



141
730
THS



This is to certify that the
thesis entitled

Academic Vocabulary at the Word and Formula Level: An
Examination of Test-Taker Discourse

presented by

Aaron Christopher Ohlrogge

has been accepted towards fulfillment
of the requirements for the

M.A.

degree in

Teaching English to Speakers
of Other Languages

Paula Winkle

Major Professor's Signature

May 13, 2009

Date

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE

ACADEMIC VOCABULARY AT THE WORD AND FORMULA LEVEL: AN
EXAMINATION OF TEST-TAKER DISCOURSE

By

Aaron Christopher Ohlrogge

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF ARTS

Teaching English to Speakers of Other Languages

2009

ABSTRACT

ACADEMIC VOCABULARY AT THE WORD AND FORMULA LEVEL: AN EXAMINATION OF TEST-TAKER DISCOURSE

By

Aaron Christopher Ohlrogge

The Academic Word List (AWL) (Coxhead, 2000) is an influential resource for EAP teaching and testing. Recent interest in formulaic language over the past decade has prompted the development a comparable set, the Academic Formula List (AFL) (Simpson-Vlach & Ellis, submitted). The AFL is a list of formulaic expressions occurring frequently in academic discourse, compiled from both spoken and written corpora, and subdivided into three sublists: the AFL Core, AFL Written, and AFL Spoken. The AFL Core list contains formulas regularly occurring in both speech and writing, while the AFL Written and Spoken lists contain formulas primarily occurring in only one modality.

The validity of such lists centers around their corpus-based origins. However, the corpora used to develop these lists consisted largely of native speaker discourse. Little is known about the use of academic vocabulary by nonnative speakers. Still less is known about what role proficiency level plays in the production of academic vocabulary. The present study examines a corpus of compositions written for a test of academic English. Amount of academic vocabulary is compared against proficiency level. Results indicate an increased use of AWL words and AFL Core formulas by higher proficiency students, but no interaction between proficiency level and use of AFL Written or Spoken formulas. Additionally, individual lexical items and formulas exhibiting substantial variation in frequency of use across proficiency level are examined.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Paula Winke, along with the members of my committee, Charlene Polio and Nick Ellis, for their guidance in the completion of this project. My gratitude is extended to all of the faculty and students of the Michigan State University Second Language Studies and MA-TESOL programs, as well as my colleagues at the University of Michigan English Language Institute, for many hours of enriching discussion. In particular I would like to thank Ute Römer and Matthew O'Donnell for their help in assembling and analyzing the corpus of texts used in this study, and Barbara Dobson, Associate Director for Testing at the University of Michigan, for her permission to use confidential testing materials and test-taker responses. Finally I would like to thank my parents, John and Carol Ohlrogge, and my wife, Audrey Johnson, for their continued support, interest, and understanding throughout the duration of this thesis.

TABLE OF CONTENTS

List of Tables	v
List of Figures	vi
Chapter 1	
Literature Review.....	1
Chapter 2	
Methods.....	14
Materials.....	14
Analyses	17
Chapter 3	
Results.....	19
Chapter 4	
Discussion	26
AWL Differences across Proficiency Levels	27
AFL Differences across Proficiency Levels.....	31
AFL Discourse Functions.....	32
Chapter 5	
Conclusions.....	35
Limitations	35
Conclusions	36
Chapter 6	
References.....	37

LIST OF TABLES

Table 1: Composition Length Descriptive Statistics	19
Table 2: Frequency of AWL and AFL Items by Proficiency Level	19
Table 3: ANOVA Results	21
Table 4: Group Differences across Proficiency Levels	22
Table 5: AWL Words Varying in Use by Proficiency Level.....	24
Table 6: AFL Formulas Varying in Use by Proficiency Level.....	24

LIST OF FIGURES

Figure 1: Frequency of AWL and AFL Items by Proficiency Level	20
---	----

Literature Review

Academic language is distinct and unique. In particular, academic discourse is known to contain a number of lexical items specific to academic vocabulary. By academic vocabulary I do not mean technical terms unique to a particular field (e.g. the term *washback* in language testing) but rather lexical items common to academic language across many fields (e.g. *analyze, previous, occur*). While researchers have been attempting to create lists of academic words since the 1970s (see Coxhead, 2000 for a review), it was not until the advent of corpus-based language studies in the late 1990s that a satisfactory and widely-accepted list came into being: the Academic Word List (AWL), developed by Coxhead (2000).

Coxhead's wordlist is based on a written corpus of 3.5 million words consisting of 414 texts written by over 400 authors, primarily composed of journal articles and academic textbooks. These articles and books were taken from four major disciplinary divisions, namely Arts, Commerce, Law and Science. Furthermore, each of the four major divisions was subdivided into seven subdivisions. For example, the subdivision for Arts included Education, History, Linguistics, Philosophy, Politics, Psychology and Sociology. All divisions and subdivisions in the corpus were of approximately equal size. This was done because Coxhead's intention was to identify those lexical items that occurred not just *frequently* in the corpus, but also *broadly*. As a result, the words appearing on the AWL had to occur a certain number of times across a majority of the divisions and subdivisions in the corpus in order to be included on the list.

The AWL contains a total of 3,111 unique lexical items contained in 570 word families. Coxhead (2000) defines word families as a stem able to exist on its own plus all

additional affixes that can be attached to the word. For example, the word family *indicate* includes the lexical items *indicate*, *indicated*, *indicates*, *indicating*, *indication*, *indications*, *indicative*, *indicator*, and *indicators*. In contrast, *special* and *specify* do not belong to the same word family because *spec* is not a stem that can exist in isolation. Also, the AWL excludes all members of the 2,000 most common word families, which are not considered to be academic in any way.

When first developed, the AWL was found to cover about 10% of the lexis in the academic corpus it was drawn from. While this figure may appear small, Coxhead (2000) points out that the AWL combined with the General Service List covers a total of 86% of her academic corpus. In other words, around 75% of the words of most texts consist of the 2,000 most common word families in English. Following its development, the AWL was further validated by investigating its coverage of a second academic corpus containing separate texts from the same disciplinary fields as the first corpus; the AWL was found to cover 8.6% of the second corpus. Furthermore, to verify that the AWL consisted of *academic* words, it was used to analyze a comparable sized corpus of fiction texts. The AWL covered only 1.4% of the lexis of the nonacademic corpus.

However, the pedagogical value of the AWL has recently come under question in the subfield of applied linguistics known as English for Academic Purposes (EAP). Briefly, EAP deals with teaching the features of English to nonnative speakers of English that are needed for success in an academic environment. Hyland and Tse (2007) observe that there is a fair amount of variation in the frequencies, and perhaps more importantly, the meanings, of particular lexical items as they occur across disciplines. Hyland (2008) makes similar arguments about lexical bundles (frequently occurring groups of three or

more words) that occur in academic writing. He points out that many of the most frequent lexical bundles that occur in a particular discipline (e.g. business, law, or natural science) do not occur frequently in other disciplines. Thus, Hyland claims, EAP teachers should focus on discipline-specific vocabulary rather than general academic vocabulary.

However, the intention of the AWL was never to isolate an exclusive set of lexical items worthy of focus in EAP instruction. Rather, they are intended as a common, shared basis or background for academic study. Many EAP instructors do not have the luxury of instructing students from only certain disciplines; quite often a range of disciplines are represented in a single EAP classroom, simply for institutional logistic and/or financial reasons (Gilquin, Granger & Paquot, 2007). Thus, the AWL provides a logical and sensible starting point for developing EAP materials (e.g Schmitt & Schmitt, 2005).

Additionally, it is generally now agreed that native or native-like vocabulary competence does not consist solely of individual lexical items. Both L1 and L2 vocabulary competence also includes knowledge of a large number of formulaic expressions (Ellis, Simpson-Vlach & Maynard, 2008; Pawley & Syder, 1983; Sinclair 1991; Wray 2002). These formulaic expressions are variously known by many names, including *formulaic language*, *formulaic sequences*, *multi-word units*, *lexical bundles*, *fixed expressions*, and, most traditionally, *idioms* (see Wray, 1999 for a review of the terminology).

It is generally agreed that formulaic expressions are highly characteristic both of written texts (Erman & Warren, 2000; Sinclair 1991) and spoken texts (Biber, Johansson, Leech, Conrad, & Finegan, 1999). Not surprisingly, types of formulaic expressions and

their discourse functions appear to vary to some degree across the speech and writing modality (Pickering & Byrd, 2008). Estimates vary, however, as to just how frequent formulaic expressions may be. As detailed by Wray (2002), recent estimates of how much of naturally occurring language is formulaic have ranged from about 5% to as much as 80%, with many estimates falling somewhere in between. Undoubtedly, such variation is largely due to differences in identification and classification of formulaic language across studies.

Formulaic expressions fulfill a variety of discourse functions in child acquisition of both L1 (Peters, 1983) and L2 (e.g. Girard & Sionis, 2003; Myles, Hooper & Mitchell, 1998) as well as adult L2 acquisition and communication (e.g. Simpson-Vlach & Ellis, submitted; Wray & Perkins, 2000). There is also extensive psycholinguistic evidence that formulaic expressions are produced, processed, and stored as whole units rather than word-by-word (Ellis, 1996; Wray 2002, 2008) and as a result are processed more quickly than novel expressions by both L1 and L2 language users (Fei & Ohlrogge, in preparation; Siyanova & Schmitt, 2008; Jiang & Nekrasova, 2007; Conklin & Schmitt, 2007; Ellis, Simpson-Vlach et al., 2008).

Furthermore, there is extensive evidence that knowledge of formulaic language is closely correlated with many standardized and general measures of proficiency. While Pawley and Syder (1983) were among the first to point out the importance of achieving “idiomatic” competence in an L2, theirs was merely a speculative, theoretical study. One early work that did consider the influence of proficiency level on formulaic or idiomatic language use was that of Yorio (1989), who compared groups of compositions written by beginning ESL students in the U.S. and advanced EFL students in Argentina. He noted a

greater tendency for the higher proficiency EFL students to use “idiomatic” or collocational phrasings in compositions. However, his was not an experimental or quantitative study, and included no specific measures of proficiency level, grades assigned to compositions, or number or types of formula use by individual subjects.

Several more recent studies have begun to examine the influence of proficiency level on formulaic knowledge. Bonk (2001) described the development of a language test designed specifically to measure collocational proficiency. The test contained items based on information found in a collocation dictionary designed for ESL students (Benson, Benson & Ilson, 1997). Subjects were required to produce, in writing, one word of a two-word collocation. Bonk’s test included items testing three types of collocations: verb-object, verb-preposition, and figurative verbs. K-R 20 values for the three subtests were 0.69, 0.47, and 0.61, respectively. The test overall had a K-R 20 value of .83. Most items on Bonk’s instrument were able to successfully discriminate between learners of different proficiency levels as determined through classical item analysis, and an IRT analysis indicated that few items on the test were misfitting. In fact, the collocational test produced a better distribution of subjects’ scores than did a subsection of a retired form of the TOEFL, which was used as an independent measure of subjects’ proficiency. Bonk claimed that the better distribution was due to the more advanced subjects being able to “max out” on the TOEFL, while possessing only limited collocational knowledge (p. 125). Additionally, scores on the collocational test were strongly correlated with external measures of proficiency, including TOEFL scores as well as teacher rankings of the participants. In a similar vein, Keshavarz and Salimi (2007) detailed the development of a multiple-choice test of lexical and grammatical

knowledge and two cloze tests, one supply and one multiple choice, with all items taken from a later edition of the same collocation dictionary used by Bonk (Benson et al., 1997). They too reported good reliability of their items, although, unlike Bonk, they did not provide an independent measurement of proficiency to compare their findings against. Nevertheless, these studies do suggest that a command of multiword expressions is a valid and testable construct in second language proficiency.

Comparisons of formulaic knowledge and proficiency are not limited to written collocational tests, however. Van Lancker-Sidtis (1993) presented native, near-native, and advanced ESL students with two recordings of a spoken idiom (e.g. *the coast was clear*). One utterance was intended by the speaker to be the literal meaning of the phrase, while the other was intended to be the idiomatic meaning of the phrase. Native speakers had little difficulty determining which recording was idiomatic and which was literal, while near-native speakers who had lived in the U.S. for many years and used English exclusively in the home and workplace still had great difficulty discriminating between the two. Advanced ESL students performed no better than chance on the task. Van Lancker-Sidtis thus concluded that auditory recognition of idioms is a learnable competency, but only at the very highest levels of L2 proficiency, just as Bonk (2001) did.

It follows then, that if knowledge of formulaic expressions constitutes part of vocabulary competence, and if competence in academic language includes knowledge and use of specifically academic vocabulary, then competence in academic language should also include knowledge and use of academic formulaic expressions. Therefore, the Academic Formula List (AFL) (Simpson-Vlach & Ellis, submitted), was created as a

companion piece to the AWL. Like the AWL, the AFL is drawn from corpora of academic English. Unlike the AWL, though, this includes both written and spoken corpora, namely the Michigan Corpus of Academic Spoken English (MICASE) (Simpson, Briggs, Ovens & Swales, 2002), an academic subsection of the British National Corpus (BNC) which includes both spoken and written texts, and a corpus of research articles (Hyland, 2004).

The creation and validation of the AFL was similar in principle to the AWL's. However, deciding what constitutes an academic formula was naturally more difficult than deciding what constitutes an academic word. Ellis et al. began by extracting n-grams of three, four, and five words in length. An n-gram is a cluster of n words that co-occur together in sequence in a text. They selected n-grams which occurred at least ten times per million words, and were found in a variety of academic fields contained in the academic corpora. Furthermore, n-grams which occurred equally as often in nonacademic corpora, both spoken and written, were excluded, ensuring that the formulas extracted were indeed academic ones. Three separate lists were created from the corpora: a core list containing formulas common to both academic speech and writing; a list containing formulas occurring primarily in spoken text; and a list containing formulas occurring primarily in written text.

While computer programs can easily identify n-grams that occur repeatedly throughout texts, the mere extraction of reoccurring n-grams has not always been a fruitful way of identifying formulaic expressions (Simpson-Vlach & Ellis, submitted; Wray, 2002). The reason is that pure frequency of occurrence leads to the "identification" of many sequences that are composed merely of three extremely common

words (e.g. *yes and um, and for the*) which have no particular discourse function or referential meaning, and are not likely to be stored or processed as intact units.

To combat this, Simpson-Vlach and Ellis (submitted) then applied a specialized statistical measure known as mutual information (Oakes 1998). Put simply, mutual information is a measure of the probability of two or more words co-occurring more often than chance in a text. For example, if one knows that the second and third words of a three word n-gram are ____ *a result*, the probability that the first word is *as* is quite high, much higher than the probability that the first word is *obtain*, another equally grammatical option. A high mutual information score indicates that a pair or sequence of words “coheres” together strongly; in other words, that sequence of words is much more likely to occur together as a group than the individual frequencies of the given words would predict.

However, statistical frequencies and mutual information scores alone were still not enough to create a fully-functional list. Some n-grams that were either extremely frequent or had very high mutual information scores were not deemed to be pedagogically relevant (e.g. *and this is, but it is*). At the same time, some of the least frequent n-grams barely making the ten per million cutoff and some of the phrases with the lowest mutual information score were identified as useful formulaic expressions (e.g. *in the present study, it is obvious that*).

To resolve this, Simpson-Vlach and Ellis (submitted) then selected a subset of the formulas identified and presented them to twenty EAP teachers and language testers. Subjects were asked to rate, on a five point scale, whether they thought the phrase constituted a “chunk,” whether the phrase had a cohesive meaning or function, and

whether they thought it was worth teaching or testing. Results indicated high reliability among the three questions for each phrase, and a multiple-regression analysis indicated that the mutual information of a phrase was nearly twice as important as the frequency at which it occurred in predicting “teaching-value.”

It follows then, that if academic vocabulary at both the word and formula level can be directly identified in a valid way, and if formulaic knowledge is a construct which can be directly tested, then productive tests of academic language, specifically Academic English, should elicit authentic academic vocabulary at both the word and formula level. Since academic discourse incorporates a lexicon that differs in some ways from discourse in general, the use of test tasks that elicit academic language increases the authenticity of tests of Academic English. Maintaining high authenticity in a language test is an excellent way to increase the validity of decisions made based on the test, because it makes it easier to generalize from the test discourse domain to the target language use domain. Furthermore, as an indication of the construct validity of a test, differences in performance of construct-specific features (e.g. academic vocabulary) across different proficiency levels should be observed (Chapelle, 1999). Specifically, there should be a direct and positive association between productive use of academic vocabulary and scores on tests of Academic English.

Surprisingly, though, few investigations into test-taker produced discourse have considered whether assessed proficiency level plays any role in the production of academic language or formulas of any type. The few works concerning formulas all conclude that greater use of idioms and collocations is indeed associated with higher levels of proficiency. Due to lack of explicit definition of the constructs of idioms and

collocations, however, their conclusions come across as tentative at best. Four such studies will be discussed below.

First, Hawkey and Barker (2004), reporting on a project to develop a common writing scale across multiple University of Cambridge ESOL (English for Speakers of Other Languages) international certificate exams, analyzed a set of compositions written by candidates for several different exams, spanning a wide range of proficiency levels. They noted, among many other linguistic features, a much higher frequency of collocations and idioms in highly rated compositions as compared to lower rated ones. Additionally, Kennedy and Thorpe (2007) examined a small corpus of compositions written for the International English Language Testing Service (IELTS) exam in order to identify specific linguistic features that characterize compositions rated at particular bands. The IELTS is an international proficiency exam jointly administered by Cambridge ESOL examinations, Australia IDP, and the British Council. Candidate scores are reported in bands ranging from 1 (lowest) to 9 (highest). In a comparison of compositions receiving scores of 8, 6, and 4, the authors subjectively observed that Band 8 compositions contained a great deal of collocational and idiomatic language, whereas compositions rated as a 6 or 4 had far less use of such language. In a parallel investigation, Read and Nation (2006) analyzed transcripts of IELTS oral examinations of candidates also rated at Bands 8, 6, and 4. They observed extensive use of idioms and collocations among candidates achieving Band 8 on the oral portion of the test, while candidates achieving bands 6 and 4 used considerably fewer idioms and collocations. Neither Kennedy and Thorpe (2007) nor Read and Nation (2006) provided any quantitative data to support their observations, however. Ohlrogge (2008) investigated a

small corpus of compositions written for the Examination for the Certificate of Proficiency in English (ECCE), a high-stakes EFL exam produced by the English Language Institute of the University of Michigan. He identified eight distinct types of formulaic expressions that occurred in student exam papers and observed substantial variation in use of particular types of formulas by proficiency level. While some types of formulaic expressions, such as explicit transitional markers (e.g. *on the other hand*) were favored and perhaps even overused by low proficiency students, other types, including collocations and idioms, were favored by high proficiency students.

A common limitation of these four studies, though, is that they have relied exclusively on researcher intuitions about what constitutes a formula, a matter which is open to claims of subjectivity. Hawkey and Barker (2004) and Kennedy and Thorpe (2007) simply mention the presence of idioms and collocations in their data but do not specify the criteria used to identify them. Read and Nation (2006) and Ohlrogge (2009) cite Wray's (2002) working definition of a formulaic sequence:

A sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar (p. 9)

This definition, while intuitively psycholinguistically useful, does not solve the ultimate problem of subjectivity: what “appears to be” prefabricated to one researcher may not appear so to another. The creation of the AFL, as a companion to the AWL, mitigates this drawback to some degree, as it provides a common, concrete reference point grounded in and validated by psycholinguistic and corpus linguistic methodology.

Returning for a moment to the AWL, it appears that there are only two, related studies that consider whether higher scoring candidates on an academic English test actually produce more academic vocabulary than lower scoring ones. Brown, Iwashita and McNamara (2005) looked at production of AWL words across four oral tasks on the TOEFL iBT. They found that while use of AWL vocabulary varied significantly between the four tasks and between forms of the same tasks, there was no significant relationship between test-taker score and the number of AWL words a TOEFL test taker produced. Furthermore, they observed extremely large variances in the totals of AWL words used by different candidates; that is, while some learners used many AWL words, others used quite few. One significant limitation of this analysis is that the AWL, as mentioned before, was compiled solely from written texts, which differ substantially from academic spoken texts in significant ways, including lexically. As a result, a further analysis was done to compare the speech samples collected by Brown et al. (2005) to lexis found in the MICASE corpus (Iwashita, 2005). Iwashita isolated vocabulary that occurs more frequently in the MICASE corpus than standard word frequency lists would suggest and compared this set of spoken academic vocabulary to the TOEFL iBT speech samples. This comparison also revealed no specific relationship between use of academic vocabulary and test-taker score. Iwashita noted, however, that since multiple, brief speaking tasks were analyzed (most lasting 1-2 minutes), these samples may have been too short to adequately sample the construct of academic vocabulary.

In order to fill in the gaps mentioned in the previous literature, the following research questions are posed:

- RQ1: Do more proficient academic writers produce more AWL items than less proficient academic writers?
- RQ2: Do more proficient academic writers use more AFL formulas than less proficient academic writers?
- RQ3: In what ways do higher and lower proficiency writers differ in their use of academic vocabulary?

Based on the literature described above, the following two hypotheses are posited:

- RQ1: Production of AWL words will vary by proficiency level in an academic writing test. Most studies of diversity in lexical output (e.g. Laufer & Nation 1995) have indicated that higher-proficiency students produce a greater range of vocabulary. I expect this to be the case in my study as well.
- RQ2: Production of AFL formulas will vary by proficiency level in an academic writing test. Following Kennedy and Thorpe's (2007) analysis of IELTS written papers, I expect a greater amount of formula use at higher levels of proficiency.

The third research question is exploratory in nature, and therefore no prediction is made in terms of how higher and lower proficiency writers will differ in their use of academic vocabulary.

Methods

Materials

The Academic English Evaluation: The Academic English Evaluation (AEE) is an in-house ESL placement test used by the English Language Institute at the University of Michigan (UM). Drawing on principles of academic English, such as those outlined by Swales and Feak (2000, 2004), the AEE assesses academic language as used in an academic context. The AEE is administered to incoming undergraduate and graduate students to the University of Michigan whose submitted TOEFL, IELTS, or MELAB scores fall below a particular cut-off. The AEE is administered before the beginning of the student's first semester at UM. During an administration of the AEE, a single prompt writing test is administered first, followed by a multiple choice video listening test, a multiple choice grammar, vocabulary and reading test, and finally a speaking test consisting of several tasks conducted individually with a single examiner. Following the test, students meet individually with a counselor, who discusses the results of the AEE with the student and details any ESL course requirements that may result from performance on the test.

The AEE Writing Test: Two different tasks exist for the AEE writing test, one for undergraduate students and one for graduate students. Visiting scholars are also given the graduate writing task. The undergraduate writing task is not part of the present study and will not be discussed further. In the graduate task, students are presented with a short textual prompt which is followed by a brief chart and table. In the writing prompt used in this study, the data presented in the chart and graph depict a public health problem in the United States, namely, an increase in the rate of childhood obesity, and a variety of factors that could be contributing to the problem, (e.g. increased consumption of fast

food, decreased hours in PE classes)¹. Students are asked to write a short report on the data provided, explaining the trend and its cause(s), and offering brief recommendations for solving the problem. Students are informed that their target audience is a professor from the student's own department other than the student's own advisor, and that they will be assessed on their ability to present the data in an academically appropriate way.

The AEE writing test is graded holistically on an 8-point scale by at least two trained raters. Raters are employees of the Testing Division of the English Language Institute who have all had previous experience scoring ESL writing for other high stakes tests. Exact agreement between raters results in a final score as given by the two raters. Adjacent scores result in a final score in between the two given scores. Nonadjacent scores are resolved by a third rater. If the third rating is adjacent to both of the first two ratings (i.e. the discrepant ratings are two points apart and the third rating falls in between the two), the third rating is taken as the final score. If the third rating is identical to either of the first two ratings, then the rating which has been assigned twice is taken as the final score.

The Corpus: A total of 310 compositions written for the AEE during Fall 2007 were selected for inclusion in the corpus; the total size of the corpus was 103,765 words. All compositions were typed into Microsoft Word and saved as .txt files. The compositions were hand-typed by the author and a research assistant. Because the focus of this study is lexical and because the corpus software program used in this study cannot

¹ In the interest of test security, permission to include the actual writing prompt as an appendix to this thesis was denied. For more information on the writing prompt or on the AEE in general, contact Barbara Dobson, Acting Director of the Testing Division of the English Language Institute and Program Manager of the AEE at <bdobson@umich.edu>.

recognize alternative spellings, all spelling errors were corrected as the electronic versions of the compositions were being created. Errors which represented a nonexistent morphological form in English (e.g. *dramatical*, *overweighted*) were not corrected. No grammatical or vocabulary errors were corrected. This corpus represents the total number of graduate AEEs administered during Fall 2007.

This corpus of 310 compositions was then split into three smaller, subcorpora, divided by proficiency level. Because adjacent scores on an 8-point scale are acceptable and lead to a final score between the two assigned scores, a total of fifteen possible final scores exist. Thus, compositions were sorted into three proficiency levels based on their final scores, with scores of 1-5 labeled as a low proficiency group, scores of 6-10 labeled as an intermediate proficiency group, and scores of 11-15 labeled as a high proficiency group. There were sixty-seven compositions that were placed into the low group, two hundred that were placed into the intermediate group, and forty-three that fell into the high group.

Analyses

Each of the three subcorpora described above varies both in the number of words it contains and the number of texts it contains. Variation in subcorpus size can be controlled either by controlling for total number of words or by total number of texts present. To address research questions one and two, variation in size was controlled by total number of words. Total counts of all AWL words and all AFL formulas were extracted from the corpus via the software program Wordsmith 5 (Scott, 2008). This method allows for the researcher to analyze the overall frequency of all AWL and AFL items by proficiency level, regardless of how many individual writers may have used a given word or formula in their composition. An ANOVA analysis was conducted using

SPSS (version 16.0) to compare the frequency of AWL and AFL items by proficiency level. Tukey's HSD and Tamhane's T2 post hoc tests were conducted to confirm the direction and significance of the results obtained. The independent variable was the proficiency level of the student, and the dependent variables were the frequencies of AWL and AFL items.

To investigate the third research question, variation in subcorpus size was controlled for by number of texts in the subcorpus, rather than total number of words in each subcorpus. The total occurrences of each individual AWL word and AFL formula within each subcorpus were calculated using an unpublished software script.² To control for variation in the number of texts in each subcorpus, the total number of occurrences of each individual word was divided by the number of texts in each subcorpus. This was done because it would be difficult to analyze what percentage of each subcorpus might be comprised of any individual word or formula, because the proportion would be exceedingly low. For example, the most common word in the English language, *the*, comprises only a tiny proportion of the total words in any corpus of naturally occurring text. A significantly less frequent word, such as a word from the AWL (or a formula from the AFL) would comprise an even smaller proportion of the total words in a corpus of naturally occurring text. However, the proportion of individual writers using a particular word or formula in a highly predictable context (e.g. responses to a single writing prompt) might reasonably be expected to be a relatively large and analyzable number. In other words, substantial and meaningful differences might be observed in the

² The software script was written by Matthew O'Donnell. It extracted the total number of occurrences of each AWL word and AFL formula individually across each of the three proficiency levels described above.

varying proportion of low, intermediate, and high proficiency writers using a particular word or formula. An arbitrary difference of 10% was selected as a cut-off criterion. That is, only those words and formulas that showed differences in use of 10% or more were selected for further analysis.

Results

Descriptive statistics for each of the three proficiency levels are shown in Table 1.

Table 1: Composition Length Descriptive Statistics

Group	Number of texts	Total # of Words	Mean Length	SD	Min	Max
Low	67	18338	273.70	79.70	105	453
Intermediate	200	68962	344.81	82.09	134	572
High	43	16465	382.91	86.48	191	561

As in many studies of second language writing, higher-rated writing samples were significantly longer than lower-rated samples. A one-way ANOVA analysis revealed statistically significant differences in length across all three levels, $F(2,307) = 27.35$, $p < .001$.

Total counts of items from the AWL and the three AFL sublists— Core, Written, and Spoken were obtained for each of the three subcorpora, as shown in Table 2 and Figure 1. Since the size of each proficiency subgroup differed, raw counts of each type of vocabulary item were adjusted by averaging total occurrences against the total number of words in each subcorpus. These ratios were then multiplied by 1,000 in order to avoid dealing with extremely small figures.

Table 2: Frequency of AWL and AFL Items by Proficiency Level

Group	Total AWL Words	AWL Frequency per 1,000 words	Total AFL Core Formulas	AFL Core Frequency per 1,000 words	Total AFL Written Formulas	AFL Written Frequency per 1,000 words	Total AFL Spoken Formulas	AFL Spoken Frequency per 1,000 words
Low	860	46.9	163	8.9	117	6.38	35	1.91
Intermediate	3541	51.3	828	12	477	6.92	160	2.32
High	1016	61.7	204	12.4	113	6.86	26	1.58

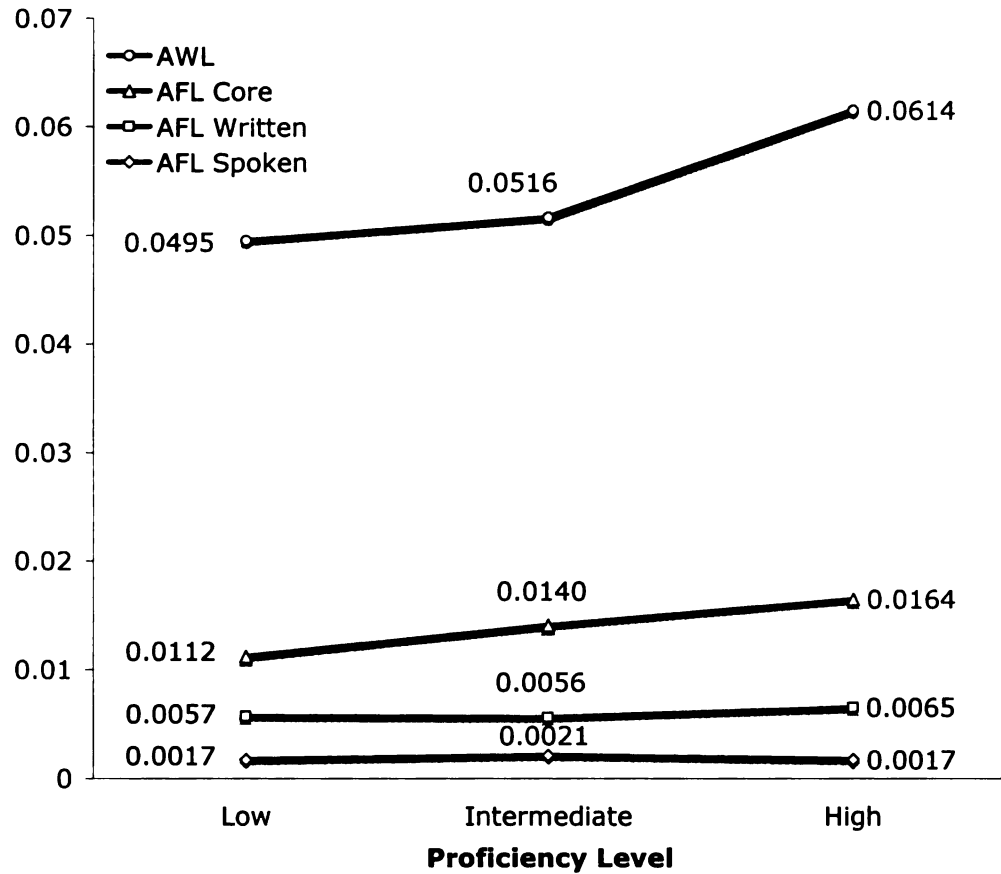


Figure 1: Frequency of AWL and AFL Items by Proficiency Level

A one-way ANOVA analysis and two post hoc analyses were conducted using SPSS Version 16.0 in order to determine whether the ratio of academic vocabulary to total word count differed significantly by proficiency level for each of the four types of academic vocabulary described in this study. Results of this one-way ANOVA test are shown in Table 3.

Table 3: ANOVA Results

Ratio	Comparison	df	F	Sig.
AWL Words To Total Word Count	Between Groups	2	6.249	0.002
	Within Groups	307		
AFL Core To Total Word Count	Between Groups	2	5.762	0.003
	Within Groups	307		
AFL Written To Total Word Count	Between Groups	2	0.573	0.564
	Within Groups	307		
AFL Spoken To Total Word Count	Between Groups	2	1.024	0.36
	Within Groups	307		

Results indicate that there was a significant interaction between proficiency level for AWL words, $F(2,307) = 6.25, p < .01$, and for AFL Core formulas $F(2,307) = 5.76, p < .01$. There was no significant interaction between proficiency level and AFL Written formulas, $F(2,307) = .573, p > .05$, or between proficiency level and AFL Spoken formulas, $F(2,307) = 1.02, p > .05$. A Levene's homogeneity of variances test was conducted in order to determine the appropriate post hoc tests to use to distinguish which proficiency level(s) differed from which other level(s). Results indicated that the assumption of equal variances was not violated for counts of AWL, AFL Spoken, or AFL Written items. However, the assumption of equal variances was violated for the AFL

Core list. As a result, Tamhane's T2 was used for the AFL Core list, while Tukey's HSD was used for the remaining three lists.

Both the AWL and AFL Core lists exhibited some variation by proficiency level. The low and high groups differed significantly in use of AWL words, and the low group differed significantly from both the intermediate and high groups in AFL Core use. The intermediate and high groups did not differ significantly from one another in AFL Core use. No significant differences were observed for AFL Written or Spoken use across the three proficiency levels. For AFL Written formulas, there is a modest but nonsignificant increase in use between low and intermediate compositions, but almost no increase between intermediate and high. For AFL spoken formulas, there is again a modest but nonsignificant increase between low and intermediate compositions, followed by a slight drop between intermediate and high compositions. A summary of the differences between groups is shown in Table 4.

Table 4: Group Differences across Proficiency Levels

Dependent Variable	(I) Group	(J) Group	Mean Difference (I-J)
AWL	Low	Intermediate	-0.0022
		High	-0.0119*
	Intermediate	Low	0.0022
		High	-0.0098*
	High	Low	.0119*
		Intermediate	.0098*
Core	Low	Intermediate	-0.0029*
		High	-0.0053*
	Intermediate	Low	.0029*
		High	0.0024
	High	Low	.0053*
		Intermediate	0.0024

Table 4: Continued

Written	Low	Intermediate	0.0001
		High	-0.0008
	Intermediate	Low	-0.0001
		High	-0.0009
	High	Low	0.0008
		Intermediate	0.0009
Spoken	Low	Intermediate	-0.0005
		High	0.0000
	Intermediate	Low	0.0005
		High	0.0004
	High	Low	0.0000
		Intermediate	-0.0004

* = difference is significant at $p < .05$

Additionally, which individual academic words and phrases might vary by proficiency level was investigated. An arbitrary cut-off point of 10% was selected as a starting point for investigation. Results of this analysis are shown in Tables 5 and 6. A total of seventeen words from the AWL, six formulas from the AFL Core, one formula from the AFL Written, and one formula from the AFL Spoken differed in frequency by ten percent or more. Words and formulas in both tables are arranged by increasing differences between low and high proficiency group use.

Table 5: AWL Words Varying in Use by Proficiency Level

	Low	Intermediate	High	Low-Intermediate Difference	Intermediate-High Difference	Low-High Difference
conclusion	7.5	14.5	2.3	7.0	12.2*	5.1
exposure	20.9	38.5	30.2	17.6*	8.3	9.3
consumption	3.0	5.5	14.0	2.5	8.5	11.0*
issue	4.5	13.0	16.3	8.5	3.3	11.8*
impact	1.5	3.5	14.0	2.0	10.5*	12.5*
role	6.0	8.0	18.6	2.0	10.6*	12.6*
major	7.5	11.5	20.9	4.0	9.4	13.5*
media	0.0	7.0	14.0	7.0	7.0	14.0*
data	40.3	35.5	25.6	4.8	9.9	14.7*
period	6.0	13.5	20.9	7.5	7.4	15.0*
trend	52.2	65.5	67.4	13.3*	1.9	15.2*
decade	3.0	9.5	20.9	6.5	11.4*	17.9*
projected	4.5	9.0	23.3	4.5	14.3*	18.8*
computer	40.3	51.5	65.1	11.2*	13.6*	24.8*
computers	3.0	13.0	27.9	10.0*	14.9*	24.9*
involved	31.3	52.0	69.8	20.7*	17.8*	38.4*
percentage	20.9	31.0	60.5	10.1*	29.5*	39.6*

* = difference of 10% or more

Table 6: AFL Formulas Varying in Use by Proficiency Level

	Low	Intermediate	High	Low-Intermediate Difference	Intermediate-High Difference	Low-High Difference	AFL Source
in order to	6.0	17.0	9.3	11.0*	7.7	3.3	Core
the rate of	47.8	50.5	39.5	2.7	11.0*	8.2	Core
the amount of	3.0	10.5	14.0	7.5	3.5	11.0*	Core
at the same time	3.0	11.5	14.0	8.5	2.5	11.0*	Core
due to the	0.0	6.5	11.6	6.5	5.1	11.6*	Core
the number of	32.8	51.0	60.5	18.2*	9.5	27.6*	Core
increase in the	1.5	11.0	30.2	9.5	19.2	28.7*	Written
first of all	4.5	19.0	7.0	14.5*	12.0*	2.5	Spoken

* = difference of 10% or more

As expected, the frequency of use of most academic vocabulary items increased with proficiency level. Fourteen of the seventeen words that exhibited substantial

variation were used most frequently by high proficiency writers. However, two words, *conclusion* and *exposure*, were used most frequently by intermediate learners, and a single word, *data*, was used most often by low proficiency learners. Likewise, six of the nine selected formulas were used most often by high proficiency learners, while three, *in order to* (AFL Core), *the rate of* (AFL Core), and *first of all* (AFL Spoken), were used most often by intermediate learners. No AFL formulas were used substantially more often by low proficiency learners.

Discussion

The present study investigated the use of academic vocabulary by nonnative speakers of English in a test of written Academic English proficiency. Results indicated that higher proficiency writers used more words from the Academic Word List and Core Formulas from the Academic Formula List than did lower proficiency writers. The use of AFL Written and Spoken formulas did not vary by proficiency level. Results also indicated that of the individual words and phrases that exhibited variation by proficiency level, most but not all were used more often by high proficiency than intermediate or low proficiency students.

The fact that higher proficiency writers used more AWL and AFL Core language provides important cross-validation evidence for both the AWL and AFL as pedagogical resources, as well as construct validity evidence for the AEE. EAP professionals are naturally concerned with selecting the most relevant and beneficial lexical items to focus on in classes and class materials. The traits that higher proficiency writers exhibit, including their patterns of vocabulary use, are presumably traits worthy of focus in the EAP classroom. In other words, the results of this study provide confirmatory evidence that words and phrases from the AWL and AFL are indeed worth teaching. Likewise, knowing that writers who use more academic vocabulary are getting higher scores on the AEE adds to the authenticity of the task, which in turn adds to the strength of content validity claims of the test (Bachman & Palmer, 1996).

The lack of significant variation across proficiency level for the AFL Written and Spoken sublists should not necessarily be seen as an indicator of the invalidity of the sublists. Since the present study dealt exclusively with written output, there is little

reason to expect that many items from the AFL Spoken list would appear in test-taker responses. The writers in this study, matriculating international graduate students to a large, Midwestern university, can reasonably be expected to be proficient enough writers to use relatively few elements of spoken language in their academic writing.

The same cannot be said of the AFL Written list, however. There is reason to expect that a constructed response test of academic writing would elicit features of written academic language, and to expect that higher-scoring responses would contain more such features. By definition, the formulas that comprise the AFL Written sublist frequently occur in academic writing but not in academic speech or nonacademic writing. As such, most language learners will have encountered these expressions primarily in only one modality instead of two (i.e. reading/writing, not speaking/listening), and in only one genre of their L2 (i.e. academic discourse). The reduced exposure that learners will have had to quintessentially written expressions as compared to expressions encountered in both speech and writing may well account for why they were less produced, and perhaps less known, than the AFL Core formulas.

AWL Differences across Proficiency Levels

Of the seventeen words showing substantial variation by proficiency level, thirteen were used most frequently by high proficiency students. As shown in Table 5, higher proficiency students used most words about 10 to 20% more frequently than low proficiency students, and intermediate students used many of these words 10% or more often than low proficiency students as well. In some cases, differences between high and low proficiency students reached nearly 40% (*involved* and *percentage* differed from high to low by 38.4% and 39.6%, respectively).

There are several factors that may account for why some words were produced at substantially different frequencies by proficiency level. In some cases, low frequency words such as *projected* and *consumption* may simply not have been known by the low proficiency group. However, most of the words favored by the high group are relatively high frequency and thus probably known, receptively as well as productively, by most of the low-level learners. These include words such as *issue*, *impact*, *role*, *major*, *media*, *period*, *decade*, and *computers*. With the exception of *media*, all of these words are among the 2,000 most common English words, which cover 76% of written academic texts (Coxhead, 2000). As a result, it is reasonable to expect that incoming graduate students to a large public university in the United States, even those who have scored in the lowest third of AEE writing test, know most, if not all, of these words.

One significant factor in explaining why these words were used much more frequently by higher scoring writers appears to be the ability of successful writers to paraphrase the lexical information present in the writing prompt. Although the words listed above do not appear in the prompt, they are semantically and/or morphologically similar to words that do, such as *decade* and *period*. Higher ability students were able to semantically paraphrase the chronological information in the prompt, as in

“In 1980-1990 there has been a sudden change in this rate, and over this *period* alone there was an increase from 7 to 12 percent. After this *period*, the childhood obesity percent has shown and is projected to increase at 4 percent per *decade*”

(Composition 90_7, high group) (italics added)

whereas lower proficiency students often referred to specific years by name repeatedly, such as in

“According to figure 1, it is obviously that childhood obesity rate in the U.S. from 1960 to 2010 is keep increasing. *In 1960*, there is only 5 percents, while 20 percents *in 2010*. Especially from *1980 to 2010*, the rate of childhood obesity is on a sharp rise” (Composition 70_94, low group) (italics added).

The word *computers* (used by 27.9% of high proficiency students compared to just 3.0% of low proficiency students) provides an example of a morphological link. The word *computers* does not appear in the writing prompt, although *computer* does. Higher proficiency students tended to use *computers* to paraphrase or generalize information from the prompt, as in “Finally, children from 5 to 12 spend more and more time in front of screens like TVs and computers” (Composition 90_6, high group) whereas lower ability students tended to repeat information from the source text verbatim, such as “The children had more hours per day of screen time by using the computer or watching television” (Composition 70_16, low group); the phrases “hours per day of screen time” and “using the computer or watching television” both appear in the writing prompt.

However, three words, *conclusion*, *exposure* and *data*, were used more often by low or intermediate level writers rather than high ones. Most occurrences of *conclusion* in the corpus are found in the phrases *in conclusion* or *as a conclusion*. Presumably these are memorized expressions that learners have explicitly learned at some point in their EFL education. However, although these two expressions are hallmarks of the traditional five-paragraph essay, neither is on the Academic Formula List. The highest proficiency writers in this study seem to have realized that published writers do not use phrases like *in conclusion* or *as a conclusion* in their academic writing; only 2.3% of high proficiency writers used the word *conclusion*, whereas 14.5% of intermediate writers did.

The case of the word *exposure* is somewhat different. This word is not part of any commonly taught written phrases, as *conclusion* is. The intermediate students' preference for this word may be due to the way it appears in the writing prompt. Table 1 of the prompt is divided into three sections: *Exercise*, *Diet*, and *Exposure to Advertising*. Although the AEE writing task does not require or explicitly call for a traditional five-paragraph essay, many students, particularly low and intermediate ones, choose to approach the task this way. As a result, these three sections of Table 1 often form the three body paragraphs of a response written in the format of a five-paragraph essay, with explicit references to each of these points in traditionally formed and placed topic sentences. The fact that low proficiency students used *exposure* less frequently than intermediate students may indicate that many of them did not know the word, or that they were not confident in their ability to use it correctly. Morphological errors resulting in English nonwords (e.g. *exposuring*, *exposured*) in this writing task suggest that some writers were unable to draw a connection between *exposure* and the correct verb form *expose*.

Finally, the word *data* was used most frequently by low proficiency students, and with subsequent decreasing frequency as proficiency level increased. Nearly twice the percentage of low students used the word as high students (40.3% versus 25.6%). Like *exposure*, *data* also appears in the writing prompt. However, rather than occurring in a title heading, as *exposure* is, *data* might be better described as occurring in task instructions, as the writing prompt informs students that "Table 1 presents data that may contribute to an explanation of this trend." Many low proficiency students tended to incorporate this instructional phrasing into their responses either verbatim or only slightly

paraphrased, whereas intermediate and high proficiency students seemed to take it for granted that they were presented with data and that their reader would have access to the data too; therefore fewer explicit references to data were needed. This may account for why usage of *data* peaked with low learners as opposed to intermediate ones.

Additionally, it seems likely the word *data* was better known by low proficiency writers compared to other AWL words, as there were relatively few errors produced involving this word.

AFL Differences across Proficiency Levels

While it is clear that higher proficiency writers used more AWL words and AFL Core formulas overall, and that the majority of individual AWL words showing variation were favored by higher proficiency students, patterns in use of specific AFL items are not as unidirectional. Of the eight formulas whose use changed substantially across proficiency level, five (*the amount of*, *at the same time*, *due to the*, *the number of* and *increase in the*) were favored by high proficiency students, while the other three (*in order to*, *the rate of*, and *first of all*) were favored by intermediate students.

Two of the misfitting formulas, *first of all* and *in order to*, are arguably the most likely of these eight to have been memorized as wholes. This is because these two phrases (of the eight under discussion) have the most distinct discourse function and would be relatively easy to teach in an L2 classroom. The misfitting of *first of all* (used by 4.5%, 19.0% and 7.0% of low, intermediate, and high proficiency writers) is likely due to the fact that more proficient writers are aware, whether implicitly or explicitly, that *first of all* is primary a spoken, not written, formula, and thus not an optimal choice for a written task. The explanation for the misfitting of *in order to* (used by 6.0%, 17.0% and

9.3% of low, intermediate and high proficiency writers) remains less clear, as this formula does not occur in the writing prompt, and is an AFL Core formula (that is, it does appear in both spoken and written academic texts produced by native speakers).

The case of the Core formula *the rate of* may be similar to that of the word *exposure* described above. *The rate of* also appears prominently in the writing prompt as the title of the first figure; as such, it is an easy phrase for lower proficiency learners, who may be less confident using their own words in an academic writing test, to borrow directly. Previous research has shown that lower proficiency writers tend to incorporate more language directly from writing prompts than do higher proficiency writers (Ohlrogge, 2009).

AFL Discourse Functions

The eight expressions that occurred in substantially different frequencies across proficiency levels were also compared to their pedagogical discourse functions as identified by Simpson-Vlach and Ellis (submitted). Following Biber, Conrad and Cortes (2004), in addition to the modality classifications described thus far (i.e. Spoken, Written, and Core) Simpson-Vlach and Ellis also divided the AFL into three main pedagogical discourse groups, each containing several layers of hierarchical subgroups. The three primary pedagogical discourse groupings are referential expressions, which refer to “physical or abstract entities, or to the textual context itself,” stance expressions, which express “attitudes or assessments of certainty,” and discourse organizers, which “reflect relationships between prior and coming discourse” (Biber, et al., 2004: p. 384, cited in Simpson-Vlach and Ellis). It is important to note that these three pedagogical discourse groupings are distributed across the modality classifications. That is, referential

expressions, stance expressions and discourse organizers are all found across the AFL Core, Written and Spoken lists.

In the present study, four of the formulas that displayed substantial variation across proficiency level are classified as referential expressions (*the amount of, increase in the, the rate of, and the number of*), and all four of these fall under the sub-heading of explicit quantitative references. This is not surprising, since the focus of the AEE writing task is a quantitatively-based data commentary task. It also suggests that AEE candidates do indeed differ significantly in their ability to discuss quantitative information in an academically appropriate way. This finding lends support to construct validity arguments for the AEE.

The other four formulas that exhibited substantial variation across proficiency level (*due to the, at the same time, in order to, and first of all*) are classified as discourse organizers. No stance expressions, Simpson-Vlach and Ellis's third major grouping which expresses important functions in academic discourse such as hedging, possibility, and modulated claims, were used more frequently by one proficiency level over another. Simpson-Vlach and Ellis do note that stance expressions tend to be more characteristic of academic speech than of academic writing. Indeed, very few stance expressions were found at any level of proficiency in the present study; *according to the* is the only stance expression used by more than ten percent of writers at any proficiency level, and its frequency does not vary substantially across proficiency level (used by 17.9, 17.0 and 14.0% of low, intermediate, and high learners, respectively.) This suggests that use of stance formulas may not be a salient criterion for discriminating higher ability EAP writers from lower ability ones. This finding is somewhat surprising because the ability

to hedge and modulate claims is featured prominently in the AEE scoring rubric and is often identified as a key skill in academic writing. It may be the case, then, that the most proficient EAP writers are expressing modality in other ways besides using formulaic expressions (e.g. with modal verbs).

As shown in Table 2 and Figure 1, for both the AFL Core and AFL Written lists, a substantial jump in frequency occurs between low and intermediate learners, and only a slight jump is present between intermediate and high. This suggests that low learners have a relatively weak command of academic formulaic expressions, one which develops later as writing proficiency increases. The increase in AFL spoken usage between low and intermediate learners suggests that intermediate learners do have some awareness of academic expressions, but are less sure about which formulas (e.g. *first of all*) are suitable for academic speech as opposed to writing. While even the low proficiency writers likely have some awareness that registers differ across speaking and writing (given that they are incoming graduate students and visiting scholars), they appear to have much less awareness of which specific lexical features are appropriate for writing as compared to speaking.

Conclusions

Limitations

The most significant limitation of the present study is that it is based upon a single administration of a single writing prompt. As discussed throughout this document, idiosyncratic features of the writing prompt used in the present study have surely influenced the results to some degree. In particular, the prompt itself contained several AWL words and several parts or wholes of AFL formulas. Naturally these were among the most commonly occurring AWL and AFL items. In addition, the particular demands of the task, which include quantitative data commentary and elements of cause-and-effect and problem-solution discourse, have likely governed the types of academic lexis typically used to respond to the prompt to some degree. Further research, drawing upon additional writing tasks and prompts, would increase the degree to which results from this type of study could be generalized to other writing contexts. Likewise, research involving data from spoken corpora, ideally also searchable by proficiency level, would be highly beneficial in determining what relationship, if any, might exist between L2 proficiency and use of spoken academic vocabulary at both the word and formula level. Unfortunately, transcription is tedious and expensive, and such corpora are not widely available except as property of large-scale testing organizations.

Another limitation is that the present study has relied upon published word and formula lists established by outside researchers. These lists are open to and have been subjected to methodological criticisms of their own (e.g. Hyland & Tse, 2007). However, in the absence of any modern, competing lists of academic vocabulary in the TESOL and applied linguistics community, the AWL and AFL may reasonably be regarded as appropriate standards of academic lexis for the time being.

Implications

The AWL has been a valuable tool for vocabulary research and EAP instruction over the past decade, and it is expected that the AFL will serve as a similar resource in the future. The results of this study indicate that such vocabulary lists can be an important resource in the field of language testing as well. For example, discrete multiple-choice items testing the form and meaning of individual AFL items are already in development at the University of Michigan English Language Institute. Additionally, the AWL and AFL might be provided to raters of tests of academic English, including the AEE, in order to raise raters' consciousness of the construct of academic language knowledge. Highlighting appropriate (or inappropriate) uses of academic vocabulary in benchmark compositions during rater training may also help to raise raters' awareness. Finally, the AWL and AFL may be useful in the development of new academic writing tasks and prompts, as proportions of academic language produced by learners can be quantified in pilot administrations of new tasks. The value that corpus-derived vocabulary lists provide to language test development, as well as EAP materials development, should not be overlooked, as they provide a comprehensive and methodologically sound basis for an important construct in academic L2 proficiency.

REFERENCES

- Bachman, L. F., & Palmer, A. S. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.
- Benson, M., Benson, E., & Ilson, R. (1997). *The BBI Dictionary of English Word Combinations*. Amsterdam: John Benjamins.
- Biber, D., Conrad, S., & Cortes, V. (2004). *If you look at ...!* Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25, 371-405.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *The Longman Grammar of Spoken and Written English*. London: Longman.
- Bonk, W. (2001). Testing ESL learners' knowledge of collocations. In T. Hudson & J. D. Brown (Eds.), *A focus on language test development: expanding the language proficiency construct across a variety of tests*. Honolulu: University of Hawai'i, Second Language Teaching and Curriculum Center.
- Brown, A., Iwashita, N., & McNamara, T. (2005). *An examination of rater orientations and test taker performance on English for academic purposes speaking tasks*. Princeton, NJ: Educational Testing Service.
- Chapelle, C. (1999). Validity in Language Assessment. *Annual Review of Applied Linguistics*, 19, 254-272.
- Conklin, K., & Schmitt, N. (2007). Formulaic sequences: are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, 29(1), 72-89.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238.
- Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking and points of order. *Studies in Second Language Acquisition*, 18, 91-126.
- Ellis, N. C., & Simpson-Vlach, R. S. (submitted). An Academic Formulas List (AFL).
- Ellis, N. C., Simpson-Vlach, R. S., & Maynard, W. C. (2008). Formulaic language in native and second-language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, 41(3), 375-396.
- Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text*, 20(1), 29-62.
- Fei, F., & Ohlrogge, A. (in preparation). Reexamining measures of formulaic sequence

processing: the interface of implicit and explicit knowledge of formulaic language.

- Gilquin, G., Granger, S., & Paquot, M. (2007). Learner corpora: the missing link in EAP Pedagogy. *Journal of English for Academic Purposes*, 6(4), 319-335.
- Girard, M., & Sionis, C. (2004). The functions of formulaic speech in the L2 class. *Pragmatics*, 14(1), 31-53.
- Hawkey, R., & Barker, F. (2004). Developing a common scale for the assessment of writing. *Assessing Writing*, 9, 122-159.
- Hyland, K. (2004). *Disciplinary Discourses*. Ann Arbor, MI: University of Michigan Press.
- Hyland, K., & Tse, P. (2007). Is There an "Academic Vocabulary"? *TESOL Quarterly*, 41(2), 235-253.
- Iwashita, N. (2005). An investigation of lexical profiles in performance on EAP speaking tasks. *Spaan Working Papers in Second or Foreign Language Assessment*, 3, 101-111.
- Jiang, N., & Nekrasova, T. (2007). The processing of formulaic sequences by second language speakers. *Modern Language Journal*, 91(3), 433-445.
- Kennedy, C., & Thorp, D. (2007). A corpus-based investigation of linguistic responses to an IELTS academic writing task. In L. Taylor & P. Falvey (Eds.), *Studies in Language Testing* (Vol. 19 IELTS Collected Papers, pp. 316-377).
- Keshavarz, M. H., & Salimi, H. (2007). Collocational competence and cloze test performance: a study of Iranian EFL learners. *International Journal of Applied Linguistics*, 17(1), 81-92.
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: lexical richness in L2 written production. *Applied Linguistics*, 16, 307-322.
- Myles, F., Hooper, J., & Mitchell, R. (1998). Rote or rule? Exploring the role of formulaic language in classroom foreign language learning. *Language Learning*, 48(3), 323-363.
- Oakes, M. (1998). *Statistics for Corpus Linguistics*: Edinburgh University Press.
- Ohlrogge, A. (2009). Formulaic expressions in intermediate writing assessment. In R. Corrigan, E. Moravcsik, H. Ouali & K. Wheatley (Eds.), *Formulaic Language. Volume 2: Acquisition, loss, psychological reality, and functional explanations* (Vol. 2, pp. 79-90). Amsterdam: John Benjamins.

- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In J. Richards & R. Schmidt (Eds.), *Language and Communication* (pp. 191-225). London: Longman.
- Peters, A. (1983). *The Units of language acquisition*. Cambridge: Cambridge University Press.
- Pickering, L., & Byrd, P. (2008). Investigating connections between spoken and written academic English: lexical bundles in the AWL and in MICASE. In D. Belcher & A. Hirvela (Eds.), *The Oral/Literate Connection: Perspectives on L2 Speaking, Writing, and Other Media Interactions*. Ann Arbor: University of Michigan Press.
- Read, J., & Nation, P. (2006). *An investigation of the lexical dimension of the IELTS Speaking Test*. Canberra: IELTS Australia.
- Schmitt, D., & Schmitt, N. (2005). *Focus on Vocabulary: Mastering the Academic Word List*. London: Longman.
- Scott, M. (2008). *Wordsmith Tools (Version 5.0)*. Oxford: Oxford University Press.
- Simpson, R., Briggs, S., Ovens, J., & Swales, J. M. (2002). *The Michigan Corpus of Academic Spoken English*. Ann Arbor, MI: The Regents of the University of Michigan.
- Simpson, R., & Mendis, D. (2003). A corpus-based study of idioms in academic speech. *TESOL Quarterly*, 37(3), 419-441.
- Sinclair, J. (1991). *Corpus Concordance Collocation*. Oxford: Oxford University Press.
- Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: a multi-study perspective. *The Canadian Modern Language Review*, 64(3), 429-458.
- Swales, J. M., & Feak, C. (2000). *English in Today's Research World*. Ann Arbor: University of Michigan Press.
- Swales, J. M., & Feak, C. (2004). *Academic Writing for Graduate Students, 2nd Edition*. Ann Arbor: University of Michigan Press.
- Van Lancker-Sidtis, D. (2003). Auditory recognition of idioms by native and nonnative speakers of English: it takes one to know one. *Applied Psycholinguistics*, 24(1), 45-57.
- Wray, A. (1999). Formulaic language in learners and native speakers. *Language Teaching*, 32(4), 213-231.

- Wray, A. (2002). *Formulaic Language and the Lexicon*: Cambridge: Cambridge University Press.
- Wray, A. (2008). *Formulaic Language: Pushing the Boundaries*: Oxford: Oxford University Press.
- Wray, A., & Perkins, M. (2000). The Functions of formulaic language: an integrated model. *Language and Communication*, 20, 1-28.
- Yorio, C. (1989). Idiomaticity as an indicator of second language proficiency. In K. Hyltenstam & L. K. Obler (Eds.), *Bilingualism across the lifespan: aspects of acquisition, maturity and loss*: Cambridge: Cambridge University Press.

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 03062 7800