

An Analysis of Research in Academic Prose between Native Speakers and Chinese Learners

Shen Lifan¹ & Chen Wencun¹

¹ School of Foreign Languages & Applied Linguistics, China West Normal University, Nanchong, China

Correspondence: Shen Lifan, School of English Teaching Methodology, China West Normal University, Nanchong, 637000, China. Tel: 86-138-9071-8577. E-mail: nanqu105@hotmail.com

Received: September 17, 2012 Accepted: February 25, 2013 Online Published: March 25, 2013

doi:10.5539/ijel.v3n2p115

URL: <http://dx.doi.org/10.5539/ijel.v3n2p115>

Abstract

This study is a corpus-based lexical study that aims to compare the use of *research* as a noun between native speakers and Chinese EAP learners in research articles in Linguistics. A self-built learner corpus of academic English (CMFD) and its parallel corpus (PQDT) are applied. Quantitative analysis of frequency and qualitative analysis of collocation of node words are used in this paper. The results reveal Chinese EAP learners use *research* more frequently than native speakers, and native speakers never use “researches” as a plural form of noun in academic writing while Chinese EAP learners use this form frequently. Compared with native speakers, Chinese learners tend to make the following errors: an overuse of *research*; using *research* as a countable noun; disorder in using of “research” and “researches”; confusedness of “much research” expressions; mixed collocation prosodies. The knowledge gained by this study can increase awareness of proper use of *research* in composition of instructors and L2 writers, leading to clearer, more accurate texts.

Keywords: corpus linguistics, English for academic purposes, AWL, collocation, research

1. Introduction

Academic vocabulary plays an important role in academic discourse. However, it is found most problematic by learners. One supervisor of the first author once pointed out postgraduates’ misuse of research (see Note 1) in their papers in her course twice. Research is indeed an important word in dissertations and theses. Moreover research is one of the most common words in the Academic Word List (see Note 2). (Research is found in sublist 1. Sublist 1 contains the most common words in the AWL. Sublist 2 contains the next most common words, and so on. There are 10 sublists totally). The noun of research is never used as a countable noun in articles written by English native speakers. However, according to the present authors’ questionnaire (see Appendix B), more than half learners of Advanced English for Academic Purposes (EAP) use “researches” in sentences where native speakers use “research”.

Based on the above phenomenon, this paper compares the usage of the word research between native speakers and Chinese learners in academic prose. It tries to find out the concrete differences on the use of research between the two. Firstly, by using both quantitative and qualitative analytic procedures to examine the frequencies and collocates of “research” and “researches” as nouns, the results gained in this study can increase awareness of proper use of research in composition of instructors and L2 writers, leading to clearer, more accurate texts. Secondly, our purpose of finding differences on the use of academic word research between native speakers and Chinese learners is to raise the awareness of learning and teaching academic vocabulary. Some scholars have stated that courses involving direct attention to language features were found to lead to better learning than courses that only focuses on incidental learning (Ellis, 1990; Long, 1988). Thus, we believe that the direct learning and teaching of the frequently-used AWL words can help students in their development of academic reading and writing abilities.

2. Literature Review

Recent years have seen the growing genre of English for Academic Purpose (EAP). According to Hyland (2002), “English for Academic Purposes refers to language research and instruction that focuses on the specific communicative needs and practices of particular groups in academic contexts. It means grounding instruction in an understanding of the cognitive, social and linguistic demands of specific academic disciplines.” This

expanding role for EAP has been accompanied by research on EAP both broad and at home. Flowerdew (2000) did research into the English language behaviors and patterns of nonnative academics in 2000. At the same time, Hyland (2000) studied the ideological impact of expert discourses, the social distribution of valued literacies, the access non-native and novice members have to prestigious genres, and found that the ways controlling specialized discourses are related to status and credibility. Recently, collocation and corpus analysis in academic writing have also attracted interest. Collocation plays an important role in lexical cohesion. Hoey (2007) argues that exposure to collocations primes or prepares us to recall their correct meaning, and use them correctly whenever we re-encounter them. And “language obtained through corpora has the advantage of being authentic and reveals uses that native speakers do not think of” (cited from BETTY LANTEIGNE & PETER CROMPTON, 2011). In addition, corpus is having a beneficial effect on contrastive studies (Connor & Moreno, 2005). Wu Jin (2011) conducted an analysis, from the point of collocation, between a self-built learner corpus of academic English and its reference corpus to investigate the depth of Chinese postgraduate students’ academic vocabulary knowledge. Viphavee Vongpumivitch, et al (2009) did a corpus-based lexical study, also from the point of collocation, to explore the frequency of the AWL words that are used in the field of applied linguistics. These two investigations are valuable as they pay attention to collocations and corpus analysis. Yet, their research in academic writing is rarely on specific words. Although Bethany Gray & Viviana Cortes (2011) did research on the pronoun in academic writing, that is “this” and “these”; at home, Zhang Xiurong & Li Zengshun (2011) examined the frequencies and discourse functions of first person pronouns (we, our, us) in research articles from a corpus-based perspective; Sun Fang & Chen Jiansheng (2011) studied the use of “however” and “therefore” in terms of their frequencies and positions in economical research articles, these words are out of Academic Word List. That is, there is a gap in the research of specific word of AWL in terms of KW’s frequency and collocation within corpora to date. Therefore, the present study tries to fill in this gap by studying the frequency and collocation of the word *research* as a noun in EAP Corpus, both of native speakers and Chinese learners.

3. Research Questions

- 1). In terms of the frequency and collocation, what are the differences between Chinese EAP learners and native speakers in using the word *research* in academic writing?
- 2). What types of errors in detail do Chinese EAP learners tend to make in using the word *research* in academic prose?

4. Method

4.1 Corpus

Michael Stubbs (2007) maintains “a corpus allows us to get the facts right, a mass examples and document things thoroughly, and document types of facts (e.g. about frequency and typicality) which are not open to introspection and which are not well described in current dictionaries and grammars”. And in contrastive studies, building comparable corpora is important. According to Connor & Moreno (2005), “Applying appropriate tertia compactionis at the design and analysis stages of contrastive research will help us build comparable corpora that can provide baseline data for meaningful cultural comparisons.” In the current study, two corpora were built. One is a sub-corpus of academic papers from China Master’s Theses Full-text Database (CMFD), which consists of theses from 10 academic disciplines and 168 special topic databases. In CMFD corpus, Linguistics discipline is chosen as the focus of the present analysis because most theses in this domain are written in English. The other is the Parallel Corpus, L1 English sub-corpus of PQDT (Master’s theses from ProQuest). PQDT is the only full-text database in China providing high quality dissertations and theses. The scope covers extensive aspects, and most dissertations and theses come from over 2000 American and European universities. In the process of corpora building, all texts are randomly selected from CMFD and PQDT, and text samples are equivalence in time, discipline, number, length and level (master). The corpora referred in this paper are described in Table 1 & Table 2.

Table 1. The corpora applied in this paper

Corpus	Time	Discipline	Number of texts	Approx. number of words
CMFD	2001-2011	Linguistics	17	305583
PQDT	2001-2011	Linguistics	17	321212

Table 2. The length of texts in corpora

Word tokens	Min	Max	Mean
CMFD	10624	43825	17975
PQDT	9333	53542	18894

4.2 Instrument and Procedures

The present study employs Antconc 3.3 as the retrieval program, and its two tools are used, that is, concordance and collocates. Using the concordance tool of Antconc 3.3, all instances of “research(es)” were located in the two corpora. All occurrences of “research(es)” were coded as nouns. Instance of “researches” that was not used as a noun, one example extracted from PQDT used as singular form of verb, was excluded from analysis. Frequencies were calculated for the total number of occurrences of “research(es)”. Besides, Chi-square Calculator is used to test the significance.

Meanwhile, semantic prosodies are used for analysis of the collocation in both CMFD and PQDT. Firth (1957) claims that some words habitually collocate with other words. According to Michael Stubbs (2007), “words may habitually collocate with other words from a definable semantic set”, “words have distinctive semantic profiles or prosodies”. And some scholars consider semantic prosody as a further level of abstraction of the relationship between lexical units (Sinclair, 1996 & 1998; Stubbs, 2001). Generally, four kinds of prosodies are used to analyze the collocation of node words with a certain span in corpora, that is, positive prosody, negative prosody, neutral prosody and mixed prosodies (Michael Stubbs, 1996). According to Partington (2004), semantic prosody falls into favourable, neutral and unfavourable prosodies. In this study, a pleasant or favourable affective meaning was labelled as positive while an unpleasant or unfavourable affective meaning was judged as negative. When what was happening was completely neutral, or the context provided no evidence of any semantic prosody, the instance was labelled as neutral. In addition, Michael Stubbs (2007) also points out: “the strength of association between words can be measured in quantitative terms.” There are many statistical tests used to measure collocational strength, e.g. the MI, z, t, log-likelihood scores. In this paper, MI-Score is applied as the role of “quantitative term” to measure “the strength of association between words”.

Additionally, SPSS is employed to offer a descriptive statistics report of “research(es)”, which aims to test any difference between Chinese EAP learner and native speakers.

5. Results

5.1 Frequencies of “Research(es)”

Table 3. Frequencies of “research(es)” in corpora

	CMFD	PQDT
Freq of research	455	365
Freq of researches	96	0
Size of corpus	305583	321212

Table 4. Log-likelihood test: frequencies of “research(es)” in corpora

KW	Chi-square	Critical value	P
research	14.86581681	6.634896601	0.01

Note. Chi-square is larger than Critical value, $p < .05$.

Table 5. Descriptive report of “research(es)” in corpora

Corpus		Research	Researches
CMFD	Mean	26.7647	8.00
	N	17	12
	Std. Deviation	15.63861	7.746
PQDT	Mean	25.6875	
	N	16	
	Std. Deviation	29.06021	

With the concordance tool, KW's (key word) frequency can be obtained. As can be seen in Table 3, “research(es)” frequencies are 455 & 96 in CMFD and 365 & 0 in PQDT. And with the frequency and corpus size filled in the table of Chi-square Calculator, significance can be tested. According to Table 4, “research” Chi-square is larger than the Critical value, and P is less than 0.01, which shows the difference is significant. In the case of “researches”, as the frequency in PQDT is 0 while the frequency in CMFD is 96, the significant difference can be easily observed. Besides, according to Table 5, although the means of “research(es)” in both corpora are near to equal, Std. Deviation (SD) of “research(es)” in PQDT is larger than that in CMFD. That is, compared with a great disparities in frequency among native speakers, there are few differences of “research(es)” frequency among Chinese EAP learners.

5.2 Collocation

Table 6. Collocates (L1) of “research(es)” in CMFD

CMFD			
research		researches	
Collocates	Stat	Collocates	Stat
Vygotskys	9.15781	extensive	9.0413
foregoing	7.83588	constructive	8.77826
theorists	7.15781	insightful	8.36322
surveyed	7.15781	earliest	8.36322
empirical	6.98788	monolingual	7.90379
present	6.24761	lexicon	7.88518
doing	6.13227	empirical	7.45633
methodology	5.90988	conduct	7.27576
future	5.87577	recent	7.1933
qualitative	5.83588	previous	6.97091
extensive	5.83588	published	6.60834
quantitative	5.72485	relevant	6.55587
further	5.47599	bilingual	6.43249
current	5.21028	future	6.20672
considerable	4.98788	enough	5.74851
blank	4.98788	psychological	5.6908
program	4.83588	further	5.6814
bilingual	4.81203	quantitative	5.60834
my	4.55493	doing	5.53033
SLA	4.49484	these	5.3536

academic	4.40292	above-mentioned	5.1734
relevant	4.35045	many	5.10584
educational	4.35045	a few	4.92028
little	4.29983	those	4.90927
phonetic	4.15781	some	4.49639
descriptive	4.15781	related	4.0145
previous	4.02853	following	3.62176
concrete	3.98788	such	3.25994
politeness	3.90988		
anxiety	3.76549		
did	3.63425		
primary	3.57285		
states	3.48538		
recent	3.40292		
related	3.39404		
this	3.3658		
our	3.32492		
linguistic	3.30606		
motivation	3.18053		
theoretical	3.02853		

Note. $f(n,c)$ (see Note 3) is greater than or equal to 1; MI is greater than 3.

Table 7. Collocates (L1) of “research” in PQDT

PQDT			
research		research	
Collocates	Stat	Collocates	Stat
time-consuming	9.54833	SLA	5.54833
seminal	7.96337	promote	5.54833
cursory	7.96337	my	5.44679
conducting	7.96337	extensive	5.37841
published	7.70034	entrepreneurial	5.2264
off-campus	7.54833	continues	5.15602
pursue	7.2264	primary	4.93362
analytical	7.2264	academic	4.92384
previous	7.10539	healthcare	4.90448
future	6.97198	varied	4.84789
diachronic	6.96337	market	4.64144
intonational	6.74098	little	4.48224
substantial	6.54833	related	4.11093
ethnographic	6.46087	this	4.07372
quantitative	6.42495	investigate	4.05648

conduct	6.37841	questionnaire	4.02477
further	6.31387	motivational	3.91898
medical	6.2264	initial	3.90448
motivation	6.12431	applied	3.71544
existing	6.0889	carol	3.32916
anxiety	6.0889	interesting	3.3004
qualitative	5.91606	current	3.05648
scientific	5.64144	present	3.01955

Note. $f(n,c)$ (see Note 2.) is greater than or equal to 1; MI is greater than 3.

Table 8. Collocates of quantitative modifier in corpora

Quantitative Modifier	PQDT	CMFD	
	research	research	researches
little	24, 25	8, 6, 10	
extensive	21	9	4, 2
considerable		10	
enough			8
many			5, 7(3)
a few			5
some			5(2), 6, 2(4)

Note. No. 1 - 17 is text ID in CMFD; No. 18 - 34 is text ID in PQDT.

5(2)= two instances of "some researches" in Text 5.

Table 9. Collocates of discipline modifier in corpora

Discipline Modifier	PQDT	CMFD	
	research	research	researches
ethnographic	34(2)		
medical	28		
bilingual		5(2)	5(2)
SLA		12, 1(2)	
phonetic		4	
linguistic		12, 11(3)	
monolingual			5
lexicon			5
psychological			13

The researchers use the collocates tool to retrieve KW's collocation. And the collocates were chosen based on the following rules: 1). The first left modifier of research was chosen to study for the aim of validity and easy processing. 2). MI-Score of collocation is greater than 3 for the statistics meaning. 3). The minimum co-occurrence frequency was set at 1. Collocation (L1) which occurs in CMFD and PQDT as a modifier of the noun "research(es)" includes the words in Table 6 & Table 7.

As shown in Table 6 & Table 7, the collocates of "research(es)" can be mainly classified into the following categories:

- Chronological modifier, such as *recent, previous, future*, etc.;
- Degree modifier, such as *further*;
- Descriptive modifier, such as *empirical, quantitative, qualitative*, etc.;
- Discipline modifier, such as *monolingual, psychological, ethnographic*, etc.;
- Pronoun modifier, such as *my, this, our*, etc.;
- Quantitative modifier, such as *extensive, some, little, enough*, etc.;
- Predicate verb, such as *conduct, do*, etc.;
- Property modifier, such as *anxiety, motivation, questionnaire*, etc..

Clearly, there are some differences between Chinese EAP learners and native speakers on these categories of collocates. First, Chinese EAP learners use more often quantitative modifier and disciplinary modifier to collocate with “research(es)” than native speakers. This excess is embodied not only in the total counts in all texts but also in individual text (Table 8 & 9). Second, Chinese EAP learners are more willing to use “researches” than “research” to denote the meaning of numerous research while native speakers use “research”. Third, as for predicate verb, “doing research(es)” is typical in CMFD, while “conducting research” is more traditional in PQDT. Fourth, Chinese EAP learners use more often *relevant* than *related* to collocate with “research(es)” while native speakers only use *related*. Fifth, native speakers say “previous research” while Chinese EAP learners prefer to “previous researches”. Last, there is a different prosody property of *research* between the two. Native speakers regard *research* as a word with non-positive prosody while Chinese EAP learners regard *research(es)* as a word with mixed (positive & neutral & negative) prosodies.

6. Discussion

The part of results in this study can answer the original research questions:

1). *In terms of the frequency and collocation, what are the differences between Chinese EAP learners and native speakers in using the word research in academic writing?*

As can be seen in Table 3, *research(es)* frequencies are different between Chinese EAP learners and native speakers, and Table 4 shows this difference is significant (“research”: $\chi^2 = 14.86581681 > 6.634896601$; $P < .05$). This difference shows Chinese EAP learners use “research” more often than native speakers. In the case of “researches”, native speakers never use it while Chinese EAP learners use it frequently in academic prose. Meanwhile, as can be seen in Table 5, the mean of “research(es)” in each corpus is mostly equal while the SD of “research(es)” in PQDT is larger than that in CMFD. It indicates there are great disparities in frequency among native speakers while Chinese EAP learners have much common in “research(es)” frequency. Besides, with the tool of keyword list of AntConc, *research* is retrieved as a negative word in highlight color. Seemingly, it implies *research* occupies a higher key-ness rank in Chinese EAP learners’ minds.

In the case of collocation, as shown in the previous section, there are also differences in several aspects. First, Chinese EAP learners use more often quantitative modifier and disciplinary modifier to collocate with “research(es)” than native speakers. This result can be observed in Table 10 in two aspects. For one thing, the total counts of quantitative modifier, disciplinary modifier in CMFD are more than those in PQDT. For another thing, the texts contained quantitative modifier and disciplinary modifier in CMFD are more than those in PQDT.

Table 10. Distribution of QM & DM in corpora

Number	CMFD		PQDT	
	QM	DM	QM	DM
Collocate	7	7	2	2
Text	8	6	3	2
Total	25	15	3	3

QM: quantitative modifier

DM: disciplinary modifier

Second, Chinese EAP learners are more willing to use “researches” than “research” to denote the meaning of much research while native speakers use “research”. As can be seen in Table 8, Chinese EAP learners tend to use collocations of “some researches”, “a few researches”, “extensive researches”, “many researches”, “enough researches”, while native speakers use collocations of “little research”, “extensive research”. Besides, as for predicate verb, “doing research(es)” is typical in CMFD, while “conducting research” is more traditional in PQDT. In addition, descriptive modifiers, *relevant* and *related*, are both in the list of collocates in CMFD while only *related* is in PQDT, and *relevant*'s Stats (4.35045 & 6.55587) are higher than those of *related* (3.39404 & 4.0145) in CMFD. That is, Chinese EAP learners use more often *relevant* than *related* to collocate with “research(es)” while native speakers only use *related*. As for the chronological modifier *previous*, native speakers say “previous research” while Chinese EAP learners prefer “previous researches” (Stat: 6.97091) to “previous research” (Stat: 4.02853). Last, from the point of prosody, according to Table 7, most of collocates used by native speakers are neutral word, several negative word, but no positive word. It seems to indicate native speakers regard research as word with non-positive prosody. However, in Table 6, Chinese EAP learners use not only neutral and negative words but also positive words, *insightful*, *constructive*, for instance, to collocate with “researches”.

2). *What types of errors in detail do Chinese EAP learners tend to make in using the word research in academic prose?*

Based on the above contrastive analysis, we can see some errors in *research* among Chinese EAP learners. First, there is an overuse of *research* among Chinese EAP learners in academic prose. Chinese EAP learners use “research” more often than native speakers. As for the form of “researches”, native speakers never use it while Chinese EAP learners use it frequently in academic prose (see Table 3 & 4 & 5). The first potential reason is native speakers use alternately “study” & “studies” more often than Chinese EAP learners in their academic prose (935 hits to 864 hits). The second possible reason is that there are different perceptions in *research* between the two, which in turn causes another error. That is, Chinese EAP learners tend to use *research* as a countable noun. In fact, native speakers use *research* as an uncountable noun and it is unconventional for them to use “researches” as the plural form in academic writing. This difference can be found in Table 6 & Table 7, *little* is used to modify “research” by native speakers whereas *many*, *a few*, *these*, *those* to “researches” by Chinese EAP learners. Here are the examples:

- (1) “*It provides convenience for conducting many researches.*” (CL in CMFD)
- (2) “*Similar responses from different subjects were found by a few researches (e.g., Jenkins 1970, cited in Aitchison 1987; Kent & Rossanoff 1910, cited in Jay 2004).*” (CL in CMFD)
- (3) “*Those researches have made great achievements and set the norms of the use of genitive, among which the researches made by A Comprehensive Grammar of the English Language (CGEL) and Longman Grammar of Spoken and Written English (LGSWE) are distinguished.*” (CL in CMFD)
- (4) “*However, the validity of these researches, which investigated the issue through general analysis with no particular case involved and a lack of data, had somewhat been affected.*” (CL in CMFD)
- (5) “*However, there is little research on the aetiology, course, prognosis or treatment of post schizophrenic depression.*” (NS in PQDT)

Besides, there is a disorder among some learners on the use of the word *research*. Specifically, *research* was used as a countable and uncountable noun alike. This can be seen in Table 8. In text 6, there is a co-occurrence of the collocations of “little research” and “some researches”. Similarly, “little research” and “enough researches” co-occur in text 8. This phenomenon signifies there is a chaotic state on the word *research* in some Chinese EAP learners' minds.

Additionally, due to a wrong perception on *research* about its number, an error takes place when Chinese EAP learners express the meaning of “much research”. As can be seen in Table 8, Chinese EAP learners tend to use collocations of “some researches”, “a few researches”, “extensive researches”, “many researches”, “enough researches”, while native speakers use collocations of “little research”, “extensive research”.

Last, there is an error of collocation prosody among Chinese EAP learners. Concretely, native speakers regard *research* as a word with non-positive prosody but Chinese EAP learners regard *research* as a word with mixed (positive & neutral & negative) prosodies. Most of collocates used by native speakers are neutral word, several negative word, but no positive word (See Table 7). However, in Table 6, Chinese EAP learners use not only neutral and negative words but also positive words, *insightful*, *constructive*, for instance, to collocate with *researches*. Here are the text examples:

(6) “Due to the complexity of genitive structures, there have been a great number of *insightful researches* into it conducted by foreign linguists from the perspective of semantics, syntax and corpus linguistics.” (CL in CMFD)

(7) “Linguists have conducted *constructive researches* in this field from different perspectives and have made great achievements.” (CL in CMFD)

According to Agustin Llach, as L2 learners become more proficient and as they face cognitively challenging writing tasks, lexical errors do not disappear; instead, the types of errors change (Cited from CAROL SEVERINO, 2012). In this study, Chinese EAP learners tend to make the following errors: an overuse of *research*; using *research* as a countable noun; disorder on the use of “research” and “researches”; confusedness of “much research” expressions; mixed collocation prosodies. The research team members consider the following to be the possible sources of error.

1) Cross-linguistic influence. i) Intralingual transfer, for one thing, Chinese EAP learners have mixed plural form of noun with the verb form of the third person singular. In English, *research* can be used as both a noun and a verb, and “researches” as a verb form of the third person singular is commonly used by native speakers. This leads to that the verb singular morpheme of -es is mistaken for the noun plural morpheme of -es by Chinese EAP learners. For another, there is a word similar to *research* in term of word form, *search*, which is a countable noun and has its plural form of “searches”. This also can cause a potential negative transfer among Chinese EAP learners. ii) Interlingual transfer, in Chinese, it is common to say “一项研究, 两项研究, 几项研究”, which is different from “a research, several research” in English. From the point of Chinese EAP learners, “一项研究, 两项研究, 几项研究” is unmarked, and “a research, several research” is marked. This case is “where the native language shows an unmarked setting and the target language a marked one”, which is “the most obvious case of transfer” (Rod Ellis, 1999). Because the setting of parameter is idiosyncratic (i.e. marked), Chinese EAP learners fall back on their L1 knowledge of “研究” in the process of learning *research*.

2) Insufficient input. Gass & Selinker (2008) argue that “input of some sort is necessary in order for acquisition to take place” and “there are three sources of input: (a) teacher, (b) materials, and (c) other learners”. Chinese EAP learners failure in acquisition of *research* seems to imply they have a insufficient exposure to this word through any source of input. Or, either the quantity or quality of input is not enough, which causes the error in *research* acquisition.

3) Lack of Awareness. Noticing hypothesis was proposed by Schmidt. Underlying the hypothesis is the idea of noticing a gap. Schmidt and Frota (1986) suggested that “a second language learner will begin to acquire the target like form if and only if it is present in comprehended input and ‘noticed’ in the normal sense of the word, that is consciously”. It highlights the role of attention, which is as important as input in the process of SLA. At one hand, attention has a diminished effect for proficiency. That is, Chinese EAP learners are more likely to pay attention to a specific word in early stages of learning. On the other hand, Chinese EAP learners are lack of register awareness, awareness of academic writing. Both aspects result in the incorrect use of *research* among Chinese EAP learners in academic writing.

7. Conclusions

A limitation of this study is the size of the sample used for the analysis. The scope of study is limited to linguistic field and the samples are insufficient in quantity. Accordingly, there is a 0 hit of “researches” in PQDT and it is not favorable to test its significance with Chi-square Calculator. This limitation also results in not a significant collocation of *research* in CMFD and PQDT. That is to say some features of collocation of *research* analyzed in this paper are not surely generalized to all Chinese EAP learners. Thus, future researchers may want to expand the size of their corpus to be as large as possible in order to increase the generalizability of their findings and to see if their results would be similar to ours.

The goal of this study is to explore the frequency and collocation of headword of the Academic Word List *research* noun in published academic research articles of both native speakers and Chinese EAP learners. The analysis shows that Chinese EAP learners use “research” more frequently than native speakers, and native speakers never use “researches” as a plural form of noun in academic writing while Chinese EAP learners use this form frequently. Compared with native speakers, Chinese learners tend to make the following errors: an overuse of *research*; using *research* as a countable noun; disorder on the use of “research” and “researches”; confusedness of “numerous research” expressions; mixed collocation prosodies. The potential causes are cross-linguistic influence, insufficient input and lack of awareness.

The findings of this paper can raise awareness of the proper use of *research* for writing instructors and students.

Increased awareness may in turn lead to a more conscious effort to think about language use in order to create clearer and more accurate texts for readers. In addition, increased awareness of the proper use of *research* may promote the development of reading skills for student writers by helping them to efficiently and accurately comprehend of native speakers articles. Last, our results can raise the awareness of learning and teaching academic vocabulary. We believe that the direct learning and teaching of the frequently-used AWL words can help students in their development of academic reading and writing abilities.

Acknowledgments

We would like to thank the editor, Diana Xu, and two anonymous reviewers for their timely feedback on drafts of this manuscript, and their constant encouragement and help. At every stage in the writing process their wisdom and insight have been an inspiration.

We would also like to thank Prof. Gu Tongqing of China West Normal University for her help and advice on improving the method in this article. We remain solely responsible for the remaining errors in this article.

References

- Academic wordlist. Retrieved from <http://www.victoria.ac.nz/lals/resources/academicwordlist>
- Connor, U. M., & Moreno, A. I. (2005). Tertium Comparationis: A vital component in contrastive research methodology. In P. Bruthiaux, D. Atkinson, W. G. Eggington, W. Grabe, & V. Ramanathan (Eds.), *Directions in Applied Linguistics: Essays in Honor of Robert B. Kaplan*. Clevedon (pp. 153-164). England: Multilingual Matters.
- Ellis, R. (1990). *Instructed Second Language Acquisition*. London: Blackwell.
- Ellis, R. (1999). *Understanding Second Language Acquisition* (p. 206). Shanghai: Shanghai Foreign Language Education Press. <http://dx.doi.org/10.1017/S0272263199002077>
- Firth, J. R. (1957). A synopsis of linguistic theory, 1930-1955. *Studies in Linguistic Analysis, Special Volume*, Philological Society, 1-32. Oxford: Blackwell.
- Flowerdew, J. (2000). Discourse community, legitimate peripheral participation, and the nonnative-English-speaking scholar. *TESOL Quarterly*, 24(1), 127-150. <http://dx.doi.org/10.2307/3588099>
- Gray, B., & Cortes, V. (2011). Perception vs. evidence: An analysis of this and these in academic prose. *English for Specific Purposes*, 30, 31-43. <http://dx.doi.org/10.1016/j.esp.2010.06.004>
- Hoey, M. (2007). Lexical Priming and Literary Creativity. In M. Hoey, et al. (Eds.), *Text, Discourse and Corpora: Theory and Analysis* (pp. 7-29).
- Hyland, K. (2000). *Disciplinary discourses: Social interactions in academic writing*. London: Longman.
- Hyland, K., & Hamp-Lyons, L. (2002). EAP: Issues and Directions. *Journal of English for Academic Purposes*, 1(1), 2-3. [http://dx.doi.org/10.1016/S1475-1585\(02\)00002-4](http://dx.doi.org/10.1016/S1475-1585(02)00002-4)
- Lanteigne, B., & Crompton, P. (2011). Analyzing Use of Thanks to You: Insights for Language Teaching and Assessment in Second and Foreign Language Contexts. *Research in Language*, 9(2). <http://dx.doi.org/10.2478/v10015-011-0018-9>
- Long, M. (1988). Instructed interlanguage development. In L. Beebe (Ed.), *Issues in Second Language Acquisition*. New York: Newbury House.
- Partington, A. (2004). "Utterly content in each other's company": Semantic prosody and semantic preference. *International Journal of Corpus Linguistics*, 9(1), 131-156. <http://dx.doi.org/10.1075/ijcl.9.1.07par>
- Schmidt, R., & Frota, S. (1986). Developing basic conversational ability in a second language: A case study of an adult learner of Portuguese. In R. Day (Ed.), *Talking to Learn: Conversation in Second Language Acquisition* (p. 311). Rowley, MA: Newbury House.
- Severino, C. (2012). Lexical Errors and Accuracy in Foreign Language Writing by AgustÍN Llach, MarÍA Pilar. *The Modern Language Journal*, 96(3). http://dx.doi.org/10.1111/j.1540-4781.2012.01367_2.x
- Sinclair, J. (1996). The search for units of meaning. *Textus*, IX, 75-106.
- Sinclair, J. (1998). The lexical item. In E. Weigand (Ed.), *Contrastive Lexical Semantics* (pp. 1-24). Amsterdam: John Benjamins.
- Stubbs, M. (1996). *Text and Corpus Analysis: Computer Assisted Studies of Language and Culture*. Wiley-Blackwell Publishers.

- Stubbs, M. (2001). *Words and Phrases*. Oxford: Blackwell.
- Stubbs, M. (2007). Collocations and Semantic Profiles: On the Cause of the Trouble with Quantitative Studies. In W. Teubert, & R. Krishnamurthy (Eds.), *Corpus Linguistics: Critical Concepts in Linguistics*. London & New York: Routledge.
- Sun, F., & Chen, J. (2011). A Study on the Use of “However” and “Therefore” in Chinese Scholars’ Academic Writings. *Journal of Luoyang Institute of Science and Technology (Social Science)*, 26(6).
- Susan, M. G., & Selinke, L. (2008). *Second Language Acquisition: An introductory course* (3rd ed., pp. 369-480). Taylor & Francis.
- Vongpumivitch, V., Huang, J., & Chang, Y. (2009). Frequency analysis of the words in the Academic Word List (AWL) and non-AWL content words in applied linguistics research papers. *English for Specific Purposes*, 28, 33-1. <http://dx.doi.org/10.1016/j.esp.2008.08.003>
- Wu Jin. (2011). 中国研究生产出性学术词汇知识深度的语料库研究 (A corpus-based study on the depth of Chinese postgraduate students’ academic vocabulary knowledge). *Foreign Language Education*, 32(2).
- Zhang, X., & Li, Z. (2011). An Empirical Research on Frequency and Discourse Functions of First Person Pronouns in Research Articles. *Journal of Xi’an International Studies University*, 2011(2).

Appendices

Appendix A. Sublists of the Academic Word List

Each word in italics is the most frequently occurring member of the word family in the Academic Corpus. For example, *analysis* is the most common form of the word family *analyse*. British and American spelling is included in the word families, so *contextualise* and *contextualize* are both included in the family *context*.

Sublist 1 contains the most common words in the AWL. Sublist 2 contains the next most common words, and so on. There are 60 families in each sublist, except for sublist 10 which has 30.

Sublist 1 of the Academic Word List

...

require	responsiveness
<i>required</i>	unresponsive
requirement	<i>role</i>
requirements	roles
requires	<i>section</i>
requiring	sectioned
research	sectioning
researched	sections
researcher	<i>sector</i>
researchers	sectors
researches	<i>significant</i>
researching	insignificant
respond	insignificantly
responded	significance
respondent	significantly
respondents	signified
responding	signifies
responds	signify
<i>response</i>	signifying
responses	<i>similar</i>
responsive	dissimilar

similarities	unstructured
similarity	<i>theory</i>
similarly	theoretical
<i>source</i>	theoretically
sourced	theories
sources	theorist
sourcing	theorists
<i>specific</i>	vary
specifically	invariable
specification	invariably
specifications	variability
specificity	variable
specifics	<i>variables</i>
<i>structure</i>	variably
restructure	variance
restructured	variant
restructures	variants
restructuring	variation
structural	variations
structurally	varied
structured	varies
structures	varying
structuring	

...

Appendix B. Questionnaire

A Questionnaire about Perception and Usage of *Research*

The following questionnaire is designed for research on different perceptions and usages of *research* (noun) between English native speakers and Chinese learners. Please answer each question honestly and frankly according to your own opinion. There are no “correct” answers. All the data collected will be highly confidential and will be used for the research only.

serial number: _____ (by investigator)

Sex: _____

Major: _____

Grade: _____

Age: _____

The following questions are based on “research” in academic prose, please choose one appropriate answer according to your own opinion.

Part one: Short answer questions.

1. Do you think “research” is a countable noun ____?
 - A. Yes (Skip down to Question 2)
 - B. No (Skip down to the next page)
 - C. I have no idea (Skip down to the next page)
2. What’s the plural form of “research” ____?
 - A. researches (Skip down to 2a)
 - B. research (Skip down to 2b)
- 2a. How frequently do you think “researches” is used? If percentage is used to describe its frequency, which one would you like to choose ____? (Skip down to the next page)
 - A. more than 50%
 - B. 50%-30%
 - C. 30%-10%
 - D. less than 10%
- 2b. How frequently do you think “research” is used? If percentage is used to describe its frequency, which one would you like to choose ____? (Turn to the next page)
 - A. more than 50%
 - B. 50%-30%
 - C. 30%-10%
 - D. less than 10%

Part two, you are presented in writing situation and you are required to choose one appropriate sentence from A, B or C according to the style. Mark your choice by writing the corresponding letter in the brackets.

1. A. **Research shows** that heroes are not particularly achievement-oriented or driven by the need for approval.
 B. **Research show** that heroes are not particularly achievement-oriented or driven by the need for approval.
 C. **Researches show** that heroes are not particularly achievement-oriented or driven by the need for approval.
 1 ()
2. A. **Some research shows** broccoli may even ease headaches.
 B. **Some research show** broccoli may even ease headaches.
 C. **Some researches show** broccoli may even ease headaches.
 2 ()
3. A. This study, together with **his scholarly research** into the Welsh and other Gaelic languages, formed his life’s work.
 B. This study, together with **his scholarly researches** into the Welsh and other Gaelic languages, formed his life’s work.
 3 ()
4. A. **More researches are** needed to understand addiction pattern in non-daily smokers.
 B. **More research are** needed to understand addiction pattern in non-daily smokers.
 C. **More research is** needed to understand addiction pattern in non-daily smokers.
 4 ()
5. A. **Recent research** on the swimming speeds of fish **shows** that they soon get tired, hence the success of the trawl.
 B. **Recent research** on the swimming speeds of fish **show** that they soon get tired, hence the success of the trawl.
 C. **Recent researches** on the swimming speeds of fish **show** that they soon get tired, hence the success of the trawl. (Please turn to the next page)
 5 ()
6. A. The broad conclusions of that pioneering work remain undisturbed, but **subsequent research has** expanded and somewhat altered their empirical support.
 B. The broad conclusions of that pioneering work remain undisturbed, but **subsequent research have**

expanded and somewhat altered their empirical support.

- C. The broad conclusions of that pioneering work remain undisturbed, but *subsequent researches have* expanded and somewhat altered their empirical support.

6 ()

7. A. *Most research shows* that simply exercising, without changing your diet, doesn't lead to losing weight.

B. *Most research show* that simply exercising, without changing your diet, doesn't lead to losing weight.

C. *Most researches show* that simply exercising, without changing your diet, doesn't lead to losing weight.

7 ()

8. A. Kays and London give the results of *extensive researches* and experiments particularly related to compact forms of heat exchanger.

B. Kays and London give the results of *extensive research* and experiments particularly related to compact forms of heat exchanger.

8 ()

9. A. Market and prospects for baby foods have come under the scrutiny, which is responsible for *so much contemporary research* in many fields.

B. Market and prospects for baby foods have come under the scrutiny, which is responsible for *so many contemporary research* in many fields.

C. Market and prospects for baby foods have come under the scrutiny, which is responsible for *so many contemporary researches* in many fields.

9 ()

This is the end of the questionnaire and thanks for your cooperation, good day!

Notes

Note 1. The noun *research* is studied in this paper. Research in italic refers to the lemma of the word, including its two variations: "research" and "researches".

Note 2. The Academic Word List (AWL) was developed by Averil Coxhead as her MA thesis at the School of Linguistics and Applied Language Studies at Victoria University of Wellington, New Zealand. The list contains 570 word families which were selected according to principles. The AWL was primarily made so that it could be used by teachers as part of a programme preparing learners for tertiary level study or used by students working alone to learn the words most needed to study at tertiary institutions.

Note 3. $f(n,c)=f(\text{node, collocate})$: joint frequency of node and collocate.