-parametric statistics were used for data analysis. The results of the Wilcoxon signed-ranks test on the six levels of the VLT showed that there was a significant increase in all levels after nine months, from September 2005 to March 2006 at the 1,000 level (z=-3.140, p=<.01), 2,000 level (z=-3.070, p=<.01), 3,000 (z=-2.706, p=<.01), 5,000 level (z=-2.365, p=<.01), 10,000 level ( z=-2.706, p<.01), and AWL (z=-3.199, p<.01). Forty-seven percent of the students completed studying the GSL and 10% completed the AWL and were studying words from other sources. Table 1 shows an increase over three years in the GSL and AWL word levels that students were studying deliberately through the Hand-Computer Method.

<table>
<thead>
<tr>
<th>Year</th>
<th>1</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>144</td>
<td>1257</td>
</tr>
<tr>
<td>Maximum</td>
<td>1257</td>
<td>off-list</td>
</tr>
<tr>
<td>Mean</td>
<td>544</td>
<td>2270</td>
</tr>
</tbody>
</table>

Table 1. GSL/AWL Word Levels.

The findings revealed significant improvement at all frequency levels of the VLT and increased extensive reading levels over the three years of the study (Zenuk-Nishide et al., 2005, 2006; Shimoda & Johnston, 2007; Zenuk-Nishide (2011)). Given this, it is important for teachers to plan, give strategy instruction, test, and teach vocabulary both deliberately and incidentally (Nation, 2008) in order to provide a meaningful learning experience, increase student autonomy, and create a path for success.
References
# Author Index

<table>
<thead>
<tr>
<th>Name</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albright, S. W.</td>
<td>173</td>
</tr>
<tr>
<td>Alves, M.</td>
<td>107</td>
</tr>
<tr>
<td>Anthony, L.</td>
<td>19, 21, 146, 199</td>
</tr>
<tr>
<td>Elgort, I.</td>
<td>82</td>
</tr>
<tr>
<td>Ewert, D.</td>
<td>153</td>
</tr>
<tr>
<td>Eyckmans, J.</td>
<td>37, 53</td>
</tr>
<tr>
<td>Fitzpatrick, T.</td>
<td>10, 21, 184</td>
</tr>
<tr>
<td>Gallacher, A.</td>
<td>186</td>
</tr>
<tr>
<td>Gesa, F.</td>
<td>123</td>
</tr>
<tr>
<td>Gibson, A.</td>
<td>188</td>
</tr>
<tr>
<td>Gilabert, R.</td>
<td>47</td>
</tr>
<tr>
<td>Gonzalez-Fernandez, B.</td>
<td>135</td>
</tr>
<tr>
<td>Gonzalez, M. C.</td>
<td>107</td>
</tr>
<tr>
<td>Green, M.</td>
<td>189</td>
</tr>
<tr>
<td>Grogan, M.</td>
<td>84</td>
</tr>
<tr>
<td>Gu, P. Y.</td>
<td>191</td>
</tr>
<tr>
<td>Gyllstad, H.</td>
<td>29, 99</td>
</tr>
<tr>
<td>Han, S.</td>
<td>137</td>
</tr>
<tr>
<td>Henriksen, B.</td>
<td>86</td>
</tr>
<tr>
<td>Hestetræet, T.</td>
<td>201</td>
</tr>
<tr>
<td>Heylen, K.</td>
<td>139</td>
</tr>
<tr>
<td>Hirai, A.</td>
<td>76</td>
</tr>
<tr>
<td>Ho, K-J.</td>
<td>193</td>
</tr>
<tr>
<td>Horst, M.</td>
<td>119</td>
</tr>
<tr>
<td>Hsu, Y.</td>
<td>65</td>
</tr>
<tr>
<td>Hu, H. M.</td>
<td>155</td>
</tr>
<tr>
<td>Husky, K.</td>
<td>145</td>
</tr>
<tr>
<td>Ishii, T.</td>
<td>157</td>
</tr>
<tr>
<td>Jafari, N.</td>
<td>55</td>
</tr>
<tr>
<td>Kanazawa, Y.</td>
<td>32</td>
</tr>
<tr>
<td>Kasahara, K.</td>
<td>76</td>
</tr>
<tr>
<td>Kawakami, M.</td>
<td>129</td>
</tr>
<tr>
<td>Kawauuchi, Y.</td>
<td>63</td>
</tr>
<tr>
<td>Kestemont, B.</td>
<td>139</td>
</tr>
<tr>
<td>Kida, S.</td>
<td>29</td>
</tr>
<tr>
<td>Knight, D.</td>
<td>21</td>
</tr>
<tr>
<td>Kramer, B.</td>
<td>26, 57, 121, 127</td>
</tr>
<tr>
<td>Kremmel, B.</td>
<td>78</td>
</tr>
<tr>
<td>Kunikoshi, A.</td>
<td>145</td>
</tr>
<tr>
<td>Lafleur, L.</td>
<td>159</td>
</tr>
<tr>
<td>Laufer, B.</td>
<td>12</td>
</tr>
<tr>
<td>Lodi, A.</td>
<td>163</td>
</tr>
<tr>
<td>Lu, C.</td>
<td>91, 195</td>
</tr>
<tr>
<td>Lucovitch, D.</td>
<td>67</td>
</tr>
<tr>
<td>Matsushita, T.</td>
<td>41</td>
</tr>
<tr>
<td>Matte, B.</td>
<td>161</td>
</tr>
<tr>
<td>McGuire, S.</td>
<td>197</td>
</tr>
<tr>
<td>McLean, S.</td>
<td>26, 57, 121, 127</td>
</tr>
<tr>
<td>Miralpeix, I.</td>
<td>123</td>
</tr>
<tr>
<td>Miyata, Y.</td>
<td>63</td>
</tr>
<tr>
<td>Mochizuki, M.</td>
<td>45</td>
</tr>
<tr>
<td>Morita, M.</td>
<td>49</td>
</tr>
<tr>
<td>Morris, S.</td>
<td>21</td>
</tr>
<tr>
<td>Munby, I.</td>
<td>113</td>
</tr>
<tr>
<td>Naganuma, N.</td>
<td>43</td>
</tr>
<tr>
<td>Nakata, T.</td>
<td>117</td>
</tr>
<tr>
<td>Nation, P.</td>
<td>19, 23, 26, 34, 69, 199</td>
</tr>
<tr>
<td>Nishimura, Y.</td>
<td>49</td>
</tr>
<tr>
<td>Nomura, M.</td>
<td>143</td>
</tr>
<tr>
<td>Noreillie, A-S.</td>
<td>139</td>
</tr>
<tr>
<td>Nylander, E.</td>
<td>99</td>
</tr>
<tr>
<td>Oghigian, K.</td>
<td>146</td>
</tr>
<tr>
<td>Okumura, T.</td>
<td>23</td>
</tr>
<tr>
<td>Orenha-Ottaiiano, A.</td>
<td>59</td>
</tr>
<tr>
<td>Peters, E.</td>
<td>139</td>
</tr>
<tr>
<td>Pinchbeck, G. G.</td>
<td>26, 72</td>
</tr>
</tbody>
</table>
Author Index

R
Rashid, N. 178
Read, J. 94
Rosalia, C. 101
Rott, S. 70
Ruegg, R. 43
Ruiz Tada, M. 47

S
Sabino, M. A. 163
Sanhueza, C. 74
Schmitt, D. 17
Schmitt, N. 19, 90
Shimoda, J. 209
Shin, J-A. 137
Skjelde, K. 201
Stengers, H. 53
Stewart, J. 121, 127, 133
Stoeckel, T. 23, 133
Strong, B. 165
Stubbe, R. 167
Sugimori, N. 45
Suzuki, S. 169

T
Tall, P. 203
Tamura, Y. 49
Tanaka, N. 88
Tay, D. 93
Thomas, D. 205
Thomson, H. 97
Thwaites, P. 35
Tomei, J. 176
Tono, Y. 15
Toomer, M. 82
Toriida, M-C. 207

V
Vajirasarn, A. 171
Vasylets, O. 47
Vitta, J. P. 173

W
Wadden, P. 23
Wang, C. 93
Wang, F. L. 105
Waring, R. 111, 175
Webb, S. 39
Westbrook, P. 86

X
Xie, H. 105

Y
Yamauchi, Y. 145
Yin, Z. 69
Yoshii, M. 176

Z
Zenuk-Nishide, L. 209
Zou, D. 105
Notes
Please say “hello” and “thank you” to our friendly student volunteers!