

## **THE USE OF ACADEMIC WORDS AND FORMULAE IN L1 AND L2 SECONDARY SCHOOL WRITING**

**Taha Omidian**

*Independent researcher*

**Natalia Beliaeva**

*Victoria University of Wellington*

**Linda Todd**

*Wellington East Girls' College*

**Anna Siyanova-Chanturia**

*Victoria University of Wellington*

### **Abstract**

*It is widely acknowledged that learning English in the context of academic studies implies mastering academic vocabulary. Such vocabulary is not merely made up of single words, but also consists of multi-word items. Despite the fact that linguists have long called for the inclusion of multi-word items in vocabulary studies and have also pointed out the importance of these units for language learning, few studies in this area have attempted to investigate the use of academic vocabulary at both word and phrase levels. To bridge this gap, the present study examined the use of academic words from the Academic Word List (Coxhead, 2000) and academic formulae from the Academic Formulas List (Simpson-Vlach & Ellis, 2010) in essays produced by first language (L1) and second language (L2) school-age students. To this end, a corpus of L2 school essays was compiled and compared with a matched L1 corpus. Substantial differences were observed between the two corpora in terms of the use of academic words. However, both L1 and L2 writers were found to be comparable in their use of academic formulae. These findings and their implications are discussed in the context of teaching and learning academic vocabulary in K-12 settings (primary and secondary education).*

*Keywords:* native speakers, L2 learners, corpus, academic words, academic formulae

### **Introduction**

Achieving proficiency in academic language can be very demanding for students, especially for those whose academic success is contingent, to a large extent, on their

ability to communicate effectively in their field of study. Research in this area has primarily focused on pathways through which students can achieve such proficiency. One promising avenue that has long been a topic of interest in studies and educational circles in this area is academic vocabulary knowledge. Knowledge of such vocabulary can be conducive to gaining academic literacy in that it provides students with varied lexical choices that are particularly useful for communication in academic discourse (Corson, 1997; Coxhead & Nation, 2001; Li & Schmitt, 2009; Snow & Kim, 2007). Over the past four decades, several academic vocabulary lists have been developed to meet the specific vocabulary needs of students and help them build a varied repertoire of the vocabulary required for undertaking their school or university studies (Campion & Elley, 1971; Coxhead, 2000; Gardner & Davies, 2013; Martinez & Schmitt, 2012; Praninskas, 1972; Simpson-Vlach & Ellis, 2010; Xue & Nation, 1984). These lists are developed on the basis that, in order for students to be able to successfully participate in academic conversations, they need to be familiar with words and sequences commonly used in academic texts (Blachowicz & Fisher, 2000; Coxhead, 2000; Durrant, 2016; Gardner & Davies, 2013; McKeown, Beck, Omanson & Pople, 1985; Simpson-Vlach & Ellis, 2010; Stahl & Fairbanks, 1986). However, expanding vocabulary and developing the ability to use this knowledge for academic purposes can be a relatively slow and difficult process (Cohen, Glasman, Rosenbaum-Cohen, Ferrara & Fine, 1998; Nation, 2000; Schmitt, 2000).

Based on various conceptualizations of what constitutes a vocabulary, two different strands of research have investigated the learning of academic vocabulary. On the one hand, studies have considered single-word lexical items as the unit of analysis for assessing learners' academic vocabulary knowledge (Bailey, 2007; Brock, Lapp, Salas, Townsend, 2009; Corson, 1997; Lee, 2003; Staehr, 2008), estimating the vocabulary size needed for undertaking academic tasks (Baumann & Graves, 2010; Ishii & Schmitt, 2009; Laufer & Nation, 1995; Webb & Nation, 2008), and exploring the lexical distributions and densities of written texts produced by learners (Laufer, 1994; Laufer & Nation, 1995; Lee, 2003; Muncie, 2002). On the other hand, studies have based their analyses on multi-word items, such as formulaic sequences, lexical bundles, and collocations (Siyanova & Schmitt, 2008; Biber, Conrad, & Cortes, 2004; Chen & Baker, 2010; Schmitt, 2004; Staples, Egbert, Biber, & MacClair, 2013). Although the findings from these studies have greatly informed the theoretical and applied research into academic vocabulary knowledge, and English language education in general, the need for a study that considers both aspects of the mental lexicon – single- and multi-word items – exists. The goal of the present investigation was to explore the use of academic words from the Academic Word List (Coxhead, 2000) and academic formulae from the Academic Formulas List (Simpson-Vlach & Ellis, 2010) in essays produced by first language (L1) and second language (L2) adolescent writers.

## **Academic vocabulary**

Academic vocabulary is typically classified into two sub-categories: technical and sub-technical vocabulary (Cowan, 1974; Hiebert & Lubliner, 2008; Nation, 2001). Technical (or discipline-specific) vocabulary is a set of words or sequences which are exclusive to individual academic disciplines and differ by subject area. Words and sequences such as *antiproton*, *estrogen*, *anomie*, *ultra high-energy cosmic rays*, *Navier-Stokes equations*, *flow and heat transfer*, and *English for Academic Purposes* can be found among the terminologies specifically used in disciplines such as physics, biology, sociology, and applied linguistics. Although this vocabulary can be of particular importance for students to gain acquaintance with technical terminologies unique to a certain content area, the acquisition of such knowledge is likely to occur in parallel with the development of students' overall understanding of their field of study (Cowan, 1974; Farrell, 1990; Nagy & Townsend, 2012). The second category of academic vocabulary, however, appears to be more challenging for students to acquire. Sub-technical or general academic vocabulary, which is the focus of this study, is composed of lexical items (i.e., words or sequences) frequently used in a wide range of academic genres, yet relatively uncommon in non-academic registers (Corson, 1997; Coxhead, 2000; Coxhead & Nation, 2001). These lexical items differ from those of technical vocabulary in that they are not tied to any particular subject area, though they may exhibit some degree of semantic disparity across disciplines (Durrant, 2014; Hyland & Tse, 2007).

As was mentioned above, the use of general academic vocabulary is relatively infrequent in colloquial language. This makes it virtually impossible for students to acquire ample knowledge of such vocabulary through incidental learning and exposure to non-academic contexts (Hyland & Tse, 2007). This complication becomes even more serious in primary and secondary education (henceforth K-12), where students have limited exposure to academic language, compared to more academically competitive environments, such as university and college (Nagy & Townsend, 2012). Research has shown that, for students to be able to fully understand sub-technical words and phrases, having multiple exposures to academic texts is crucial (Blachowicz & Fisher, 2000; Stahl & Fairbanks, 1986). Moreover, it has also been argued that the students who are even at a greater disadvantage in the acquisition of such vocabulary are EFL/ESL students. According to Townsend and Collins (2008), these students may experience difficulties coping with the intricacies of academic language due to the daunting task of developing language proficiency while undertaking school tasks. In testing an academic vocabulary intervention with linguistically diverse adolescent students, Lesaux, Kieffer, Kelley, and Harris (2014) found that the effects of the intervention were larger for students whose primary home language was not English. Furthermore, Hakuta, Butler, and Witt (2000) and Short and Fitzsimmons (2007) also demonstrated that, in order to be able to do well in academic content areas, most adolescent English language learners need four to seven years of instruction before they can develop academic literacy. These findings

suggest that students studying in K-12 settings have special vocabulary needs. More research could contribute to understanding the language needs of this demographic.

## **Academic vocabulary knowledge and development**

Understanding academic vocabulary knowledge and how it develops requires attention to a number of key factors involved in the acquisition of a vocabulary. First, vocabulary acquisition is a continual process, which can take place over many months and years (Lindsay & Gaskell, 2010; Speidenberg & Zevin, 2006). It is during this protracted process that the knowledge of lexical items gradually develops and shifts from receptive to productive mastery, where the learner can finally use the stored item in its proper context (Laufer, 1991; Laufer & Paribakht, 1998; Levitzky-Aviad & Laufer, 2013; Melka, 1997). A considerable amount of research has focused on vocabulary acquisition at a receptive level, and the general consensus is that reading and listening can substantially contribute to expanding L2 vocabulary knowledge (Hirsh & Nation, 1992; Horst, 2005; Horst, Cobb, & Meara, 1998; Nation, 2006; Rott, 1999; Pigada & Schmitt, 2006; Saragi, Nation, & Meister, 1978; Shu, Anderson, & Zhang, 1995; van Zeeland, 2017). However, a number of studies have shown that, in order for language learners to be able to make appropriate use of such knowledge, they have to move beyond just the comprehension of input. Laufer (2005) found that only 16 to 35 percent of the vocabulary knowledge of learners in her study was active and ready to be used productively. This led Laufer to conclude that relying solely on the receptive knowledge of words is insufficient for learners to be able to use them correctly in their proper context. In a similar study, Laufer and Paribakht (1998) examined the gap between the receptive and productive knowledge of adult English language learners through investigating three types of vocabulary knowledge (i.e., passive, controlled active, and free active). They observed that the productive knowledge of words developed more slowly and less predictably than the receptive knowledge. The findings from these studies suggest that there is a chasm between the receptive and productive knowledge of a word. This gap becomes even wider when we enter the territory of abstract words, such as those of academic vocabulary. According to Corson (1995, p. 180), abstract words, “created by the introduction of an academic culture of literacy”, demand more mental activation than those belonging to everyday meaning systems. As Schmitt (2014) states, the underlying reason for this mental activation being demanding is that many of the word knowledge aspects required for this process to take place are implicit in nature. That is to say, developing the ability to use abstract words requires having ample knowledge of the characteristics that are inextricably related to them. According to Nation (2013), characteristics such as register constraints, frequency intuitions, and collocational relationships are word knowledge aspects which are unlikely to be mastered solely through recognizing a word’s form and meaning. The above suggests that achieving active knowledge of abstract words can be very demanding, and that a deeper knowledge of words is required to gain mastery of productive skills such as writing. (See Schmitt, 2014 for a detailed discussion of vocabulary depth.)

The second consideration is that vocabulary is not merely composed of individual words. In fact, one of the major findings of corpus-based studies is the pervasive presence of multi-word items in the English language (Moon, 1997; Nattinger & DeCarrico, 1992; Schmitt, 2004, Sinclair, 1987, 1991; Wray, 2000, 2002). According to some estimates, these groups of words take up around 20 to 50 percent of spoken and written discourse (Biber, Johansson, Leech, Conrad, & Finegan, 1999; Erman & Warren, 2000; Foster, 2001; Howarth, 1998; Sorhus, 1977). While some of these sequences are intrinsically colloquial and informal in tone (e.g., phrasal verbs), others such as those included in the Academic Formulas List are frequently used in academic registers (Biber, *et al.*, 1999; Simpson-Vlach & Ellis, 2010). Research has shown that there are some differences between native and non-native speakers of English in terms of their active knowledge and productive use of these items. Schmitt, Grandage, and Adolphs (2004) investigated the psycholinguistic validity of recurrent word combinations by administering a number of multi-word items extracted from a corpus to a group of native and non-native speakers of English. They found that native participants, compared to their non-native counterparts, performed substantially better in terms of their accurate production of the target items. Based on these findings, the authors concluded that there is a substantial difference between native and non-native speakers of English in terms of the way in which they store, retrieve, and produce word strings, such as formulaic sequences. Such evidence suggests that the combinatory nature of words should not be considered a marginal phenomenon in vocabulary studies and that researchers should acknowledge “the fact that many English language users (e.g., L2 learners, L1 children) struggle to recognize, acquire, and utilize such items” (Gardner, 2007, p. 260).

We can conclude from this discussion that in order to gain a better understanding of academic vocabulary knowledge and its use among language speakers, both single- and multi-word components of such vocabulary should be investigated. Therefore, the present research explores the use of academic words and sequences in school essays produced by L1 and L2 adolescent writers. The following research questions guided the present study:

1. Is there a significant difference between L1 and L2 adolescent writers in terms of the diversity and richness of the lexical items used in their school essays?
2. Is there a significant difference between L1 and L2 adolescent writers in their use of general academic words from the Academic Word List (AWL)?
3. Is there any significant difference between L1 and L2 adolescent writers in their use of academic formulae from the Academic Formulas List (AFL)?

## **Corpus description and methodology**

## Corpora

The present study used data from two corpora: L1 corpus and L2 corpus. The L1 corpus was a sample of A-level essays extracted from the Louvain Corpus of Native English Essays (LOCNESS). The essays for this corpus were selected in such a way that their topics broadly corresponded to those of essays in the L2 corpus. The L2 corpus was collected from a group of English as a Second Language (ESL) students, attending English for Academic Purposes classes in an all-girls' school in Wellington, New Zealand (the classes were designed specifically for ESL students, no native speaker students were attending these classes). The girls (age range 16-18 years old, Year 12 and 13) were from a variety of first language backgrounds (Mandarin, Dari, Hindi, Gujarati, Khmer, Arabic, Greek). They immigrated to New Zealand with their families as teen-agers (between 10 and 16 years old), although most had spent between one and three years in New Zealand prior to the data collection taking place. The participants were unpaid volunteers. Because, at the time of data collection, most participants were under the age of 18, permission was sought from the school board and the families, where necessary. The research was carried out fulfilling ethical requirements in accordance with the standard procedures of the school and Victoria University of Wellington.

The essays in both corpora were argumentative in nature, covering potentially controversial topics, such as *the impact of health issues and poverty on education, depression, obesity, immigration, transport, lottery*, and so on, topics that require the writer to establish a position on the issue in question. A computer spellchecker was used to identify and correct orthographic errors in the L2 corpus. Following this, any text which did not appear to belong to the prose of the essays, such as references, direct quotations, and annotations were removed and were excluded from the final word count. On average, L2 essays were longer in length, which led to the inclusion of more essays into the L1 corpus with the aim of reaching a matched word count for both corpora. Table 1 provides a description of the two corpora used in this study.

**Table 1. Description of the corpora**

Corpus	Average Length	N. of Texts	Size
L1	309.6	34	22000
L2	970.6	21	22169

## Analysis

In order to address the three research questions of the study, the following characteristics of the two corpora were investigated: 1) lexical diversity and richness, 2) the use of academic words, and 3) the use of academic formulae.

In order to provide a lexical text analysis of the two corpora, the *VocabProfile-Compleat* BNC-COCA 1-25 programme (Cobb, n.d.) was used. The vocabulary

profiles of essays in each corpus were created and were then compared with regard to three factors: the overall number of word families and types in each corpus, the number of tokens per family and type, and the use of words from different frequency groups (i.e., K1-K25 frequency levels). Since the two corpora were closely matched for token number, any significant differences in the overall numbers of word families and types would be informative. Therefore, these characteristics were compared using Chi-square tests. In addition, the differences between the two corpora, in terms of the number of tokens per family and type, were estimated using *t*-tests. For this particular analysis, the relevant mean characteristics were calculated using the vocabulary profiles of the respective essays. Finally, in the case of frequency groups, contingency table tests were performed to analyze the association between each corpus and the frequency distribution of word families, types, and tokens from each frequency level. Following this, a series of pairwise comparisons were made between the two corpora in order to explore their differences in the use of words from different frequency groups. For these analyses, the frequency groups K1-K5 were compared individually, whereas K6-K25 groups were collapsed into two frequency categories: K6-K10 and K11-K25. This was deemed necessary because the total number of words in the K6-K25 groups was very low (i.e., below 100 tokens in each frequency group), and some of the frequency groups were not represented at all.

To address the second research question, the two corpora were compared to investigate the extent to which L1 and L2 adolescent writers draw on the academic words (AWs) from the AWL. Coxhead's (2000) AWL consists of the 570 most frequent word families (i.e., approximately 3000 words) in a wide range of academic disciplines. The AWL is included in the *VocabProfile* programme (Cobb, n.d.), a web-based adaptation of Heatley, Nation & Coxhead (2002), which was used to analyze the two corpora. The total number of AWL families, types, and tokens used in the two corpora were compared using Chi-square tests. Also, the differences between the mean type-token ratios (TTR) of the AWs were estimated using *t*-tests. Moreover, the text-based RANGE software (Cobb, n.d.) was employed to evaluate the degree of overlap between the two corpora in terms of the use of AWs from the AWL.

Finally, to address the third research question, the use of academic formulae (AFs) in the two corpora was investigated. The aim of the analysis was to compare the total number of AFs from the AFL in the corpora and to assess the overlap. The AFL includes 207 phrases common in both academic written and spoken language (Core AFL), as well as 200 word combinations exclusive to academic written language (Written AFL) and 200 sequences specific to academic spoken language (Spoken AFL). In line with the general aim of the study (i.e., to compare L1 and L2 essays in terms of the AFs used), only the items included in the Core and Written AFL were used. In addition, following Simpson-Vlach and Ellis (2010), we too classified the AFs in the two corpora into three functional groups: referential, stance, and discourse. The total number of AFs in each corpus was compared, as well as the

number of AFs belonging to each of these three groups. The analysis used the same statistical methods employed for AWs.

## Results

### Lexical diversity and richness

As stated above, the two corpora were closely matched for token number. The L2 corpus contained only 169 words more than the L1 corpus, which did not generate a statistically significant difference ( $\chi^2=0.64663$ ,  $df=1$ ,  $p=0.4213$ ). However, as shown in Table 2, the differences in the number of word families and types were much more noticeable and were found to be statistically significant ( $\chi^2=36.43$ ,  $df=2$ ,  $p<0.001$ ). In fact, the L1 corpus was found to contain significantly more word families ( $\chi^2=26.59$ ,  $df=1$ ,  $p<0.001$ ) and word types ( $\chi^2=13.00$ ,  $df=1$ ,  $p<0.001$ ) than the L2 corpus. Table 2 presents the use of word families, types, and tokens in the two corpora.

**Table 2** *The total number of word tokens, types and families in the corpora*

Corpus	Tokens	Types	Families
L1	22000	3446	2073
L2	22169	3156	1754

As expected from the above results, the type-token ratio (TTR) of L1 essays was found to be significantly higher than that of L2 essays (L1 mean TTR=0.433,  $SD=0.074$ , L2 mean TTR=0.331,  $SD=0.057$ ;  $t=5.817$ ,  $df=52.577$ ,  $p<0.001$ ). Also, comparing the mean numbers of tokens per family and tokens per type in the two corpora showed similar results (see Table 3). However, no significant differences between the mean types per family ratios were found.

**Table 3.** *The use of word families, types and tokens by L1 and L2 writers*

Ratio	L1 mean ( $SD$ )	L2 mean ( $SD$ )	$t$ value	$df$	$p$ value
Tokens per family	2.735 (.464)	3.507 (.663)	-4.872	33.941	<.001
Tokens per type	2.374 (.397)	3.096 (.522)	-5.558	36.12	<.001
Types per family	1.139 (.040)	1.135 (.047)	.319	39.003	.751

Following this analysis, we examined the lexical frequency profiles of the two corpora. The analysis revealed significant differences in the use of words belonging to different frequency groups (see Table 4). A contingency table Chi-square test also showed that the number of tokens from each frequency group was not independent of the corpus ( $\chi^2=37.752$ ,  $df=6$ ,  $p<0.001$ ).

As can be seen in Table 4, the two corpora differ in terms of the use of word tokens from different frequency groups. While the two corpora showed no significant differences in the use of word tokens from K1 and K2 levels ( $\chi^2=4.041$ ,  $df=1$ ,

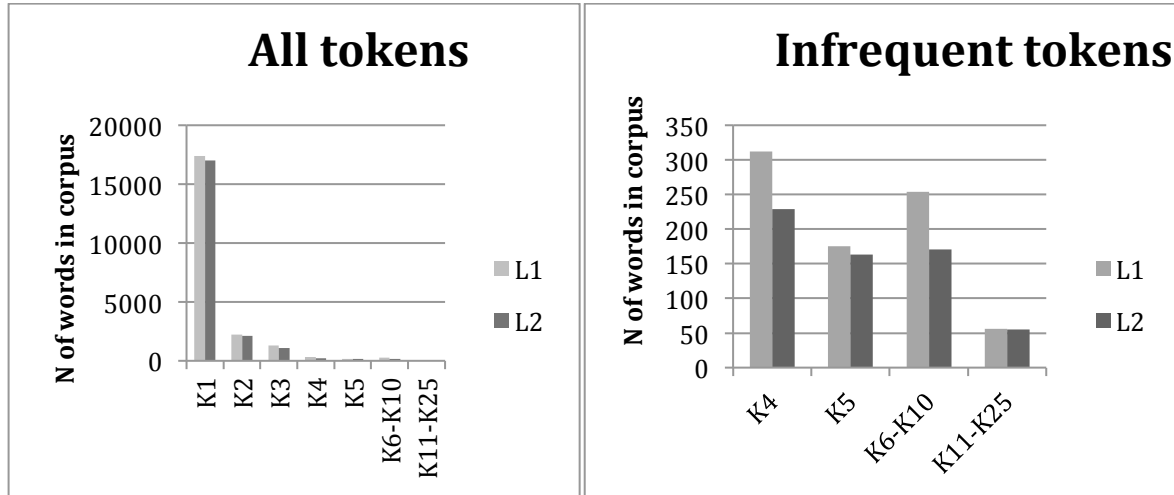
$p=0.044$  for K1 tokens,  $\chi^2=2.366$ ,  $df=1$ ,  $p=0.124$  for K2 tokens), the L1 corpus was found to contain significantly more word tokens from lower-frequency levels, such as K3 ( $\chi^2=20.089$ ,  $df=1$ ,  $p<0.001$ ), K4 ( $\chi^2=12.902$ ,  $df=1$ ,  $p<0.001$ ) and K6-K10 ( $\chi^2=16.294$ ,  $df=1$ ,  $p<0.001$ ). However, the differences between the numbers of K5 tokens, as well as tokens belonging to K11-K25 levels appeared to be statistically insignificant ( $\chi^2=0.276$ ,  $df=1$ ,  $p=0.599$  for K5 tokens,  $\chi^2=0.009$ ,  $df=1$ ,  $p=0.924$  for K11-K25 tokens).

**Table 4. Lexical frequency profiles of L1 and L2 corpora**

Frequency level	Family (%)		Type (%)		Token (%)		Cumulative % tokens	
	L1	L2	L1	L2	L1	L2	L1	L2
K1	761 (36.71)	711 (40.53)	1460 (42.37)	1331 (42.17)	17398 (79.08)	17024 (76.79)	79.08	76.79
K2	461 (22.24)	432 (24.62)	760 (22.05)	714 (22.62)	2206 (10.03)	2105 (9.50)	89.11	86.29
K3	406 (19.59)	339 (19.32)	581 (16.86)	462 (14.64)	1282 (5.83)	1067 (4.81)	94.94	91.10
K4	156 (7.53)	99 (5.64)	184 (5.34)	118 (3.74)	312 (1.42)	229 (1.03)	96.36	92.13
K5	93 (4.49)	59 (3.36)	106 (3.08)	71 (2.25)	175 (.80)	163 (.74)	97.16	92.87
K6	53 (2.56)	33 (1.88)	55 (1.60)	38 (1.20)	113 (.51)	73 (.33)	97.67	93.20
K7	38 (1.83)	22 (1.25)	40 (1.16)	24 (.76)	50 (.23)	34 (.15)	97.90	93.35
K8	30 (1.45)	12 (.68)	33 (.96)	14 (.44)	45 (.20)	25 (.11)	98.10	93.46
K9	26 (1.25)	11 (.63)	27 (.78)	12 (.38)	31 (.14)	15 (.07)	98.24	93.53
K10	11 (.53)	6 (.34)	11 (.32)	7 (.22)	15 (.07)	24 (.11)	98.31	93.64
K11	10 (.48)	8 (.46)	11 (.32)	8 (.25)	12 (.05)	10 (.05)	98.36	93.69
K12	7 (.34)	4 (.23)	7 (.20)	4 (.13)	7 (0.03)	4 (.02)	98.39	93.71
K13	7 (.34)	2 (.11)	7 (.20)	2 (.06)	8 (.04)	2 (.01)	98.43	93.72
K14	4 (.19)	3 (.17)	5 (.15)	3 (.10)	10 (.05)	15 (.07)	98.48	93.79
K15	4 (.19)	2 (.11)	4 (.12)	2 (.06)	5 (.02)	5 (.02)	98.50	93.81
K16	3 (.14)	2 (.11)	3 (.09)	2 (.06)	11 (.05)	2 (.01)	98.55	93.82
K17		2 (.11)		2 (.06)		2 (.01)		93.83
K18	1 (.05)	2 (.11)	1 (.03)	2 (.06)	1 (.00)	7 (.03)		93.86
K19	1 (.05)	2 (.11)	1 (.03)	2 (.06)	1 (.00)	3 (.01)		93.87
K20		1 (.06)		1 (.03)		2 (.01)		93.88
K21	1 (.05)	1 (.06)	1 (.03)	1 (.03)	1 (.00)	1 (.00)		93.88
K22								
K23								
K24		1 (.06)		1 (.03)		2 (.01)		93.89
K25								
Off-list			149	335	317	1355 (6.11)	100.00	100.00

			(4.32)	(10.61)	(1.44)			
Total	2073	1754	3446	1756	22000	22169 (100)	100.00	100.00
			(100)	(100)	(100)			

Figure 1 illustrates the frequency distributions of word tokens in the two corpora.



**Figure 1.** Frequency distributions of word tokens in L1 and L2 corpora.

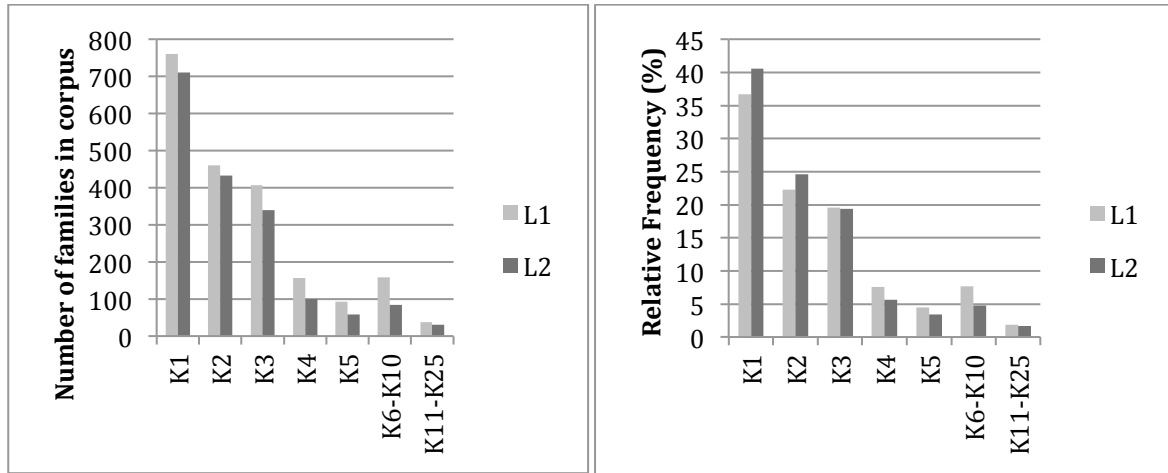
Similar results were observed for word families and types. The Chi-square tests for contingency tables found significant relationships between each corpus and the number of word families and types from each frequency group ( $\chi^2=26.254$ ,  $df=6$ ,  $p<0.001$  for types,  $\chi^2=26.173$ ,  $df=6$ ,  $p<0.001$  for families). As was the case with tokens, most of the significant differences in the frequency distribution concerned less frequent words. As is shown in Table 5, the L1 corpus included significantly more word families and types from K3-K10 levels. Table 5 presents the differences between the two corpora in terms of the use of word families and types from different frequency groups.

**Table 5.** The differences in the frequency distributions of families and types

Frequency level	Types			Families		
	L1	L2	$\chi^2$	L1	L2	$\chi^2$
K1	1460	1331	6.475*	761	711	1.960
K2	760	714	1.681	461	432	1.082
K3	581	462	13.934**	406	339	6.254*
K4	184	118	14.391**	156	99	12.655**
K5	106	71	6.765*	93	59	7.392*
K6-K10	166	95	19.197**	158	84	22.457**
K11-K25	40	30	1.225	38	30	0.774

\* $p<0.05$  \*\* $p<0.001$

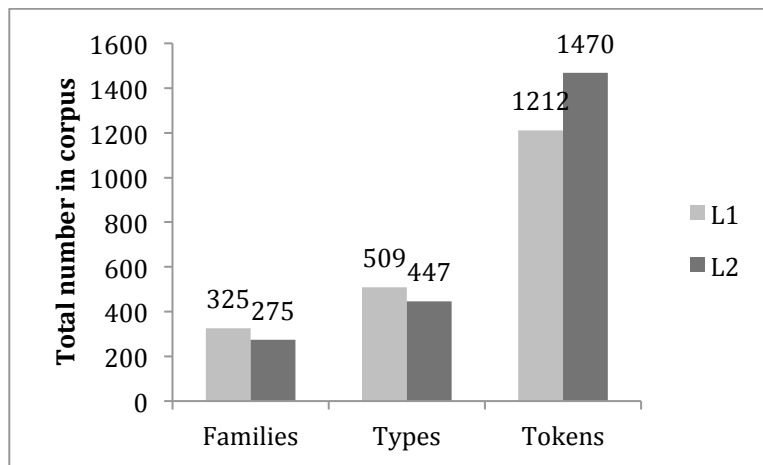
In addition, we also calculated the relative frequency distribution of word families in the two corpora. The results showed that, compared to the L1 corpus, the distribution of word families in the L2 corpus was more skewed towards families from higher-frequency levels (i.e., K1 and K2). The significance of these findings were all confirmed at the .01% level using a test of equal or given proportions ( $\chi^2=143.11$ ,  $df=1$ ,  $p<0.001$  for K1 families,  $\chi^2=66.503$ ,  $df=1$ ,  $p<0.001$  for K2 families).



**Figure 2.** Frequency distributions of word families in the two corpora

### The use of academic vocabulary

As shown in Figure 3, compared to the L2 corpus which contains more AWL tokens, the L1 corpus appears to include more AWL families and types. The results of pairwise Chi-square tests revealed that these differences were all significant at the 5% level



**Figure 3.** The number of AWL units in the two corpora

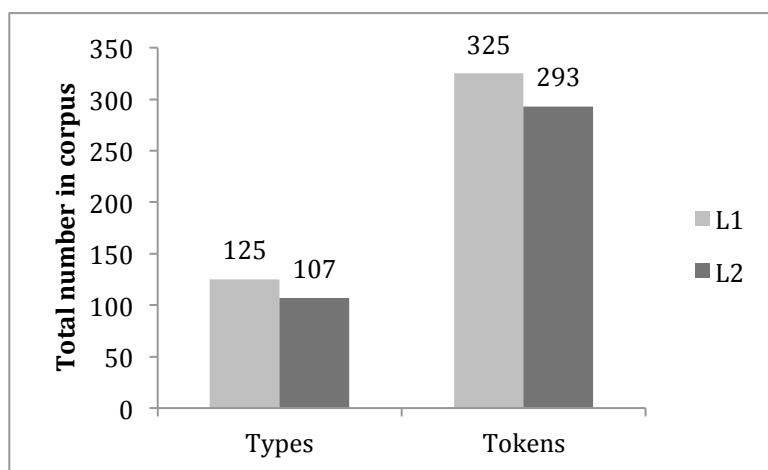
( $\chi^2=21.391$ ,  $df=1$ ,  $p<0.001$  for tokens,  $\chi^2=4.278$ ,  $df=1$ ,  $p=0.039$  for types,  $\chi^2=4.324$ ,  $df=1$ ,  $p=0.038$  for families). Also, the results of a Chi-square test of independence

confirmed that the number of AWL words was not independent of the corpus ( $\chi^2=28.01$ ,  $df=2$ ,  $p<0.001$ ). These results were consistent with the significantly higher mean TTR found in the L1 corpus (L1 mean=0.789,  $SD=0.117$ , L2 mean=0.610,  $SD=0.169$ ,  $t=4.3463$ ,  $df=33.694$ ,  $p<0.001$ ).

It is further worth noting that a considerable number of AWL items overlapped between the two corpora. In particular, 189 AWL families appeared in both corpora, which comprised 58.15% of the AWL families in the L1 corpus and 68.73% of the AWL families in the L2 corpus. The proportion of the overlap in the overall number of AWL types, however, was found to be lower. Specifically, 196 AWL types which appeared to take up 38.51% of all the AWL types in the L1 corpus and 43.85% of the AWL types in the L2 corpus.

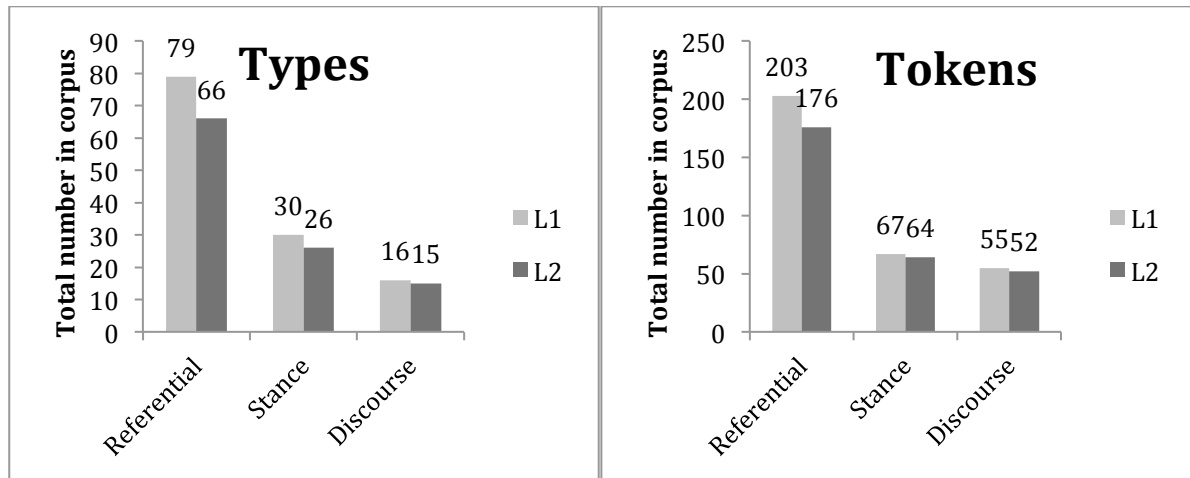
### The use of academic formulae

In contrast to the results reported in the previous section, the analysis of the use of AFs from the AFL revealed a surprising degree of similarity between the two corpora. First, the total number of AF types and tokens used in each corpus was found to be very similar in both corpora. Although the overall use of AF types and tokens was higher in the L1 corpus, the differences between the two corpora were not statistically significant ( $\chi^2=1.371$ ,  $df=1$ ,  $p=0.242$  for types,  $\chi^2=1.775$ ,  $df=1$ ,  $p=0.183$  for tokens). Figure 4 illustrates the number of AF types and tokens used in each corpus.



**Figure 4. The number of academic formulae in L1 and L2 corpora**

The analysis of the two corpora in terms of the use of AFs from different functional categories (i.e., referential, discourse, and stance) also revealed a similar trend. As can be seen from Figure 5, the number of referential, stance, and discourse AFs (both types and tokens) used in the L1 corpus was slightly higher than that in the L2 corpus. The differences, however, were not statistically significant ( $\chi^2=0.087$ ,  $df=2$ ,  $p=0.957$  for types,  $\chi^2=0.312$ ,  $df=2$ ,  $p=0.855$  for tokens).



**Figure 5. The distribution of academic formulae by functional category**

In addition, a high proportion of AF types was found to overlap in the two corpora (i.e., 76 AF types, which comprised 60.8% of the AFs in the L1 corpus and 71.03% of the AFs in the L2 corpus). Further, a high degree of overlap was found in the use of AF types from each of the three categories. Table 6 provides information about the degree of overlap between the two corpora in terms of the use of AF types from all the three functional categories.

**Table 6. Degree of overlap between L1 and L2 corpora across three categories**

Functional categories	N. shared AF types	L1 (%)	L2 (%)
Referential	49	62.02	74.24
Stance	18	60	69.23
Discourse	9	56	60

Moreover, the analysis of the frequency of occurrence of the AFs found in each corpus revealed interesting results. As can be seen from Table 7, six of the 12 most frequent AFs used in each corpus were found to be common between the two corpora (e.g., *a/the number of*, *due to the*, *for example*, *as well as*, *it is not*, *in order to*). It is also worth noting that both corpora contained a number of low-frequency AFs which occurred only once in each corpus (i.e., 56 AFs in the L1 corpus and 52 AFs in the L2 corpus). Among these low-frequency AFs, 12 were also found to be shared by the two corpora (e.g., *a variety of*, *exactly the same*, *in the form of*, *means that the*). However, it should be noted that the AFs which appeared only once in either of the two corpora were not found in the high-frequency list of the other corpus.

**Table 7. *The most frequent AFs in the two corpora (token frequency > 5)***

Top 12 AFs in the L1 corpus		Top 12 AFs in the L2 corpus	
AF	N tokens	AF	N tokens
a/the number of	16	a/the number of	24
due to the	12	in order to	12
for example	11	is that	11
such as the	10	there is a/an/no	9
they [did/do] not	8	this is [a/an/not]	9
(more) likely to (be)	8	for example	8
as well as	8	is not [a/the]	8
it is not	7	it is not	8
in order to	7	(the) fact that (the)	7
that there [are/is (a)]	7	the use of	7
(as) part of [a/the]	6	as well as	7
is that [it/the/there]	6	due to the	7

## Discussion

This study aimed to investigate the production of academic words and sequences in compositions written by school-age L1 and L2 students. Our main focus was to determine the degree to which students in K-12 settings make use of academic vocabulary in their essays, as well as the extent to which L1 and L2 adolescent writers differ in terms of the use of these words and sequences.

To explore these differences, we first analyzed the two corpora in terms of their lexical diversity and lexical richness. Our results showed that the diversity of the words used in L2 essays was significantly lower compared to that of L1 essays. This finding suggests that the L1 adolescent writers in this study seemed to be better at expressing themselves with the vocabulary knowledge that they had at their disposal. As Laufer (1994) points out, a high level of lexical diversity “shows how well a person can express himself with whatever vocabulary he has” (p. 30). Following this, we also compared the two corpora in terms of their lexical profiles. Our analyses showed that the use of word families and types from K3-K10 frequency levels was significantly lower in L2 essays. Schmitt and Schmitt (2012) considered the frequency levels of K4 to K9 as mid-frequency groups and emphasized the prominent role of these words in ensuring effective communication. In addition, the authors also stress that the K3 level is the last frequency group which can be considered as ‘high-frequency’, and that this is the frequency level from which learners’ performance starts to dwindle. Our findings support this assertion as the differences between the two corpora started to reach statistical significance from the K3 level, with the L2 writers making significantly less use of word families and types from this particular

frequency group. Schmitt and Schmitt also propose that anything beyond the first 9,000 word families should be considered as low-frequency vocabulary, which is “not typically useful enough to warrant an explicit focus” (p. 498). Therefore, according to our findings, it appears that the words belonging to the frequency bands of K3 to K9 should receive extra attention from content area teachers teaching L2 students. This conclusion is supported by the findings of previous research regarding the lack of attention to these particular groups of words in ESL classrooms (Horst, 2010; Horst, Collins, & Cardoso, 2009; cited in Schmitt & Schmitt, 2012).

We further found that the two corpora were comparable in terms of the use of word families and types from lower-frequency levels, such as K11 to K25. In fact, our results showed that both L1 and L2 writers made limited use of words from these particular frequency groups. This is not surprising as the words belonging to these frequency levels are relatively rare in discourse and that it is unlikely for school-age students to have multiple exposures to such words through every-day activities outside the classroom. However, a word of caution is needed when interpreting these findings. That is, we do not necessarily suggest a shift towards teaching low-frequency words in content classrooms. In the analysis of the vocabulary size required for understanding high-school level science textbooks, Coxhead, Stevens, and Tinkle (2010) found that learners equipped with the knowledge of words from K9 frequency groups (i.e., mid-frequency word levels) would be able to understand 98.10% of the running words presented in these books. The authors also showed that many of the remaining unfamiliar words were technical words related to the analyzed subject area (i.e., science). Therefore, it can be concluded that, instead of teaching some random low-frequency words only because they are missing in learners’ vocabulary, attention should instead be directed to those that play a supportive role in understanding the text.

Our further analyses focused on the differences between the two corpora in terms of the use of AWs from the AWL. We found that the essays written by L2 writers contained a significantly lower number of AW families and types. Our results also showed that the degree of diversity in the use of these words was significantly higher in the L1 corpus than L2 corpus. These findings are important in that they indicate the need for more attention to academic vocabulary needs of L2 school-age students in content classrooms. The importance of such consideration is echoed in a study conducted by Townsend, Filipini, Colins, and Biancarosa (2012), who investigated the importance of academic word knowledge in the academic success of linguistically and socio-economically diverse middle-school students. The authors found an empirical link between students’ general academic vocabulary knowledge and their academic achievement, highlighting the importance of heightening teachers’ awareness of the decisive role of academic words in ESL students’ school success. Considering the high stakes involved in learning academic words for ESL adolescent students, one can interpret these findings as suggesting that this particular demographic can greatly benefit from appropriate scaffolding focused on the explicit

instruction of general academic words. In doing so, teachers should be aware of the fact that, in order for students to be able to use academic words in their writing, they need multiple exposures to these words, for which there are limited opportunities in K-12 settings compared to tertiary education contexts (Nagy & Townsend, 2012).

Finally, following the analysis of academic words in the two corpora, we compared the compositions with regard to the use of academic sequences from the AFL. Unlike the results reported for AWs, the use of academic multi-word items was found comparable in L1 and L2 essays. In fact, the analysis of the top 12 most frequent AFs used in each corpus revealed a considerable degree of overlap between the two corpora. Further, by looking at the functions of the used AFs in each corpus, we found that both L1 and L2 writers made more use of referential sequences, which is in line with the findings of previous research, indicating the ubiquity of these word combinations in academic texts (Biber, *et al.*, 1999; Biber, 2006; Chen & Baker, 2010). These findings point to the fact that L2 writers in this study, just as their L1 counterparts, knew and were able to use academic sequences in their writing. Interestingly, these results appear to go against those reported in the literature regarding the underuse of multi-word items among L2 learners (Adel & Erman, 2012; Siyanova & Schmitt, 2007, Chen & Baker, 2010; Durrant & Schmitt, 2009; Granger, 1998; Wray, 2002). However, this contradictory evidence can be explained with reference to two intertwined factors in second language acquisition: age and exposure to an L2 environment. As explained in previous sections, the L2 participants from whom the data were aggregated were all immigrant school-age students, growing up in a naturalistic L2 environment (i.e., New Zealand). As opposed to ESL/EFL adult learners, L2 children of this age learn language faster and, as a result of their integration into the L2 environment, they mostly adopt a native-like approach towards acquiring different properties of the language, such as formulaic sequences. In an in-depth analysis of the effect of sociocultural adaptation on the acquisition of formulaic sequences, Dornyei, Durow, and Zahran (2004) found that acculturation and integration into a native English-speaking environment can play a pivotal role in the acquisition of formulaic sequences by L2 learners. This level of engagement with the target language appears to have substantial linguistic gains for school-age children. According to Kuiper, Columbus, and Schmitt (2009), during this age, the language used in social activities becomes more sophisticated and involves the use of phrasal items. This process can be facilitative to directing children's attention to the highly phrasal nature of language and help them build an intuition for storing and retrieving authentic sequences. Therefore, in the case of our findings, it can be argued that, although the analyzed sequences (i.e., academic formulae) are unlikely to be encountered in activities outside the classroom, it is possible for L2 students in this study to have developed a preference for 'picking up' and utilizing conventional sequences as a result of their sustained engagement with the target language in a naturalistic L2 environment. This conclusion appears to be in line with the findings of Siyanova-Chanturia and Spina (2015), who found a

relationship between rich language exposure and L2 speaker intuitions of units above the word level.

## Conclusion

The crucial role of academic vocabulary in achieving academic literacy applies not only to single words but also to multi-word items. The results presented in this study offer some insights into this assertion by demonstrating that different conceptualizations of what constitutes a vocabulary can affect our understanding of learners' vocabulary competence in certain registers. That is to say, although the ESL learners in the present investigation exhibited poorer knowledge of academic vocabulary at the word level compared to their L1 peers, the analysis of their performance in using the phrasal aspect of academic vocabulary yielded a pattern of results comparable to L1 speakers. This is in line with previous research indicating that ignoring the phrasal aspect of the mental lexicon can have important ramifications for the validity of pedagogical theories and practices in the area of vocabulary (Gardner, 2007). Taken together, our findings, albeit involving some limitations – such as the use of small corpora – should be viewed as an attempt to underscore the importance of considering both single words and multi-word items in the research into academic vocabulary.

## References

- Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*, 31(2), 81-92. DOI: 10.1016/j.esp.2011.08.004
- Bailey, A. L. (2007). Introduction: Teaching and assessing students learning English in school. In A. L. Bailey (Ed.), *The language demands of school: Putting academic English to the test* (pp. 1–28). New Haven, CT: Yale University Press.
- Baumann, J. F., & Graves, M. F. (2010). What is academic vocabulary? *Journal of Adolescent & Adult Literacy*, 54(1), 4-12. DOI:10.1598/JAAL.54.1.1
- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam: Benjamin.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at . . . : Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371-405. DOI: 10.1093/applin/25.3.371
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow, England: Longman.
- Blachowicz, C. L. Z., & Fisher, P. (2000). Vocabulary instruction. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 503-523). Mahwah, NY: Erlbaum.
- Brock, C., Lapp, D., Salas, R. G., & Townsend, D. (2009). *Academic literacy for English learners: High quality instruction across content areas*. New York: Teachers College Press.
- Campion, M. E., & Elley, W. B. (1971). *An Academic Word List*. Wellington. New Zealand: New Zealand Council for Educational Research.
- Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning and Technology*, 14(2), 30-49.

- Cobb, T. (2009). The compleat lexical tutor (version 8.0) [Computer software]. Retrieved from <http://www.lextutor.ca/>
- Cobb, T. Web Vocabprofile [accessed May 2016 from <http://www.lextutor.ca/vp/> ], an adaptation of Heatley, Nation & Coxhead's (2002) Range.
- Cohen, A., Glasman, H., Rosenbaum-Cohen, P. R., Ferrara, J., & Fine, J. (1988). Reading English for specialized purposes: Discourse analysis and the use of standard informants. In P. Carrell, J. Devine, & D. Eskey (Eds.), *Interactive approaches to second language reading* (pp. 152–167). Cambridge: Cambridge University Press.
- Corson, D. (1997). The learning and use of academic English words. *Language Learning*, 47(4), 671–718. DOI: 10.1111/0023-8333.00025
- Cowan, J. R. (1974). Lexical and syntactic research for the design of EFL reading materials. *TESOL Quarterly* 8(4), 389–399. DOI: 10.2307/3585470
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213–238. DOI: 10.2307/3587951
- Coxhead, A., & Nation, I. S. P. (2001). The specialized vocabulary of English for academic purposes. In J. Flowerdew & M. Peacock (Eds.), *Research perspectives on English for academic purposes* (pp. 252–267). Cambridge: Cambridge University Press.
- Coxhead, A., Stevens, L., & Tinkle, J. (2010). Why might secondary science textbooks be difficult to read? *New Zealand Studies in Applied Linguistics*, 16(2), 35–52.
- Dörnyei, Z., Durow, V., & Zahran, K. (2004). Individual differences and their effects on formulaic sequence acquisition. In N. Schmitt (Ed.), *Formulaic sequences* (pp. 87–106). Amsterdam: John Benjamins. DOI: 10.1075/llt.9.06dor
- Durrant, P. (2014). Discipline and level specificity in university students' written vocabulary. *Applied Linguistics*, 35(3), 328–356. DOI: 10.1093/applin/amt016
- Durrant, P. (2016). To what extent is the Academic Vocabulary List relevant to university student writing? *English for Specific Purposes*, 43, 49–61. DOI: 10.1016/j.esp.2016.01.004
- Durrant, P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *IRAL-International Review of Applied Linguistics in Language Teaching*, 47(2), 157–177. DOI: 10.1515/iral.2009.007
- Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text*, 20(1), 29–62. DOI: 10.1515/text.1.2000.20.1.29
- Farrell, P. (1990). Vocabulary in ESP: *A lexical analysis of the English of electronics and a study of semi-technical vocabulary* (CLCS Occasional Paper No. 25). Dublin, Ireland: Trinity College, Centre for Language and Communication Studies.
- Foster, P. (2001). Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching, and testing* (pp. 75–94). Harlow, England: Longman.
- Gardner, D. (2007). Validating the construct of word in applied corpus-based vocabulary research: A critical survey. *Applied Linguistics*, 28(2), 241–265. DOI: 10.1093/applin/amm010
- Gardner, D., & Davies, M. (2013). A new academic vocabulary list. *Applied Linguistics*, 35(3), 305–327. DOI: 10.1093/applin/amt015
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A.P. Cowie (Ed.), *Phraseology: Theory, Analysis, and Applications* (pp. 79–199). Oxford: Oxford University Press.
- Hakuta, K., Butler, Y.G., & Witt, D. (2000). *How long does it take English learners to attain proficiency?* Santa Barbara, CA: University of California Linguistic Minority Research Institute.
- Heatley, A., Nation, I.S.P. & Coxhead, A. (2002). RANGE and FREQUENCY programs. Available at <http://www.victoria.ac.nz/lals/staff/paul-nation.aspx>

- Hiebert, E. H., & Lubliner, S. (2008). The nature, learning, and instruction of general academic vocabulary. In A. E. Farstrup & S. J. Samuels (Eds.), *What research has to say about vocabulary instruction* (pp. 106–129). Newark, DE: International Reading Association.
- Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language*, 8(2), 689-689.
- Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *Canadian Modern Language Review*, 61(3), 355-382. DOI: 10.3138/cmlr.61.3.355
- Horst, M. (2010). How well does teacher talk support incidental vocabulary acquisition? *Reading in a Foreign Language* 22(1), 161–180.
- Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, 11(2), 207-223.
- Horst, M., Collins, L., & Cardoso, W. (2009). Focus on vocabulary in ESL teacher talk. Paper presented at the annual conference of the American Association for Applied Linguistics, Denver, CO, March 2009.
- Howarth, P. (1998). The phraseology of learners' academic writing. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and application* (pp. 161-186). Oxford, England: Oxford University Press.
- Hyland, K., & Tse, P. (2007). Is there an “academic vocabulary”? *TESOL Quarterly*, 41(2), 235-253. DOI: 10.1002/j.1545-7249.2007.tb00058.x
- Ishii, T., & Schmitt, N. (2009). Developing an integrated diagnostic test of vocabulary size and depth. *RELC Journal*, 40(1), 5-22. DOI:10.1177/0033688208101452
- Kuiper, K., Columbus, G., & Schmitt, N. (2009). The acquisition of phrasal vocabulary. In S. Foster-Cohen (Ed), *Language acquisition* (pp. 216-240). Basingstoke: Palgrave Macmillan. DOI: 10.1057/9780230240780\_10
- Laufer, B. (1991). The development of L2 lexis in the expression of the advanced learner. *The Modern Language Journal*, 75(4), 440-448. DOI: 10.1111/j.1540-4781.1991.tb05380.x
- Laufer, B. (1994). The lexical profile of second language writing: does it change over time? *RELC journal*, 25(2), 21-33. DOI: 10.1177/003368829402500202
- Laufer, B. (2005). Focus on form in second language vocabulary learning. *Eurosla yearbook*, 5(1), 223-250. DOI: 10.1075/eurosla.5.11lau
- Laufer, B., & Nation, I. S. P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307-322. DOI: 10.1093/applin/16.3.307
- Laufer, B., & Paribakht, T. S. (1998). The relationship between passive and active vocabularies: Effects of language learning context. *Language Learning*, 48(3), 365-391. DOI: 10.1111/0023-8333.00046
- Lee, S. H. (2003). ESL learners' vocabulary use in writing and the effects of explicit vocabulary instruction. *System*, 31(4), 537-561. DOI: 10.1016/j.system.2003.02.004
- Lesaux, N. K., Kieffer, M. J., Kelley, J. G., & Harris, J. R. (2014). Effects of academic vocabulary instruction for linguistically diverse adolescents: Evidence from a randomized field trial. *American Educational Research Journal*, 51(6), 1159-1194. DOI: 10.3102/0002831214532165
- Levitzky-Aviad, T., & Laufer, B. (2013). Lexical properties in the writing of L2 learners over eight years of study: single words and collocations. *Eurosla Monographs Series*, 2, 127-148.
- Li, J., & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing*, 18(2), 85-102. DOI: 10.1016/j.jslw.2009.02.001
- Lindsay, S., & Gaskell, M. G. (2010). A complementary systems account of word learning in L1 and L2. *Language Learning*, 60(s2), 45-63. DOI: 10.1111/j.1467-9922.2010.00600.x
- Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics*, 33(3), 299-320. DOI: 10.1093/applin/ams010

- McKeown, M. G., Beck, I. L., Omanson, R. C., & Pople, M. T. (1985). Some effects of the nature and frequency of vocabulary instruction on the knowledge and use of words. *Reading Research Quarterly*, 20(5), 522–535. DOI:10.2307/747940
- Melka, F. (1997). Receptive vs. productive aspects of vocabulary. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 84-102). New York, NY: Cambridge University Press.
- Moon, R. (1997). Vocabulary connections: Multi-word items in English. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 40-63). Cambridge, England: Cambridge University Press.
- Muncie, J. (2002). Process writing and vocabulary development: Comparing lexical frequency profiles across drafts. *System*, 30(2), 225-235. DOI: 10.1016/S0346-251X(02)00006-4
- Nagy, W., & Townsend, D. (2012). Words as tools: Learning academic vocabulary as language acquisition. *Reading Research Quarterly*, 47(1), 91-108. DOI: 10.1002/RRQ.011
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63(1), 59-82. DOI: 10.3138/cmlr.63.1.59
- Nation, I. S. P. (2013). *Learning Vocabulary in Another Language* (2nd ed.). Cambridge: Cambridge University Press.
- Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford, England: Oxford University Press.
- Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language*, 18(1), 1-28.
- Praninskas, J. (1972). *American university word list*. London: Longman.
- Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition*, 21(4), 589-619.
- Saragi, T., Nation, I. S. P., & Meister, G. F. (1978). Vocabulary learning and reading. *System*, 6(2), 72-78. DOI: 10.1016/0346-251X(78)90027-1
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- Schmitt, N. (Ed.). (2004). *Formulaic sequences: Acquisition, processing, and use*. Amsterdam: John Benjamins Publishing.
- Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, 64(4), 913-951. DOI: 10.1111/lang.12077
- Schmitt, N., Grandage, S., & Adolphs, S. (2004). Are corpus-derived recurrent clusters psycholinguistically valid. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 127-151). Amsterdam: John Benjamins Publishing.
- Schmitt, N., & Schmitt, D. (2012). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, 47(4), 484-503. DOI: 10.1017/S0261444812000018
- Seidenberg, M. S., & Zevin, J. D. (2006). Connectionist models in developmental cognitive neuroscience: Critical periods and the paradox of success. In Y. Munakata & M. H. Johnson (Eds.), *Attention and performance XXI: Processes of change in brain and cognitive development* (pp. 315–347). Oxford: Oxford University Press.
- Short, D. J., & Fitzsimmons, S. (2007). *Double the work: Challenges and solutions to acquiring language and academic literacy for adolescent English language learners*. Washington, DC: Alliance for Excellent Education.
- Shu, H., Anderson, R. C., & Zhang, H. (1995). Incidental learning of word meanings while reading: A Chinese and American cross-cultural study. *Reading Research Quarterly*, 76-95. DOI: 10.2307/747745
- Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31(4), 487-512. DOI: 10.1093/applin/amp058

- Sinclair, J. (1987). Collocation: a progress report. In R. Steele and T. Threadgold (Eds.), *Language Topics: Essays in Honour of Michael Halliday Vol. II* (pp. 319-331). John Benjamins, Amsterdam.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford, England: Oxford University Press.
- Siyanova, A., & Schmitt, N. (2007). Native and nonnative use of multi-word vs. one-word verbs. *IRAL-International Review of Applied Linguistics in Language Teaching*, 45(2), 119-139. DOI: 10.1515/IRAL.2007.005
- Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *Canadian Modern Language Review*, 64(3), 429-458. DOI: 10.3138/cmlr.64.3.429
- Siyanova-Chanturia, A., & Spina, S. (2015). Investigation of native speaker and second language learner intuition of collocation frequency. *Language Learning*, 65(3), 533-562. DOI: 10.1111/lang.12125
- Snow, C., & Kim, Y. (2007). Large problem spaces: The challenge of vocabulary for English language learners. In R.K. Wagner, A. E. Muse, & K. R. Tannenbaum (Eds.), *Vocabulary acquisition: Implications for reading comprehension* (pp. 123-139). New York: Guilford.
- Sorhus, H. B. (1977). To hear ourselves: Implications for teaching English as a second language. *English Language Teaching Journal*, 31(3), 211-221.
- Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal*, 36(2), 139-152. DOI: 10.1080/09571730802389975
- Stahl, S. A., & Fairbanks, M. M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research*, 56(1), 72-110.
- Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes*, 12(3), 214-225. DOI: 10.1016/j.jeap.2013.05.002
- Townsend, D., & Collins, P. (2009). Academic vocabulary and middle school English learners: An intervention study. *Reading and Writing*, 22, 993-1019. DOI:10.1007/s11145-008-9141-y
- Townsend, D., Filippini, A., Collins, P., & Biancarosa, G. (2012). Evidence for the importance of academic word knowledge for the academic achievement of diverse middle school students. *The Elementary School Journal*, 112(3), 497-518. DOI: 10.1086/663301
- Université catholique de Louvain, Centre for English Corpus Linguistics. (2010, April 5). *Centre for English Corpus Linguistics: LOCNESS*. Retrieved May 2014, from <http://www.uclouvain.be/en-cecl-locness.html>
- van Zeeland, H. (2017). Christopher Brumfit Thesis Award Winner 2014-Hilde van Zeeland: Four studies on vocabulary knowledge in and from listening: Findings and implications for future research. *Language Teaching*, 50(1), 143-150.
- Webb, S., & Nation, I. S. P. (2008). Evaluating the vocabulary load of written text. *TESOLANZ Journal*, 16, 1-10.
- Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied linguistics*, 21(4), 463-489. DOI: 10.1093/applin/21.4.463
- Wray, A. (2002). *Formulaic language and the lexicon, Vol. 5*. Cambridge, England: Cambridge University Press.
- Xue, G., & Nation, I. S. P. (1984). A university word list. *Language Learning and Communication*, 3(2), 215-229.