Lexical diversity and the use of academic and lower frequency words in the academic writing of EFL students

Neda Akbari
University of Canberra

This study focuses on lexical diversity and the use of academic and lower frequency words in essays written by EFL (English as a Foreign Language) students enrolled in Years 1 and 2 at the undergraduate university level. The purpose of this study is to find out the extent to which EFL students become more proficient in their use of academic and lower frequency words and make more diverse choices in their writing after one year of undergraduate university education in English. The study also compares essays written by EFL students and NS (native speaker) students to determine inter-language differences. Essays written by 62 EFL students and 198 NS students at Years 1 and 2 were analyzed for this study. The findings showed no statistically significant difference between the essays written by EFL students in Year 1 and those written in Year 2, either in terms of lexical diversity or in terms of the use of academic and lower frequency words. EFL students in both year levels had a preference for highly frequent words (words in the 1k frequency band). This is in contrast to the NS students, whose use of academic and some lower frequency words improved in Year 2. The findings also showed a statistically significant difference between the essays written by the EFL and the NS students in both year levels. The EFL students made less diverse lexical choices and used fewer words in various frequency bands than the NS students. Findings are discussed and recommendations are offered to EFL students and their educators on how to focus on these aspects of academic writing.

Keywords: Lexical diversity, academic words, lower frequency words, academic writing, EFL students
Introduction

Existing research shows that large numbers of NNSs (non-native speakers) of English struggle with academic writing in English at various educational levels (Cooley & Lewkowicz, 1995; Dong, 1998; Hanauer & Englander, 2011; Yeh, 2010). This study focuses on EFL (English as a foreign language) students undertaking an undergraduate course in English speaking universities in Ukraine, a non-English speaking country, where students have limited exposure to English outside the classroom. Whilst throughout their secondary education many of these students have studied and produced academic writing, such as essays, mostly in their L1 (first language), they are now expected to write essays that meet academic standards as well as discipline and genre specific requirements in English. Although EFL students might be able to translate the basic structure and the social function of various sections of the essay genre from their L1 into English (Lillis & Curry, 2010; Swales, 2004), some important aspects of their writing are language-specific and require knowledge and practice in the foreign language. For instance, to successfully write in the academic register of English, students need to demonstrate, among other things, a sophisticated command of the lexical, syntactic, and discourse aspects of the English language that are specific to the academic register (and their relevant disciplines). This means students must have a large academic and technical vocabulary at their disposal that allows them to communicate their message. In addition, displaying a lexically sophisticated degree of writing presupposes that students have a productive knowledge of words, including knowledge of lexical form, meaning, and use of words (Nation, 2001). Such knowledge is harder to develop than receptive knowledge of words (Schmitt, 2008). The scarcity of experience in writing academic essays in English during secondary education and the difficulty of establishing productive knowledge of language may present challenges to these EFL students during their years at university. One aspect of academic writing that has received considerable attention in the recent literature is vocabulary building, which is the focus of this study.

Research has shown strong links between high quality academic writing, lexical diversity and the use of academic and lower frequency words in samples produced by NSs (native speakers) of English and ESL (English as a second language) students at various educational levels (Coxhead, 2012; Crossley, Weston, Sullivan, & McNamara, 2011; Dong, 1998; McNamara, Crossley, & McCarthy, 2010; McNamara, Crossley, & Roscoe, 2013). Lexical diversity and the use of lower frequency words reflect sophisticated and skillful language production (Uccelli, Dobbs, & Scott, 2013) since they are indicative of the students’ access to a broad range of vocabulary, which is helpful for communicating the intended message. These lexical aspects of academic writing have also been found to affect, among
others, readers’ perception of a writer’s credibility, competence, and communicative effectiveness (Dillard & Pfau, 2002; Ransdell & Wengelin, 2003). Furthermore, use of academic words is very important in academic writing as it indicates membership of the intellectual academic community (Coxhead, 2012; Ivanič, 1998).

Lexical diversity refers to “the proportion of words in a language sample that are not repetitions of words already encountered” (Jarvis, 2013, p. 88). In other words, it is the positive counterpart of redundancy. For example, *we run up and down the slope of that hill every morning before sunrise* (adapted from Jarvis, 2013) is lexically a diverse sentence as it does not display word repetition. Less frequent words are those that are not among the most commonly occurring words in the language, such as *prefer* instead of *like* or *been relocated* instead of *gone*. Words can be divided into high-, mid-, and low-frequency categories. High-frequency words comprise the most frequent 3,000 word families, a small number of word families which occur very frequently in English texts (Schmitt & Schmitt, 2014). High-frequency words make up around 95% of most written texts (Nation, 2006; Nation & Anthony, 2013). Words in the 4,000 to 9,000 frequency bands are in the mid-frequency category (Schmitt & Schmitt, 2014). Knowledge of mid-frequency words is crucial for language learners as knowledge of high- and mid-frequency words together provides around 98% coverage of most written texts (Nation, 2006; Nation & Anthony, 2013). Words in the 9000+ frequency band are in the low-frequency category (which is not the focus of this study). Additionally, academic words are those that are frequently used in the academic context across disciplines and genres, such as in textbooks, lectures, and research papers, for instance, words such as *activate*, *bound*, and *classify* (in the NAWL (New Academic Word List) of Browne et al., 2013).

Despite underscoring the importance of lexical aspects for the quality of writing, previous research has mostly focused on ESL contexts where ESL students study in English in English-speaking countries (e.g., Coxhead, 2012; Crossley et al., 2011; Dong, 1998; McNamara et al., 2010; McNamara et al., 2013; Yu, 2009). Demonstrating a good command of the lexical aspects of writing may be more challenging for EFL students, specifically for those in EFL contexts such as Ukraine where students mostly write in their L1 (Russian and Ukrainian) during secondary education. Also, these students tend not to encounter English in communicative use outside the classroom, which perhaps prevents them from having the opportunity to be exposed to and use some of the words they have encountered in the classroom. In addition, unlike in many ESL contexts where students come from various language backgrounds, in EFL contexts such as the one exemplified in Ukraine, students share the same L1, which means some will use their L1, rather than English, both when communicating with one another outside the classroom (including for the purposes of academic communication).
and in personal communications with their lecturers. It is, therefore, important to investigate the lexical aspects of writing produced by EFL students and find out the extent to which university education in English may assist these students to use greater numbers of academic and lower frequency words and make more diverse lexical choices in their writing.

Lexical aspects of writing in EFL contexts have been investigated in studies of students at the primary (e.g., Espinosa, 2005; Olinghouse & Wilson, 2013) and secondary (e.g., Henriksen & Danelund, 2015; Laufer, 1998; Jarvis, 2002) educational levels, perhaps more so than at the tertiary level (e.g., Hasselgren, 1994; Laufer & Nation, 1995). For the purpose of comparability, only results for EFL students in upper-secondary and in tertiary educational levels are reported here. Studies that examine EFL students in secondary education have reported vocabulary knowledge deficiencies in EFL students’ written production. Henriksen and Danelund (2015) investigate the vocabulary knowledge of secondary school Danish EFL students (Grade 10 and Grade 12) and their written productions. Their research shows low levels of receptive and productive vocabulary knowledge in general for EFL students who have studied English for 7 to 8 years. More than 80% of the Grade 10 EFL students did not receptively master the 2,000 most frequent words in English and more than 88% of them did not productively master them. In addition, the authors find that EFL students, even the high-level ones, do not use the vocabulary they know in their written production. Grade 12 students score higher on both receptive and productive vocabulary tests than Grade 10 students, yet they produce the same proportion of high- and low-frequency words in their essays. Further, the number of low-frequency and academic words employed by Grade 12 students is very low.

Likewise, Laufer (1998) investigated three types of vocabulary knowledge, namely, passive, controlled active, and free active knowledge, among Israeli EFL secondary-school students (Grades 10 and 11, who have studied English in school for 6–7 years). Her study shows that after one year of school instruction, students’ passive and controlled active vocabulary progresses but their free active vocabulary does not progress at all. Despite this partial progress, students still use the same proportion of frequent and non-frequent words in a free writing task. Further, Hasselgren’s (1994) study on Norwegian EFL students (first year university students and upper sixth-formers) shows that these students, even the high-proficiency ones, mostly use familiar L2 lexical items including high-frequency words, words resembling L1 words, and words learned early on in their writing. In another study, Laufer and Nation (1995) found that their least proficient students made use of more words in the 1k frequency band, while the most proficient group made use of less high-frequency words and more academic words.
Research in EFL contexts is mostly carried out on secondary school EFL students who attend classes in English for only a few hours per week. This limited exposure to English has perhaps influenced results to date. The current study investigates the lexical aspects of writing among EFL undergraduate students whose entire university course is mainly in English. Despite the fact that these students do not encounter English in communicative use outside the classroom, the fact that instruction is in the medium of English may improve their chances of learning new words. In addition, previous research investigated essays written on a general topic and under time pressure, factors that are likely to affect their vocabulary profile. In contrast, this study examines essays written by EFL students for assessment in their disciplinary subjects, and under no time pressure. Essays produced in the disciplines and under no time pressure are more likely to have a better vocabulary profile (i.e., include greater numbers of academic and lower frequency words) than general essays (Coxhead, 2011; Yu, 2009).

The present study

The purpose of this study is to find out the extent to which EFL students become more proficient in their use of academic and lower frequency words and make more diverse lexical choices in their writing after one year of undergraduate university education in English. To this end, it investigates the essays written by EFL students in Year 1 and Year 2 of an undergraduate degree in social sciences. The study also compares a number of essays written by EFL students in both year levels to those written by NS students in the same year levels, to determine inter-language (native and non-native speaker) differences. The research questions are:

1. What are the differences between the essays written by Year 1 and Year 2 EFL students in reference to lexical diversity and the use of academic words and words in the 1–9k\(^1\) frequency bands?
2. What are the differences between the essays written by NS and EFL students (in both year levels) in reference to lexical diversity and the use of academic words and words in the 1–9k frequency bands?

If, after one year of studying in English and writing essays in English, EFL students make more diverse lexical choices and use fewer high-frequency words but more academic words and words of various frequency bands compared with EFL students in their first year, this would indicate that EFL students are able to expand their productive vocabulary knowledge in an EFL university context. However, if

\[^{1}\text{k = 1000}\]
there is no difference between essays written by Year 1 and Year 2 EFL students, it would be recommended that EFL students and their educators focus explicitly on the lexical aspects of writing.

Methods

Participants

The EFL participants in this study were 38 Year 1 (17 male and 21 female; age range: 18–31) and 24 Year 2 (11 male and 13 female; age range: 19–32) students, all native speakers of Ukrainian and/or Russian studying for an undergraduate degree in social sciences in an English-speaking university in Ukraine. The disciplines included business, economics, management sciences, marketing, international law, and sociology. In the participants’ university, teaching and assessment are carried out for the most part in English. The students had little experience in writing essays in English during their secondary education and had taken an English composition course at university. Prior to their university education, their primary and secondary education was mainly in Ukrainian and/or Russian, and they had varying levels of English language proficiency. The essays written by these students were produced as part of their midterm or end-of-term assessment and under no time pressure, and were assessed by their disciplinary lecturers following standard assessment criteria. Only essays with scores of 60% or higher were selected for this study. The essays produced in the disciplines are more likely to have a better vocabulary profile (i.e., include more academic and lower frequency words) than essays produced for EFL classes with general topics (Yu, 2009). Similarly, highly scored essays are more likely to display greater numbers of different words (McNamara et al., 2010).

In addition to investigating essays written by these students, this study also examined writing samples of native speakers of English in Year 1 \((N = 98)\) and Year 2 \((N = 100)\) for comparison purposes. These samples (mostly essays but also including problem questions, case studies, and explanations) were selected from the BAWE (British Academic Written English) corpus (2004–2007) based on their discipline (social sciences: business, economics, sociology, law, and politics), year level (Years 1 and 2), score (60% or higher), and L1 of the student (only essays written by those whose L1 was English were selected).\(^2\) The writing produced by both EFL and NS students had different topics. Nevertheless, as they were disciplinary

\(^2\) It should be noted that the BAWE contributors all received high grades for their writing, and for this reason the BAWE samples are not typical NS output.
texts (in contrast to general texts), it is likely that students were familiar with the
disciplinary topics and thus were able to have a better vocabulary profile. Students’
high scores also indicate their familiarity with their topics. Research has shown
that familiarity with the topic is positively correlated with vocabulary profile of
the composition (O’Loughlin, 1995; Yu, 2009). For instance, in Yu’s study, stu-
dents who took the writing test for professional certification in nursing produced
compositions of higher diversity for the plastic surgery topic, due to familiarity
with the topic.

Task, measures, and data analysis

This study reports the results of an investigation of the essays written by EFL
students as part of their mid- and end-of-semester assessment. These essays, all
ranging between 250 and 500 words, were then compared with those written by
NS students in the same disciplines and year levels. McCarthy’s (2005) MTLD
(Measure of Textual Lexical Diversity) was used to quantify lexical diversity. This
was necessary because the essays written by EFL students had different lengths and
were mostly shorter than those written by NSs. MTLD does not vary as a function
of text length, which allows for comparison of texts of different lengths. This is
achieved by holding “TTR [type-token ratio] constant (usually at .72) while calcul-
ating the average number of words in any segment of text that remains above the
TTR cutoff value” (Jarvis, 2013, p. 94). The higher the MTLD value, the more lexi-
cally diverse a text is. Despite MTLD’s limitations (it does not evaluate the given
text as a unified whole; see Jarvis 2013 for details), it was chosen over vocd3 as the
latter works better with texts 150 words or shorter (McCarthy & Jarvis, 2010). For
the lexical diversity analysis, student essays were imported into the TextInspector
program (n.d.) an online tool for analyzing text difficulty, where MTLD was cal-
culated for each essay. The MTLD measures were then compared across languages
and year levels. ANOVA in SPSS was used to find out whether the inter-group
differences were statistically significant.

The use of academic words and words in the 1–9k frequency bands was ex-
amined by means of the LexTutor program (n.d). This program employs various
word lists to create vocabulary profiles of texts. The essays were submitted to this
program for a word frequency analysis using the BNC (British National Corpus)
1–20k frequency bands (2007). Words in the 1k frequency band are the 1000
most frequently used words in English; words in the 2k frequency band are the

3. Vocd also measures the range and variety of vocabulary used in a text. It “outputs an LD [lexi-
cal diversity] index that is calculated through a series of TTR [type-token ratio] samplings and
curve fittings” (McCarthy & Jarvis, 2007, p. 460).
The BNC is a 100 million word text corpus of samples of written (90%) and spoken (10%) modern British English. It includes texts in various subject areas, genres, and registers, thus representing a wide cross-section of English. The written part of the BNC includes, for example, extracts from newspapers, journals for all ages and interests, and academic books. According to Nation (2012), the mainly formal written nature of the BNC affects the occurrence of words in the high- and low-frequency bands. For example, some very familiar words such as *cat*, *hello*, and *sun* occur in the lower frequency bands, and more formal words such as *civil* and *commissioner* occur in the first 1k frequency band. As the current study investigates written academic samples, a formal written register, this limitation is unlikely to affect this study.

The essays were also submitted to this program for an analysis of the use of academic words using the NAWL (New Academic Word List) (Browne et al., 2013). This list consists of 963 lemmas that are widely used in academic texts. It is derived from an academic corpus of about 288 million words comprising text from academic books and journals from the UK and US, mainly from the Cambridge English Corpus (its major component). It thus provides a large coverage of academic words and covers a wide range of disciplines and topics. This is in contrast to Coxhead’s (2000) AWL (Academic Word List) in which two narrow subject areas of Commerce and Law account for 50% of the entire corpus (Hyland & Tse, 2007). The NAWL was also preferred over Gardner and Davies' (2014) Academic Vocabulary List, which is derived solely from an American English corpus, the Corpus of Contemporary American English. It should be noted that around 33% of lemmas in the NAWL are among the most frequent words (1–3k) in the BNC 1–20k lists. This should be taken into consideration for teaching purposes. The LexTutor program calculated the percentage of academic words and words in the 1–9k frequency bands for each essay. The percentages were then compared across languages and year levels. Likewise, ANOVA was performed to find out the statistical significance of the inter-group differences.

**Findings**

As shown in Table 1, the findings of this study with regard to the lexical diversity in the essays produced by EFL students in Years 1 and 2 of their undergraduate degree demonstrate that students in both year levels had a lexical diversity rating between 60 to 65. This is in contrast to an average lexical diversity rating between 84 to 85 for the NS groups. The difference between the EFL and NS groups was statistically significant at both year levels (Year 1: \( t(133) = 11.13, p = 0.00 \); Year 2: \( t(123) = 21.35, p = 0.00 \). There was no statistically significant difference in terms
of lexical diversity between essays written by EFL students in Year 1 and Year 2 ($t(60) = 2.22, p = 0.51$). Nor was there a statistically significant difference between those written by NS students in Year 1 and Year 2 ($t(197) = 1.65, p = 0.20$).

Table 1. Mean and standard deviation for lexical diversity

<table>
<thead>
<tr>
<th>Year level</th>
<th>Mean MTLD (SD)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Year 1</td>
<td>65.22 (20.25)</td>
<td>84.37 (18.03)</td>
</tr>
<tr>
<td>Year 2</td>
<td>60.41 (16.77)</td>
<td>85.42 (19.57)</td>
</tr>
</tbody>
</table>

*Note: MTLD = Measure of textual lexical diversity; SD = standard deviation; EFL = English as a foreign language; NS = native speaker.*

Table 2 shows the findings of this study in reference to the use of academic words and words in the 1–9k frequency bands in essays written by EFL students in Years 1 and 2. Students in both year levels had a preference for highly frequent words. Around 82% of words used in essays written by Year 1 students and 85% of words used in essays written by Year 2 students were among the most frequently used English words, i.e. words in the 1k frequency band. This is in contrast to the 60% of words used in essays written by Year 1 NS students and 57% of words used in essays written by Year 2 NS students. Interestingly, there was no statistically significant difference between the Year 1 and Year 2 EFL students in reference to the use of academic words or words in the 1–9k frequency bands: 1k ($t(60) = 1.63, p = 0.50$); 2k ($t(60) = 1.03, p = 0.10$); 3k ($t(60) = 2.04, p = 0.15$); 4k ($t(60) = 0.70, p = 1.01$); 5k ($t(60) = 1.02, p = 0.86$); 6k ($t(60) = 0.62, p = 0.53$); 7k ($t(60) = 0.50, p = 0.61$); 8k ($t(60) = 1.41, p = 0.15$); 9k ($t(60) = 1.00, p = 0.31$); academic words ($t(60) = 1.65, p = 0.35$).

The findings of this study show that, in contrast to the EFL students, Year 2 NS students did better than Year 1 NS students in the use of some high- and mid-frequency words and also in the use of academic words. There was a statistically significant difference between the two groups in reference to words in the 1k ($t(197) = 2.97, p = 0.003$), 3k ($t(197) = 2.70, p = 0.007$), 4k ($t(197) = 3.59, p = 0.00$), and 5k ($t(197) = 2.86, p = 0.005$) frequency bands and in reference to academic words ($t(197) = 3.35, p = 0.001$). Year 2 students made use of fewer words in the 1k frequency band but more words in the 3–5k frequency bands and also more academic words. No statistically significant difference was observed between the two groups in reference to words in the 2k ($t(197) = 1.63, p = 0.10$), 6k ($t(197) = 0.62, p = 0.53$), 7k ($t(197) = 0.50, p = 0.61$), 8k ($t(197) = 1.41, p = 0.15$), or 9k ($t(197) = 1.00, p = 0.31$) frequency bands. These findings also indicate that simple exposure to words within the university context does not seem to help
EFL students, unlike NS students, use fewer highly frequent words but more mid-frequency as well as academic words in their writing after one academic year.

Table 2. Mean and standard deviation for the use of academic words and words in the 1–9k frequency bands

<table>
<thead>
<tr>
<th>Words</th>
<th>Year 1 EFL (SD)</th>
<th>Year 1 NS (SD)</th>
<th>Year 2 EFL (SD)</th>
<th>Year 2 NS (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1k</td>
<td>82.16 (7.53)</td>
<td>59.78 (6.24)</td>
<td>84.66 (5.10)</td>
<td>57.21 (5.93)</td>
</tr>
<tr>
<td>2k</td>
<td>7.78 (2.38)</td>
<td>17.05 (2.19)</td>
<td>8.66 (2.82)</td>
<td>17.53 (1.98)</td>
</tr>
<tr>
<td>3k</td>
<td>1.64 (1.31)</td>
<td>4.37 (1.33)</td>
<td>1.54 (1.21)</td>
<td>4.88 (1.29)</td>
</tr>
<tr>
<td>4k</td>
<td>1.15 (1.48)</td>
<td>4.74 (1.69)</td>
<td>1.41 (1.74)</td>
<td>5.27 (0.90)</td>
</tr>
<tr>
<td>5k</td>
<td>0.27 (0.98)</td>
<td>2.83 (1.11)</td>
<td>0.29 (0.55)</td>
<td>3.26 (0.99)</td>
</tr>
<tr>
<td>6k</td>
<td>0.77 (0.89)</td>
<td>1.79 (0.87)</td>
<td>0.58 (0.58)</td>
<td>1.87 (0.82)</td>
</tr>
<tr>
<td>7k</td>
<td>0.18 (0.47)</td>
<td>1.14 (0.67)</td>
<td>0.20 (0.50)</td>
<td>1.09 (0.55)</td>
</tr>
<tr>
<td>8k</td>
<td>0.24 (0.43)</td>
<td>0.79 (0.71)</td>
<td>0.25 (0.60)</td>
<td>0.93 (0.62)</td>
</tr>
<tr>
<td>9k</td>
<td>0.17 (0.59)</td>
<td>0.50 (0.54)</td>
<td>0.04 (0.20)</td>
<td>0.57 (0.49)</td>
</tr>
<tr>
<td>Off-list</td>
<td>2.37 (2.46)</td>
<td>4.87 (2.22)</td>
<td>1.16 (1.40)</td>
<td>5.43 (2.70)</td>
</tr>
<tr>
<td>Academic words</td>
<td>1.10 (1.14)</td>
<td>4.63 (1.58)</td>
<td>1.12 (1.19)</td>
<td>5.27 (1.07)</td>
</tr>
</tbody>
</table>

Note: NS = native speaker; EFL = English as a foreign language; k = 1000 words; SD = standard deviation.

In response to Question 2, the findings of this study show that, in both year levels, the NS students did better than the EFL students in terms of the use of academic words and words in the 1-9k frequency bands (see Table 2). There was a statistically significant difference between Year 1 NS and EFL students in the use of words in the 1k ($t(133) = 12.03, p = 0.00$), 2k ($t(133) = 12.15, p = 0.00$), 3k ($t(133) = 2.83, p = 0.005$), 4k ($t(133) = 11.03, p = 0.00$), 5k ($t(133) = 8.90, p = 0.00$), 6k ($t(133) = 4.84, p = 0.00$), 7k ($t(133) = 6.77, p = 0.00$), and 8k ($t(133) = 4.38, p = 0.00$) frequency bands and in the use of academic words ($t(133) = 10.72, p = 0.00$). NS students made use of fewer words in the 1k frequency band but more words in the 2–8k frequency bands as well as more academic words. No statistically significant difference was observed between the two groups for words in the 9k frequency band ($t(133) = 1.13, p = 0.25$). Similarly, a statistically significant difference was observed between Year 2 NS and EFL students in the use of words in the 1k ($t(123) = 20.87, p = 0.00$), 2k ($t(123) = 18.03, p = 0.00$), 3k ($t(123) = 11.51, p = 0.00$), 4k ($t(123) = 15.28, p = 0.00$), 5k ($t(123) = 14.06, p = 0.00$), 6k ($t(123) = 7.25, p = 0.00$), 7k ($t(123) = 7.15, p = 0.00$), 8k ($t(123) = 4.84, p = 0.00$), and 9k ($t(123) = 5.13, p = 0.00$) frequency bands and in the use of academic words ($t(123) = 16.62, p = 0.00$). NS students employed fewer words in the 1k frequency band but more words in the 2–9k frequency bands as well as more academic words.
Discussion

As observed in the findings section, the EFL students had less diverse lexical choices in their essays and made use of fewer academic words and words in lower frequency bands than the NS students; instead, they mostly used highly frequent English words. In addition, no progress was observed in Year 2 EFL students in reference to the lexical diversity or the use of academic and lower frequency words. In contrast, Year 2 NS students used more words in various frequency bands as well as more academic words. These findings indicate that exposure to academic and lower frequency words at university does not have the same effect on NS students and EFL students, insofar as the latter fail to import such words into their productive language use, such as in academic essays. These findings are especially important when considering that lexical issues have also been observed in the writing of EFL students enrolled at higher educational levels such as a master’s degree or a PhD, and that both NNS students and their educators acknowledge the importance of lexical sophistication in academic writing (Cooley & Lewkowicz, 1995; Dong, 1998; Hanauer & Englander, 2011).

The findings of this study, showing no statistically significant difference between the writings of EFL students in Years 1 and 2 but a better vocabulary profile for NS students at the same year levels, are consistent with those reported in the literature for upper-secondary and tertiary EFL students (e.g., Hasselgren, 1994; Henriksen & Danelund, 2015; Laufer, 1998). This is despite the EFL students of this study taking their entire undergraduate university course mainly in English, and despite producing disciplinary essays for assessment under no time pressure. Demonstrating lexical diversity and exploiting academic and lower frequency words in academic writing require students to have productive knowledge of these words. For this, students need to establish a productive link from meaning to the written form, that involves knowledge of form (word parts, spelling), meaning (what word form to use to express the meaning; what other words could be used), and use (grammatical functions, collocations, constraints on use, e.g., register, frequency) of words (Nation, 2001). For instance, in order for an academic word such as ‘simplify’ to be selected for use in an academic essay, in addition to knowing the word parts and whether it expresses the intended meaning, EFL students also need to know, among others, whether the word can be used in the academic context and what words commonly occur with it. This is much more difficult and time-consuming to learn than forming a receptive link from form to meaning (Schmitt, 2008). This is because in order to apply words in their productive use, students need to have precise knowledge of words, whereas for receptive use, knowing a few distinctive features may be sufficient. In addition, language learners tend to have more practice with receptive use of words than the productive one (Nation, 2001). It is, therefore,
possible that, due to the difficulty of establishing productive links from meaning to form, the Year 2 EFL students of this study demonstrate a vocabulary profile that is similar to that of the Year 1 EFL students, thus lagging behind NS students.

It is thus conceivable that EFL students have knowledge of more mid-frequen-
cy and academic words than is demonstrated in their writing (as suggested in Henriksen & Danelund, 2015). As Nation (2001) points out, although some words might be known by language users, they may not be chosen for use, due to many reasons. It is possible that, as EFL students manage to communicate their message and receive high scores using mainly the most frequent words in English, they are content to exploit these words instead of making diverse lexical choices and making use of more mid-frequency and academic words. The EFL participants of this study used mainly high-frequency words in their disciplinary academic writing. For year 2 EFL students, high-frequency words made up 94.86% of their writing, while mid-frequency words accounted for only 2.77% and academic words for only 1.12%. Similarly, for Year 1 students, high-frequency words made up 91.58% of their writing, while mid-frequency words accounted for only 2.78% and academic words for only 1.10%. Yet students in both year levels received high scores from the disciplinary lecturers. Similar to the secondary education systems in Denmark and Israel (Henriksen & Danelund, 2015; Laufer, 1998), there is no specific focus on the lexical aspects of language use in English-speaking universities in Ukraine. Despite being offered courses that assist them with English language proficiency and English composition, EFL students do not often receive feedback on the lexical aspect of their writing in these courses or from their disciplinary lecturers, nor are they encouraged to use more academic and lower frequency vocabulary in their writing. As Laufer (1998) notes, in communicative-oriented EFL contexts, like in English-speaking universities in Ukraine, there is perhaps no incentive for EFL students to take the risk of using more difficult words. They therefore use the safe-playing strategy (Henriksen & Danelund, 2015) and exploit only the words they are highly familiar with.

Since the EFL students in this study lagged behind NS students in reference to the lexical diversity and the use of academic and lower frequency words at both year levels, since no progress was observed in the writing of Year 2 EFL students, and since the lexical aspects of writing have been found to be predictors of writing quality (e.g., McNamara et al., 2010), it is important that these EFL students (and those in similar contexts) and their educators pay specific attention to them. There are lists available that contain words that are commonly used in academic texts, such as Coxhead’s (2000) AWL, Browne et al.’s (2013) NAWL, or Paquot’s (2010) Academic Keyword List. Educators may need to encourage EFL students to develop productive knowledge of these words and apply them in their academic writing. According to Paquot (2010), students need to learn academic words with
attention to their semantic properties, syntactic positioning, collocations, and register differences, among other things. Teaching and learning words in the 4–9k frequency bands is perhaps more challenging due to the large volume of words involved. One way of focusing on these words is through encouraging students to extensively read graded readers, which provide repetition and varied encounters with the same words (Nation, 2015). Extensive reading can help students with both the learning and use of lower frequency words (e.g., Horst, 2005; Pigada & Schmitt, 2006). Extensive reading can also be accompanied by productive word-focused tasks for further gains (see e.g., Laufer, 2003). There are various techniques for developing productive knowledge of these words that can be used by both EFL students and their educators (see, for instance, Coxhead, 2012; Hirsh & Coxhead, 2009). It is also important to consider having a lexical assessment of student writing by disciplinary lecturers throughout the undergraduate university cycle, so that students will be pushed to use academic and lower frequency words in their writing. As Swain (1995) claims, language learners need to be pushed to use all their resources so that they progress beyond a certain level of competence. Pushed production has benefits for students as it encourages them to, for instance, think carefully about word usage and exploit more specific words (Coxhead, 2011).

Despite its contribution to EFL lexical research, this study has some limitations. The participants were Year 1 and Year 2 EFL students. Unlike for NSs of English, who have shown progress in their vocabulary profile after one academic year, for EFL students, the one year gap may be too short a gap, given that there is no explicit focus on the lexical aspects of writing in their university education. It would, therefore, be beneficial to examine the vocabulary profile of EFL students in their final year at university to find out whether, and to what extent, students may make progress over a longer period of time. In addition, the NS essays were written by native speakers of English who also received high scores for their essays. This may well contribute to the significant difference found in vocabulary profiles of the EFL and NS groups. It would be worthwhile to compare the essays written by EFL students in general to those written by highly proficient ones, to find out whether and to what extent their vocabulary profile may be related to their proficiency levels. This could not be done in this study due to the small number of EFL participants.

Conclusions

This study has raised the possibility that EFL university students do not show any progress with respect to the vocabulary profile of their writing during their undergraduate university years. Enrolled at an English-speaking university in Ukraine,
the EFL students surveyed in this study made less diverse lexical choices and used fewer academic and lower frequency words in their essays compared to NS students. Instead they relied mostly on highly frequent English words. In addition, no difference was observed between the essays written by EFL students in Year 1 and Year 2 in this regard. The findings are perhaps due to the difficulty of establishing productive links from meaning to written forms, EFL students' ability to communicate their message and receive high scores by using only high-frequency words, and the lack of focus on the lexical aspects of language use in the English-speaking universities in Ukraine. Based on these findings, it is perhaps of value for EFL students and their educators to pay specific attention to the lexical aspects of language/writing during university education. Disciplinary lecturers may also need to undertake a lexical assessment of student writing throughout the entire university cycle. Finally, recommendations for the use of various resources for teaching and learning of academic and lower frequency words were provided.

References


Author’s address

Neda Akbari
University of Canberra
University Drive, Bruce
ACT 2617
Australia

Ned.akba@gmail.com