Developing Students’ Vocabulary Knowledge in Content Subjects: 
A Computational Linguistic Approach

Kirk Dowswell 
College of Technological Innovation 
Zayed University 
Dubai, United Arab Emirates 
email: kirk.dowswell@zu.ac.ae

Abstract— Undergraduate students are exposed to discipline-specific lexis and concepts, particularly when studying in a second language. Current research suggests that most students find it difficult to fully comprehend academic reading material because they lack the requisite vocabulary, i.e., 5,000 to 8,000 word families for achieving 95% to 98% comprehension, respectively. It has also been suggested that teaching vocabulary explicitly is not an efficient use of classroom time. Thus, in order to enhance vocabulary acquisition and, ultimately, improve the reading comprehension skills of second language learners, this pilot study evaluated the use of a modified version of the Vocabulary Self-Selection Strategy (VSS+) as a self-directed learning tool. The study was conducted in an Arab higher education institution where undergraduate students studied Information Technology (IT) in English. It was anticipated that this unique intervention would improve vocabulary acquisition with minimal use of classroom teaching time. Results indicated that students were actively engaged with the wiki as a learning tool and there was a noticeable improvement in their vocabulary knowledge. Overall, the study has implications for teachers, as well as learners

Keywords: data-driven learning; vocabulary self-collection strategy; vocabulary learning; involvement load hypothesis; teaching with wikis.

I. INTRODUCTION

The ability to read academic texts in English is one of the most challenging issues facing second language students studying at English-medium institutions (EMI). Many Emirati students who have studied in Arabic-medium schools are not equipped with these skills and thus find it difficult to obtain direct entry into these institutions. Consequently, many federal EMI’s in the United Arab Emirates (UAE) require students who have not achieved an International English Language Testing System (IELTS) score of 5.0 attend pre-baccalaureate programs that focus on English academic literacy and language skills. Students can spend up to two years in these programs and can only progress by achieving the requisite IELTS score (5.0) and a pass in their final examination. IELTS broadly defines a Band 5 student as a modest user (IELTS, 2016).

Reading proficiency is assessed throughout these academic English courses and students are expected to have achieved a satisfactory level of academic literacy skills that will allow them to cope with the English language textbooks and content specific material encountered in their baccalaureate studies. At first glance, it would seem that most students do cope and have the language and academic literacy skills to meet subject requirements. However, feedback from content faculty and recent pre-graduation data in the institution where the study described in this paper took place, has revealed that over 60% of students who graduated in 2015, exited the institution with an IELTS level less than Band 6 on Reading. This indicates no or minimal improvement has been achieved by these students despite four years of subject specific study in the English language (Internal Dean’s Council Report, December 21st, 2015).

IELTS broadly defines a Band 6 student as a competent user (IELTS, 2016) and Band 6.5 as the entry level demanded by many universities in the UK; this is largely considered the minimum level that a student should possess to cope with the rigors of academic study. Based on these definitions of language competence, many of the English as a Second Language (ESL) students studying at the institution where this study took place are struggling to cope with academic texts and are graduating at a level of less than ‘competent’. Furthermore, the majority of these students are struggling with academic and technical language, once they have ceased to receive academic language support.

This situation stimulated several discussions at the institutional level about reading and the level of attainment, (as measured by the IELTS examination), that students need in order to function effectively and prompted the following questions:

1. Do ESL students studying at an EMI have sufficient academic literacy skills to be able to cope with their subject-specific reading material?

2. If not, what strategies can be employed to improve ESL students’ academic literacy skills in content courses?

The present study addressed these questions by first conducting a short survey with the IT content teaching faculty, asking how they taught academic reading and content specific vocabulary in the classroom and what problems the lack of technical vocabulary knowledge caused their students. The majority of faculty indicated that a significant number of students did, indeed, face difficulties, however, a number of faculty felt it was not their place to rectify them (Internal College of Technological Innovation’s Language Task Force Survey, January 2014). This inability of many content faculty
members to consider the difficulties faced by their students in their subjects and the students’ inadequate academic literacy skills led to the development of the next phase of this research project, described in this paper.

It was anticipated that by using contemporary text mining techniques to develop the students’ academic vocabulary and by rethinking our approach to the development of academic literacy skills, particularly in the area of vocabulary acquisition, we should be able to reduce the difficulties faced by ESL students when studying content and allow them to utilize the higher order thinking skills needed to function in today’s multi-literate society, as identified in Bloom’s Revised Taxonomy [1]. By employing these techniques, coupled with the implementation of “Language Across the Curriculum” strategies, the aim would be to develop the students’ Cognitive Academic Language Proficiency skills (CALP) [2] along with, what Volmer [3] calls, students’ ‘Conceptual Literacy’ and ‘Discourse Competence’ skills.

In Section 2 of this paper, the literature review will address the issue of developing vocabulary knowledge. Section 3 will discuss the methodology used to develop the vocabulary and finally Section 4 of the paper will conclude with a discussion of the results and future implications for this pilot project.

II. BACKGROUND

Academic research, [4], highlights the difficulties long-term (7 or more years of language instruction) second language learners have in reading academic texts. Difficulties they highlighted include those identified by Cummins [2] where students find it problematic to distinguish between spoken and academic language. The studies also provide some guidance on how these academic literacy skills, particularly vocabulary instruction, can be taught in content courses and thereby improve reading comprehension.

Research has demonstrated that there is a clear link between word knowledge and the ability to comprehend texts. This relationship was recently examined by Laufer and Ravenhorst-Kalovski [5] who not only suggested that increased vocabulary knowledge could lead to an improvement in reading comprehension, but also proposed two thresholds for text coverage and comprehension. This showed that for a student to understand 98% of a text, a knowledge of 8,000 word families is required and a knowledge of 4,000 to 5,000 word families for 95% coverage. This reinforces an earlier study by Hu and Nation [6] who proposed that for unknown vocabulary knowledge not to be a major hindrance to text comprehension, knowledge of roughly 98% of the lexis is required.

This awareness of the importance of academic discourse and, in particular, academic vocabulary is thus deemed necessary for students’ success in university study. Consequently, it is paramount that content faculty be made aware of interventions that could enhance students’ comprehension of academic texts. However, as the faculty survey revealed, many content teachers do not have the time or inclination to engage in strategies to improve their students’ text comprehension. Therefore, many faculty rely on basic glossaries available in the course textbooks and assume that students will make use of these. To rectify this situation, the present study incorporated data driven techniques to develop subject-specific keyword vocabulary lists and then created an intervention based on an extended version of the Vocabulary Self-Collection Strategy Plus (VSS+) [7], the Involvement Load Hypothesis (ILH) [8] and a class wiki.

The VSS+ and wiki, when combined with the ILH formed an important part of the vocabulary retention strategy. The Involvement Load Hypothesis (ILH), [8] an idea that postulates that words processed with greater learner involvement are retained longer than those processed with a lower involvement load. The construct, labeled ‘task-induced involvement’, incorporates the cognitive components of ‘Search’ and ‘Evaluation’ and the motivational component of ‘Need’. For example, the act of selecting a target word and finding its meaning demonstrates ‘Need’. When followed by the action of searching for the definition and translating it into Arabic, the ‘Search’ process is fulfilled. Finally, the evaluation of the ‘word sense’ or context meets the ‘Evaluation’ criteria. Such activities warrant a high level of engagement from the student on the Involvement Index scale (i.e., Presence of factors: No factor=0, Moderate=1, and Strong=2). This study evaluated the level of vocabulary knowledge development between two groups of students: a control group following the same course with no intervention and a group engaging in the wiki tasks which, ranked high on the Involvement Index scale.

III. METHODOLOGY

The aim of this research was to develop and apply a pedagogical framework for the teaching and learning of IT content-specific vocabulary. It employed a quasi-experimental research design to test the impact of the VSS+ framework and it explored the following research questions:

1. Is there a significant difference between the VSS+ wiki intervention group and the control group that received traditional instruction?
2. What do the participants in the VSS+ wiki group think about the VSS+ wiki framework as a means of learning vocabulary?

A. Participants

A total of 8 male and 21 female university students majoring in Information Technology at an EMI university in the UAE participated in the study. The students, all ESL learners with Arabic as their first language, ranged in age from 20-30 years old. A control group of 11 female students were taught by a separate instructor and followed the traditional course of instruction. The intervention group consisted of two intact groups: one female class of 10 students and one male class of 8 students who were taught using the VSS+ by myself. Of the students participating in the intervention group, 17 completed a pre- and post-vocabulary knowledge (VKS) test as well as a questionnaire. From the control group, only 5 students completed the pre- and post-VKS test.

B. The Intervention - Selecting Appropriate Lexis for the Content Course

Studies cited by Cobb [9] suggest that the first 2,000 most frequent words, coupled with the 570 word families in the
Academic Word List (AWL), can bring the coverage of an academic text up to approximately 90%. To increase comprehension to the minimum coverage of 95% the students will need to develop a word knowledge of 4,000 to 5,000 word families, [5].

The current debate on the benefits of rich vocabulary instruction, [10] as opposed to “genuine academic reading for the readers’ own purposes, [11] has prompted educators to explore what intervention would be more successful with ESL students. However, evidence provided, [12] showed that Arab learners find vocabulary acquisition extremely challenging primarily because a limited number of words in English are borrowed from Arabic and also because the Arab teaching pedagogy is traditionally based on rote learning and minimal engagement in extensive reading activities. Based on this evidence, the present study employed ‘rich’ vocabulary instruction strategies in the style of the VSS+ intervention as these would be deemed useful for Arab ESL students.

Having decided on the method of intervention, it was necessary to compile academic words specific to the pilot content course, i.e., IT in Global and Local Cultures. There are numerous definitions of academic vocabulary but, [13] state that academic words can be categorized into two distinctive areas: general and discipline-specific words. General words are used across disciplines whereas discipline-specific words tend to be used in specific disciplines. As learning discipline-specific words does not always guarantee full comprehension of discipline-specific texts, it also advisable to incorporate scaffolding techniques to make the text more meaningful, [10]. Thus, a corpus of key academic words (general and discipline specific) was created using the SketchEngine application, [14].

The use of corpora in language teaching and learning, sometimes referred to as ‘data-driven learning’ a model created by Johns (1990), as cited in, [15] has greatly simplified the process of analyzing language and enabled the creation of frequency lists based on the course textbook. Once the lists were created and analyzed, keywords were chosen based on the following criteria: relevance to subject, academic word list and, finally, frequency level as per the VocabProfiler, [16].

C. Vocabulary Self-Collection Strategy

The Vocabulary Self-Collection Strategy (VSS+) is an approach that can be used for “general, basal reading or content area development”, [17]. In this study, the focus was on content vocabulary development, with the main purpose being to develop students’ understanding of subject-specific words and concepts. The instructional strategies used to achieve this involved specific instruction of the reading followed by the students selecting key terms that have been identified in the corpus analysis of the reading. The process, however, was slightly adapted to incorporate technology and use techniques similar to those adopted by, [7] with their VSS+ framework. After reading the text and discussing the major concepts in class, the students used a wiki to develop a subject glossary based on the vocabulary they selected. This allowed them to explore the words in much greater depth, as

The rationale for the use of the wiki was twofold. Firstly, it was expected that the students would be involved in specific activities that required reading the text, creating an easily available glossary and engaging in word focused tasks that encourage deep learning. Secondly, the word tasks were collaborative and loosely followed, [18] the six steps of vocabulary instruction. It is also generally recognized that what learners do with words is an important part of vocabulary retention and it is necessary for learning tasks to involve the components of “need, search and evaluation”, [8]. It was, therefore, postulated that the adaptation of the VSS+ /ILH intervention using the wiki framework discussed above would enhance content vocabulary acquisition.

As discussed above, the Vocabulary Self-Collection Strategy (VSS+) and Involvement Load Hypothesis (ILH) approaches, were used in conjunction with the class wiki tool. The intervention took place over a period of twelve weeks or six teaching units. The students in the intervention group were instructed on the use of the wiki as an autonomous learning tool. A vocabulary template was developed for the students to investigate the meaning of their selected lexies (see Fig. 1). For each new topic covered in the course, students selected unknown vocabulary from the list of Key Words In Context (KWIC) words highlighted in the corpus or words they selected on their own from class readings. After reading the text and discussing the major concepts in class, the students used the wiki to develop a subject glossary based on the vocabulary they selected. Students in the class had the opportunity to edit the wiki entries as they saw fit. As an incentive, a small percentage of the final course grade was awarded for participation based on the number of wiki entries.
D. Data Collection Instruments & Procedures

The following two vocabulary tests were administered as pre- and post-tests to both the intervention and the control groups:

- **X-Lex**: an online test of vocabulary breadth that assesses how many words a student knows. The students in the current study were presented with a set of words from five different frequency levels (1k to 5k) one at a time, in a context-free environment. They simply decided whether they knew the meaning of each word. Based on their responses X-Lex developed a profile of vocabulary knowledge for each frequency band and then suggested which level each should be placed in, [19]. The test was completed by 13 students from the intervention group and 11 students from the control group.

- **Vocabulary Knowledge Scale, (VKS), [20]** a test of students’ knowledge of discipline specific vocabulary. This was adapted to test 130 prominent keywords extracted from the discipline specific corpus. The students in this study indicated their level of recognition of the words by selecting one of the options: a) I have never seen this word before; b) I have seen or heard of this word before; c) I think I can define this word; d) I am confident I can define this word. The test was completed by 18 students from the intervention group and 11 students from the control group.

The aim of the tests was to establish a vocabulary level for all students and to establish which of the 130 corpus keywords were known by both groups of students.

Additionally, at the end of the study, a questionnaire was used with the intervention group to assess the students’ perceptions on the use of the wiki as an autonomous learning tool. The questionnaire consisted of 8 statements and asked participants in the intervention group to rate each statement on a 5-point Likert scale (Agree, Strongly agree, Neither Agree nor Disagree, Disagree, Strongly Disagree). The questionnaire was administered at the end of the 12-week intervention and it was completed by 17 students.

IV. RESULTS & DISCUSSION

A. Pre- and Post-Test Comparisons

The first research question sought to investigate the effect of vocabulary instruction using the VSS+ with the intervention group and to compare the results with a control group who had received traditional vocabulary instruction.

The results were analyzed using a t-test and produced some interesting findings within the groups, but unfortunately limited information was obtained in the between-groups comparisons. This could be a result of the small amount of data available for the control group, many of whom failed to complete the post-test. It is planned to repeat the study with a much larger cohort of students and greater control over the testing procedures.

The pre- and post-test scores were calculated with both the raw scores and the percentage scores. No significant difference was found between pre- and post-VKS, between the control and intervention groups. However, the t-test results showed that the intervention group performed significantly better with regards to their knowledge of vocabulary (see Table 1), which, suggests that the use of the VSS+ strategy and the wiki improved their vocabulary knowledge. Although the pilot study did not provide conclusive results, it is evident

<table>
<thead>
<tr>
<th>VKS</th>
<th>Pre-Test (% of students)</th>
<th>Post Test (% of students)</th>
<th>Pre-Test (% of students)</th>
<th>Post-Test (% of students)</th>
<th>t-test score</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% Word recognition</td>
<td>% Word recognition</td>
<td>Mean</td>
<td>Mean</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VKS 1 (I have never seen this word before)</td>
<td>17%</td>
<td>10%</td>
<td>21.28</td>
<td>12.94</td>
<td>4.652</td>
<td>.000</td>
</tr>
<tr>
<td>VKS 2 (I have seen or heard of this word before)</td>
<td>9%</td>
<td>8%</td>
<td>11.89</td>
<td>10.67</td>
<td>.466</td>
<td>.647</td>
</tr>
<tr>
<td>VKS 3 (I think I can define this word)</td>
<td>12%</td>
<td>7%</td>
<td>15.00</td>
<td>9.00</td>
<td>2.848</td>
<td>.011</td>
</tr>
<tr>
<td>VKS 4 (I am confident I can define this word)</td>
<td>62%</td>
<td>74%</td>
<td>77.78</td>
<td>93.28</td>
<td>-9.520</td>
<td>.000</td>
</tr>
</tbody>
</table>
that using this method has the potential to enhance students’ vocabulary and ability to comprehend academic texts.

Overall, the results of the VKS support the view that the wiki was a useful tool for teaching vocabulary, when using the VSS+ and ILH strategy because the students were actively engaged in the process. Not only did the students research the meaning and ‘sense’ of the vocabulary items they selected, but they also translated each word into Arabic and added a pictorial representation, where possible. Additional test results also revealed that there were no significant statistical differences between the male and female students. Finally, they were also engaged in the glossary review process, which promoted collaborative learning and a sense of community within the class. All of the above would seem to support the conclusion that “What learners do with the word may be more important than how many times they encounter it.” [21].

B. Questionnaire

The second research question sought to discover what the intervention group thought about the class wiki and the VSS+ strategy, as a means of learning vocabulary. A questionnaire was used asking participants in the intervention group to rate a total of 8 statements on a 5-point Likert scale. The results of the questionnaire are shown in Table 2.

The feedback gathered from the questionnaire suggested that:

- Two thirds of the students in the intervention group found that a wiki is a useful vocabulary learning tool.
- 11 out of 17 students liked using the wiki as a way of exposing themselves to and learning new vocabulary.

**TABLE II. QUESTIONNAIRE RESULTS (N=17).**

<table>
<thead>
<tr>
<th>Statement</th>
<th>D</th>
<th>N</th>
<th>A</th>
<th>SA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The class wiki is a useful tool to practice new course vocabulary.</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>2. The class wiki has given me more exposure to new vocabulary.</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>3. I feel competent to peer review my classmates’ wiki entries.</td>
<td>4</td>
<td>0</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>4. Peer review of the wiki entries has been useful for vocabulary learning.</td>
<td>4</td>
<td>0</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>5. I felt insecure to make corrections to other student’s wiki entries.</td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>6. I think my motivation to the subject has now increased.</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>7. The wiki has improved the sense of community in the class.</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>8. I would like to employ the wiki in all my courses.</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

Note: SD = Strongly Disagree; D= Disagree; N=Neutral; A= Agree; SA = Strongly Agree

- A further 10 students felt competent in peer reviewing their classmates’ entries and felt that it was a useful learning exercise.
- Only one student felt insecure with editing other students’ work.
- More than half of the students felt that their motivation increased and that it improved their sense of community in class.
- Just over one third of students would like to use a vocabulary wiki in all their classes.

By and large, the wiki proved to be functional as a teaching tool and was used throughout the semester for a variety of activities. It was particularly useful for collaborative activities, once the students had become used to the many features available. However, in the future, it would be advisable to employ the wiki solely as an online glossary that can be easily accessed and utilized, as a result of the finding that some initial difficulties were encountered when it was being used for other purposes.

The use of the wiki as the main technological interface appears to have been well received, with 11 out of 17 students indicating that is was a useful tool to learn vocabulary. However, the interface design was not as easy to manipulate as had been anticipated and valuable time was spent training students on the process for creating and editing the wiki glossary. The word template, although easy to use, required quite a few ‘clicks’ to achieve the end result and, thus proved to be frustrating for some students. As the activities were associated to the Involvement Load Hypothesis, [8] the students were required to click on a variety of links, to seek out the required information and then enter the results into the template, thus making the whole process quite time consuming.

V. CONCLUSIONS

The main purpose of this pilot study was to determine the effectiveness of the VSS+ and ILH strategy, when used in conjunction with a wiki, as a means of enhancing the acquisition of content specific vocabulary by undergraduate Arab students studying at an EMI in the UAE. Each of the research questions sought to determine the level of effectiveness of this intervention. While the results from the first research question were not conclusive owing to the limited number of students in the control group completing the tests, the results for the intervention group indicated an improvement in vocabulary knowledge when the VSS+, wiki ILH intervention were used. The second research question addressed the students’ attitudes to the VSS+, wiki and ILH as a means of developing their vocabulary. The questionnaire results indicated that the majority of students found the strategy and tools to be a useful method by which to learn new IT content specific vocabulary.

The present study has implications for both teachers and learners. The VSS+/Wiki/ILH framework could be easily adapted by content-specific teachers as a method for
developing their students’ vocabulary knowledge and concepts in their specialized courses. Although this a pilot study, these results indicate that this method of autonomous learning could have long term future implications that could be beneficial to many fields of study, wherein extensive knowledge and retention of vocabulary is required. The potential to share the framework and build it into course templates is also a possibility, as many universities now use learning management systems such as BlackBoard or Moodle that incorporate wikis in their course tools. With the help of free and readily available tools, the framework described in this paper, can help learners work collaboratively and further enhance their vocabulary knowledge.

At this point, it is useful to identify some of the factors that should be taken into consideration for the follow-up study. The development of the corpus in this study, although straightforward with the Sketch Engine application, could be created just as easily with the AntConc application that is freely available for all use and has training videos readily available on YouTube. It would also be advisable to contact textbook publishers, well in advance, for a .txt version of the course textbook being used to create the corpus, as the option of scanning the whole textbook and creating the .txt file for use in the concordance is very time consuming, even when using specialized software. Finally, future studies should also analyze the long-term retention of the vocabulary, as it is subject specific and, unlike general vocabulary, it will not be recycled on a regular basis throughout the other subjects.

REFERENCES