The Effects of Lexical Input in Second Language Writing: A Corpus-Informed Approach

Huang Zeping
The Chinese University of Hong Kong

Introduction

A lack of vocabulary knowledge has been considered the main difficulty for writing in a second language (see Leki & Carson, 1994). To enhance writing performance, a corpus approach has been regarded as a viable alternative for helping learners with their lexico-grammatical patterns (Coxhead & Byrd, 2007; Flowerdew, 2010) and organizational patterns (Tribble, 2001). Tribble and Jones (1997) discussed two possibilities of incorporating corpora into L2 writing classrooms. First, teachers could examine a corpus to determine the most common words or patterns relative to a target genre, and write teaching materials based on the observed results. Second, students can be taught how to use a concordancer to explore the corpus themselves. This study was designed to investigate in what ways a corpus-informed approach has an impact on L2 writing and how L2 learners perceive the effects of corpus-informed materials on their English writing.

Literature Review

For more than two decades, the application of corpora has been regularly described as one of the most promising ideas in language teaching (see Johns, 1986; Sinclair & Carter, 2004) and a number of attempts have been made to apply corpora to L2 writing instruction. Thurstun and Candlin (1997), by utilizing a specialized corpus, designed a workbook for students' essay writing. The workbook provides both native and non-native English speaking university students with intensive exposure to some of the most important words in academic English. The workbook also introduces various rhetorical functions of academic essays—

such as stating the topic, referring to the literature, and drawing conclusions—and presents sets of concordance-based exercises on the most common vocabulary items used for carrying out these functions. The selection of vocabulary items is based on a specialized corpus of professional academic writing, an electronic collection of academic texts and papers from a range of disciplines, with a total word count of over one million words.

Grounded in discourse and genre analytic frameworks, Tribble (2002) outlined the ways in which appropriate corpus resources can be used to help learners develop competence as writers within specific academic domains. He demonstrated the use of keywords and frequency lists to identify lexico-grammatical features of the text in specific genres, and concluded that it is a plausible strategy for helping learners to understand text features in EAP and ESP writing.

Two more publications on applying corpus linguistics to the development of teaching material are that of Coxhead (2000) and Coxhead and Byrd (2007). Using a corpus linguistics approach, Coxhead (2000) generated an academic wordlist (AWL) of 570 headwords and 3000 words altogether. The idea of the AWL is to provide a shortcut for learners to expand their vocabulary by learning the most frequently-used lexis for academic writing. Coxhead and Byrd (2007) delineate a possible way, through integrating concordances with the keyword list, to prepare teachers to teach vocabulary and grammar for academic prose. In comparison to the traditional method of selecting and analyzing samples of academic prose, the authors argued that the web-based corpus approach, concordancing in particular, was more innovative and effective for helping teachers with materials development and providing them with information about academic language. They further stated that using such learning materials can benefit students to obtain the skill and knowledge needed to become effective learners of new words and their associated grammar.

Although Hyland (2003) states that the use of a corpus and a concordancer can offer "one of the most exciting applications of new technologies to L2 writing classes" (p. 167), two existing problems of corpus use in this area, in theory and in methodology, cannot be ignored. From a methodological point of view, Braun (2007) points out that most accessible corpora so far have been developed with linguistic research goals only, and they are not necessarily the resources with the most obvious pedagogical value. She stresses the values of a smaller and

genre-specific corpus, arguing that it can overcome some of the shortcomings of "mainstream" corpora, especially with regard to size and diversity of content. But realistically, these small corpora have not travelled well beyond the institutions in which they were created (Aston, 2004). From a theoretical point of view, Widdowson (2002) has criticized the use of corpora stating that while language learning is concerned with discourse and the use of language in concrete communicative situations, corpora are only a collection of texts, that is, products of language use isolated from any communicative situation.

When integrating corpora into language teaching, it seems to be more pertinent to adopt a corpus-informed rather than a corpus-based approach on the grounds that, in the former, learning takes on the advantages of the traditional teacher-student interactivity as well as the technological benefits brought by the pedagogical application of a corpus. The significance of a corpus-informed approach to language teaching is also reinforced by McCarthy (2008). He compares the difference between corpus-based and corpus-informed approaches: i.e., the corpus-based materials tend to be absolutely based on what we get from the corpus about language use, no matter whether they are useful or not. On the other hand, in a corpus-informed approach, teachers or material writers attempt to utilize corpus information in accordance with students' needs in order to filter it for pedagogical purposes.

The works cited above suggest the feasibility of developing writing materials using a corpus approach and the usefulness of corpus materials in helping students to acquire linguistic knowledge and improve the quality of their writing. However, this claim is mainly based on the writers' observations, not on empirical research. To this writer's knowledge, no research appears to have focused on the effects of corpus-informed materials on L2 writing by comparing the writing outcomes of a control and the experimental group. Hence, the need exists for empirical research to investigate whether corpus-informed materials help students improve their writing.

Research Questions

Two research questions are addressed in this study:

1. Can corpus-informed materials help L2 learners improve their overall writing quality?

2. Do L2 learners believe that corpus-informed materials help their writing?

Methodology

Corpus-informed Materials

First, a topic-specific corpus was compiled consisting of texts related to the topics of gambling and lottery. The texts were obtained from two sources: online authoritative English news websites and a small corpus named LOCNESS (Louvain Corpus of Native English Essays).

Following an innovative text-collecting approach suggested by Nelson (2009), three authoritative English news websites which contained quality articles on the desired topics were identified. They were BBC News, the Guardian and the New York Times. Once the websites were identified, a search for the keywords gambling and lottery on the websites was carried out and the relevant articles were downloaded.

The other source was from a sub-corpus of opinion essays on the topic of a national lottery written by British students and retrieved from LOCNESS. In this corpus, each text had approximately 500-600 words. Twelve essays, identified as samples of good writing by an experienced native English speaking teacher of writing, who was also an IELTS (International English Language Testing System) writing examiner, were selected for this study. Although it could be argued that essays written by native English speaking students may not be a suitable and reliable source for teaching English writing, it should also be noted that these revised texts can be deemed appropriate as they deal with the same subject field of the writing task. They are, in fact, quite close to the students' writing compared with the longer academic texts.

Next, a keyword list was generated from the topic-based corpus with the aid of a corpus tool called Wmatrix (Rayson, 2002). Target words were selected on the basis of two criteria: frequency of occurrence in the small, topic-based corpus (each word occurs at least three times), and abstract nouns often used in opinion essays (Read, 2004). According to the criteria, five words were chosen. They were controversy, criticism, objection, situation and effect. About ten concordances lines of each target word were selected and presented in the corpus-informed materials (see the appendix).

Participants

Forty third-year university students majoring in English for Business Purposes at a University in south China participated in this study. Their overall English proficiency level was upper intermediate according to the Oxford English Placement Test. The participants were randomly assigned to a control group and an experimental group, each group consisting of twenty students. A writing pretest conducted before the experiment showed there was no statistically significant difference in English writing competence between the two groups.

Procedure

The two groups were instructed to perform three writing exercises, i.e., a pre-test, an immediate posttest, and a delayed posttest (see Figure 1). Each writing test lasted 60 minutes. In the first week (Week 1), a pretest was taken by both groups. In the following week (Week 2), both groups took an immediate posttest, writing an opinion essay on the topic of lottery. Before the students took the immediate posttest, the experimental group was given a set of concordance exercises to learn the target words while the control group was only allowed to consult their dictionaries in order to learn the words. Two weeks later (Week 4), both groups took a delayed posttest, writing an opinion essay on gambling. Following the delayed posttest, questionnaires on the learners' evaluation of the corpus-informed concordance exercises were administered to the experimental group.

Figure 1. Flowchart of the Experiment
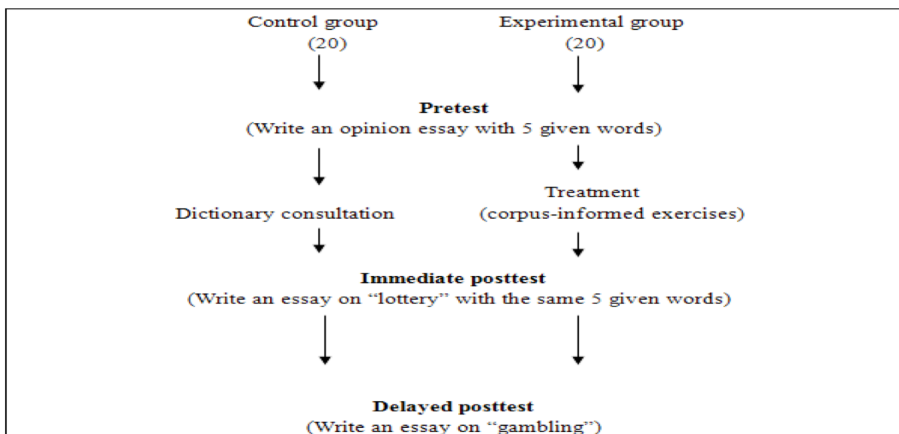
Data Analysis

Two sets of data were analyzed to investigate the effects of corpus-informed materials on learners' writing products: (1) student essays and (2) descriptive data obtained from students' questionnaire responses and students' learning journal entries. Three native English speaking teachers were invited to evaluate students' writing products. They had taught writing for more than five years and were also experienced IELTS examiners for the British Council. They marked the student essays according to the TWE (TOEFL Test of Written English Guide). The TWE test is holistically scored using a criterion-referenced scale ranging from 1 to 6. A score of 1 demonstrates incompetence in writing while a score of 6 shows clear competence in writing on both the rhetorical and syntactic levels. Each essay was scored twice, each time by a different rater. When the scoring differed by more than one point (e.g., one score of 3 and one of 5), the essays were sent to a third rater to resolve the discrepancy. A Pearson correlation coefficient was computed in order to maintain inter-rater reliability between the two sets of ratings (r=.823).

In addition to the holistic scores, the student essays were also textually analyzed to determine the extent to which the abstract nouns (controversy, criticism, objection, situation and effect) had been accurately used in the pretest, the immediate posttest, and the delayed posttest.

Finally, students were surveyed for their views on the corpus-informed approach in L2 writing. The responses to the Likert-scaled questionnaires in each category were summed and treated as interval data. The mean and standard deviation were calculated using SPSS, and its internal reliability was checked using Cronbach's Alpha. In order to enhance the presentation of the questionnaire data, students' responses were coded into three categories—"helpful," "not helpful," and "no opinion"—by placing all positive answers (5 "somewhat agree", 6 "agree" and 7 "strongly agree") into the "helpful" category, and all negative answers (1 "strongly disagree", 2 "disagree", 3 "Somewhat disagree") into the "not helpful" category.

Results

Holistic Scores

As can be seen in Figure 2, the mean scores of the control group and the experimental group in the pretest were very close to each other, which were 3.8
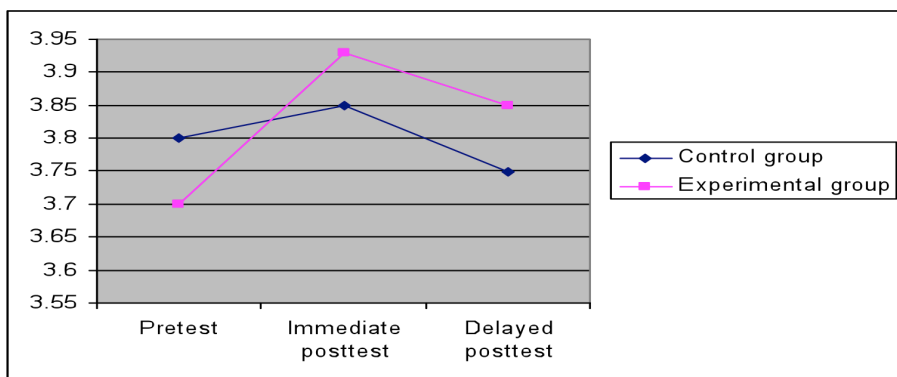
Figure 2. Mean scores of the three writing tests

Table 1. Appropriacy Scale

Scale	Category	Description
1	Appropriate	Acceptable to use the target phrase in the context
2	Less appropriate	Grammatically acceptable, but seldom used in argumentative context
3	Inappropriate	Grammatically or semantically incorrect; definitely not used in the context

and 3.7 respectively. In the immediate posttest, the average score of the control group increased to 3.85 with a mean improvement of 0.05 while the experimental group's score rose to 3.95 with a better mean improvement of 0.23. In the delayed posttest, though, the two groups' mean scores dropped slightly in comparison with the immediate posttest, but the mean of the experimental group remained higher than that of the control group.

Accurate Use of Abstract Nouns

Error-free ratios between groups and the improved use of the abstract nouns within the experimental group were compared. Two of the holistic scorers evaluated the students' use of the target nouns by categorizing them on a 3-point scale: appropriate, less appropriate, and inappropriate (see Table 1). If the use of a noun fell into the category of "less appropriate" or "inappropriate", it was characterized as an error. Figure 3 shows the error-free ratios between the control

group and the experimental group in the pretest and the immediate posttest. In
the pretest, the control group and the experimental group had similar error-free
ratios in terms of the use of the five target nouns (39.9% and 37.9% respectively).
However, in the immediate posttest, the experimental group's error-free ratios
increased to 88.2% while the control group improved only to 47.2%.

Improved use of the nouns by the experimental group was further investigated
and categorized into three types: positive change, negative change, and no
change. Positive changes were described as inappropriate or less appropriate use
of the nouns in the pretest but appropriate use in the immediate posttest. Negative
changes were appropriate use of the nouns in the pretest but less appropriate
or inappropriate in the immediate posttest, and "no change" was described as
inappropriate or less appropriate use of the nouns both in the pretest and the
immediate posttest.

As can be seen from Figure 4, the instances of positive change (42 instances
in total) outnumbered negative change (3 instances) and no change (6 instances).
Table 2 provides examples of positive changes in using controversy, objection,
and criticism by students in the experimental group. Two examples are given for
each abstract noun.

Although the occurrences of positive change far outnumbered the other two
categories, cases of negative change and no change in the immediate posttest
still existed. These cases are illustrated in Table 3, which gives a list of examples
of no change and negative change. For example, the error made by S3 in the pretest
(i.e., in criticism way) was repeated in the immediate posttest (i.e., hold a criticism

Figure 3. Error-free ratios in the pretest and the immediate posttest
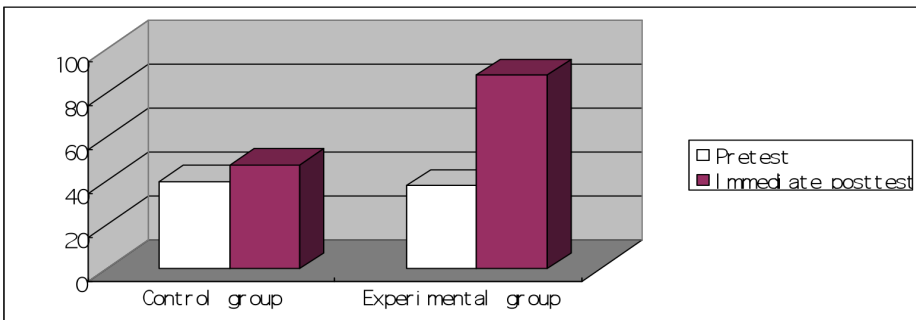
problem), where criticism was used to modify abstract nouns way and problem. It might be that this particular student intended to say in the pretest that the development of tourism should be taken into consideration critically. Similarly, in the immediate posttest, she repeated the wrong pattern by using criticism as a pre-modifier to follow a noun, while the correct use could probably be we should look at this problem critically.

Table 2. Examples of positive change in the experimental group

Student	Pretest	Immediate Posttest
S32	But when so many people pour into the beautiful scene, problem arises as well, which **makes** it *controversy*.	However, lottery games have **provoked considerable** *controversy*.
S13	In short, the *controversy* **of** the tourism is unlikely to end.	People waste a lot of money and time on it, which **stirs considerable** *controversy*.
S2	There are more and more people who **are more in** *objection* **of** tourism and the impact of tourism is in controversy.	The miseries brought about by lottery **lead to great** *objection* **raised against** lottery.
S18	Nowadays, some people **have an opinion of** *objection* towards tourism.	There are more and more people **raising** *objection* **to** the lottery.
S26	However, we should understand the situation **in** *criticism*.	**Another** *criticism* **is that** they are games which are easy to get addicted to.
S6	Between **objection and agreement**, I **choose** *criticism* **side** in all controversy.	Some people praise that lottery will be a boost to the country's domestic consumption while the lottery also **came under intense** *criticism* by other people.

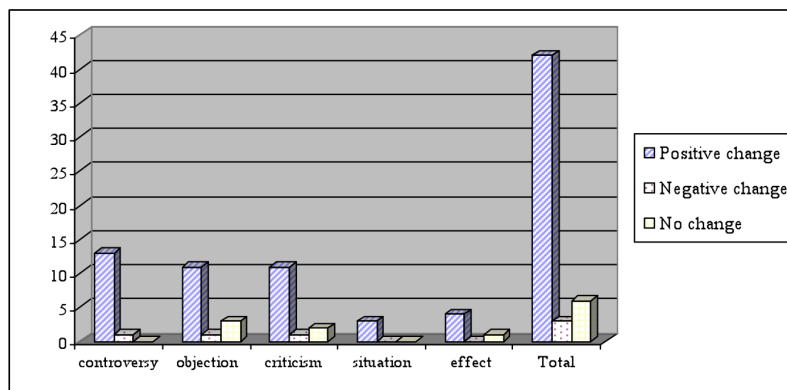Figure 4. Changes in using abstract nouns in the experimental group

Table 3. Examples of no change and negative change in the
          experimental group

Student	Pretest	Immediate Posttest
S3	So, tourism should be taken into consideration **in** *criticism* **way.**	We should **hold a** *criticism* **problem.**
S26	Though it causes many problems and makes some bad effects on our environment, **in** *objection*, it is good for the development of our world.	**The main** *objection* **of it** is based on the fact that the rate of winning the prizes is far too small.
S15	Furthermore, from tourism, the countries can show every aspect to the world and **have a better** *effect* **on** foreigners.	Many people hold the objection against them because they ignore the bad *effect* of lottery games especially **to the young children** who have less control of themselves than adults do.

Figure 5. Distribution of grammatical patterns in the immediate posttest
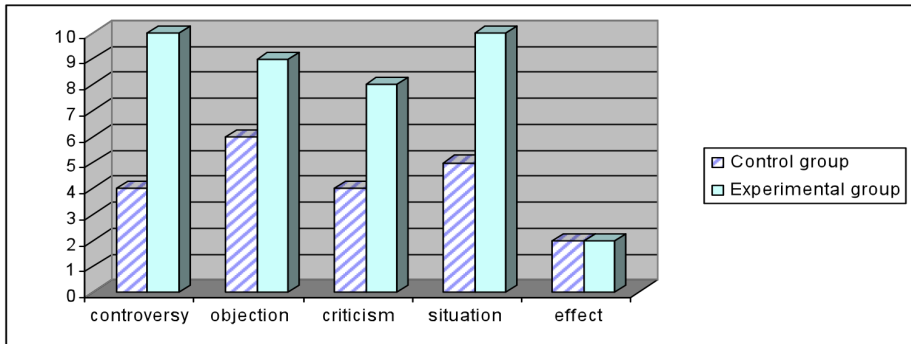
From Figure 5, we can see an overview of grammatical structures of the five target nouns used in the immediate posttest between the two groups. Overall, all the target nouns, except effect, were used with more grammatical patterns by the experimental group than by the control group. The grammatical patterns of effect were equal in both groups. They fell into two types: (1) V + effect by collocating cause and have; and (2) effect + Copular Verb BE as Subject (e.g., Another negative effect is that lottery games have caused many crimes).

Table 5. Overall evaluation on the concordance exercises (n=20)

Category	Agree (%)	Disagree (%)	No Opinion (%)	Mean	S.D.
Overall evaluation on vocabulary learning	95	5	0	5.90	0.91
Overall evaluation on L2 writing	95	5	0	5.50	0.94
Increasing confidence about using the words in L2 writing	95	5	0	5.70	0.92
Expecting more concordance exercises in future writing	85	15	0	5.50	1.27
Prefer learning words from the concordance lines than being taught by teacher	70	30	0	4.90	1.51

*1-3 = disagree, 4= no opinion, 5-7 = agree.

Table 6. Problems in doing the concordance exercises (n=20)

Category	Agree (%)	Disagree (%)	No Opinion (%)	Mean	S.D.
Time-consuming	75	20	5	4.50	1.31
Unfamiliar vocabulary	80	10	10	4.75	1.16
Cut-off sentences	80	20	0	4.85	1.63
Too many sentences	65	30	5	4.40	1.42
Limited number of sentences	30	60	10	3.45	1.19
Difficulty in formulating the overall rules of the usage of the words	45	50	5	4	1.68

*1-3 = disagree, 4= no opinion, 5-7 = agree.

Student Views of the Concordance Exercises

The follow-up survey focused on two aspects of students' attitudes towards the corpus-informed materials: (1) overall evaluation of corpus-informed materials, and (2) difficulties in doing the concordance exercises.

As noted in Table 5, vocabulary learning ranked the top among the categories. About 95% of the students were favorable toward concordance exercises, reporting

that they were helpful for vocabulary learning and increasing confidence about using the words in L2 writing.

Although the mean responses in Table 5 indicate that the majority of the students had a favorable attitude towards corpus use for vocabulary learning and L2 writing, Table 6 reveals a different perspective with 75% of the responses showing that it was time-consuming to do concordance exercises. However, half of the students reported they did not have difficulty in formulating overall usage rules for the words. As the student LJR wrote, "I think there are too [many] contents which cost our lots of time. It would be better if there is less exercise or we just [under]line the answer in the content and not need to write it out."

About 80 percent of the students had difficulty in doing the concordance exercises due to the cut-off sentences and the new words in concordances. The cut-off sentences (see the Appendix) hampered them in fully understanding the information presented in the concordance output. They remarked that the cut-off sentences prevented them from understanding the examples or the viewpoints from the sample texts in the concordance lines. As one student commented:

> However, the examples of the words using are not so perfect because some of them are just a part of a sentence, and we don't know what the whole meanings of the examples are. So these examples cannot well express the exact using of the words. Some of them just show the verbs or prepositions that can be used with them.

The cut-off sentences are the main reason for some students stating that topic-based concordance lines failed to provide them with ideas related to the writing topic. The students commented that it would be of help for them to grasp the ideas more thoroughly if they could go further into the full context of the target word and read the complete sentences or even the whole paragraph.

Discussion

A comparison of the holistic scores of the pretest, the immediate posttest and the delayed posttest did show that there was no statistically significant difference in overall writing quality between the control group and the experimental group. This finding indicates that the corpus-informed concordances did not

have a significant impact on students' writing outcomes in terms of overall writing quality.

However, L2 writing performance is subject to a number of linguistic variables besides vocabulary use. General ideas, text structures, grammatical use, and coherence and cohesion are equally vital factors that determine the overall writing quality. As an experimental study, this research was set up to elicit results within a fixed period of time. It is quite possible that the input of these five target vocabulary items would not make a difference in holistic writing scores within four weeks. In order to attain more desirable results of the effects of topic-specific corpus on overall writing ability, a long-term study with more lexical input in the form of concordances is worth carrying out.

Although there was no significant difference in the overall writing outcomes between the control and the experimental groups, the mean improvement of the experimental group in the immediate posttest was much higher than that of the control group. This could suggest that corpus-informed concordances did exert a positive effect on students' writing quality in the short term. This could be attributed to more accurate and complicated structural use of the target nouns after the treatment which may indicate that corpus-informed concordances are especially effective in helping learners to obtain lexico-grammatical patterns of the target nouns when compared with dictionary consultation. These types of learning resources have a notable effect on improving language use in L2 writing, particularly in relation to collocational uses and grammatical patterns. The effect is indeed twofold, which concerns both acquisition and production. First, it could help learners acquire a variety of collocational patterns. Second, acquisition of collocational patterns in turn would enable learners to generate more accurate and complex syntactic patterns.

Most encouraging was the students' positive views of the concordance exercises. In terms of their importance in vocabulary learning and improvement of writing ability, and enhancing confidence in using the new words in their writing, the overwhelming majority of students were strongly supportive of the concordance exercises.

The significant progress in using the target words in students' writing along with their overall positive attitudes towards corpus use testifies to the feasibility and usefulness of a corpus-informed approach to L2 writing instruction.

Conclusion

The findings in this study provide empirical evidence that corpora can be a useful resource for writing teachers to help students improve their lexico-grammatical use of vocabulary in their writing. The findings also indicate that the incorporation of corpus materials into writing instruction needs to be pedagogically mediated. In corpus-informed writing instruction, teachers play a central role from the initial stage of materials development to the implementation of corpus-related learning activities. For teacher practitioners, the statement by Johansson (2009) can be used as a gentle reminder that "corpora are important in basic research, and they have a role to play in the classroom as well. But let's not exaggerate. Corpora are no replacement for natural communication. They cannot replace the teacher" (p. 42).

References

Aston, G. (2004, July). *Corpus upon corpus: a bout of indigestion?*. Paper presented at the Sixth Teaching and Language Corpora Conference, Granada, Spain.

Braun, S. (2007). Designing and exploiting small multimedia corpora for autonomous learning and teaching. In E. Hidalgo, L. Quereda, & J. Santana (Eds.), *Corpora in the foreign language classroom: Selected papers from TaLC 2004* (pp. 31–46). Amsterdam: Rodopi.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*, 213–238.

Coxhead, A. & Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing*, *16*, 129–147.

Flowerdew, L. (2010). Using corpora for writing instruction. In M. McCarthy & A. O'Keeffe (Eds.), *The Routledge handbook of corpus linguistics* (pp. 444–457). London: Routledge.

Hyland, K. (2003). *Second language writing*. Cambridge: Cambridge University Press.

Johansson, S. (2009). Some thoughts on corpora and second language acquisition. In K. Aijmer (Ed.), *Corpora and language teaching (pp.33–45). Amsterdam/ Philadelphia: John Benjamins.*

Johns, T. (1986). Micro-Concord: A language learner's research tool. *System*, *14*(2), 151–162.

Leki, I. & Carson, J. (1994). Students' perceptions of EAP writing instruction and writing needs across the disciplines. *TESOL Quarterly*, *28*, 81–101.

McCarthy, M. (2008). Accessing and interpreting corpus information in the teacher education context. *Language Teaching*, *41*(4), 563–574.

Nelson, G. (2009). Comparing Englishes: ICE and ICELite. *Proceedings of the 6th national natural language processing research symposium* (pp. 12–17). De La Salle University, Manila.

Rayson, P. (2002). *Matrix: A statistical method and software tool for linguistic analysis through corpus comparison*. Unpublished doctoral dissertation, Lancaster University.

Read, J. (2004). Research in teaching vocabulary. *Annual review of Applied Linguistics*, *24*, 146–161.

Sinclair, J. & Carter, R. (2004). *Trust the text: Language, corpus and discourse*. London: Routledge.

Tribble, C. (2001). Small corpora and teaching writing: Towards a corpus-informed pedagogy of writing. In M. Ghadessy, A. Henry, & R. L. Roseberry (Eds.), *Small corpus studies and ELT: Theory and practice* (pp.381–408). Amsterdam/Philadelphia: John Benjamins.

Tribble, C. (2002). Corpora and corpus analysis: New windows on academic writing. In J. Flowerdew (Ed.), *Academic discourse* (pp. 131–149). Harlow, UK: Longman.

Tribble, C. & Jones, G. (1997). *Concordances in the classroom: A resource guide for teachers*. Houston, TX: Athelstan Publications.

Thurstun, J. & Candlin, C. (1997). *Exploring academic English: A workbook for student essay writing*. Sydney: NCELTR.

Widdowson, H. G. (2002, July). *Corpora and language teaching tomorrow*. Keynote lecture delivered at the Fifth Teaching and Language Corpora Conference, Bertinoro, Italy.

Appendix: Excerpt of the corpus-informed material

Pre-writing vocabulary study

Directions: Study the concordance, underline or highlight the central group of words that stand alone, as has been done in the first example. Then answer the questions which follow. Do not worry that these are cut-off sentences—just familiarize yourself with the key words.

Objection

Study the lexico-grammatical patterns

Study the concordance lines of *objection* and answer the following questions.

1.	Lottery was eventually approved. *Much of the* **objection** *to* the National Lottery came from church leaders.
2.	profits to charity, but was rejected. My *personal* **objection** *to* Camelot as the lottery organizer is that a large
3.	most famous businesses and Families had *a particular* **objection** *to* the start of the National Lottery.
4.	All rely on participating viewers who *have no great* **objection** *to winning* their moment of glory by doing their best
5.	harmless fun which won't break the bank. The *main* **objection** *to the lottery* is based on the grounds that those who
6.	than that, a tax on the poor. The *main* **objection** *came from charities* who predicted that charitable donate
7.	like horse-racing and casinos. *These* **objections** *became much greater* with the introduction of *scratch cards*
8.	are how bitterly unpopular taxation is; therefore, **objections** *raised against the lottery concerned* it being marketed
9.	or saving for a holiday or a car. *There were also* **objections** *raised to* the amount of money the proposed jackpot was
10.	and psychological drawbacks. There were two main **objections** *against* the introduction of the national lottery
11.	would have been donated to charity. *Another* **objection** *raised* was that the National Lottery would
12.	The proposal of a lottery brought about *many* **objections** *and complaints*. There were, and still are, two
13.	of conservative government ! *Despite the numerous* **objections**, the introduction of the lottery has induced a 'fever

1). Which adjectives are used before objection(s)?
   Please write down the phrases. e.g., the main objection

2). Which verbs or verbs phrases are used with objection(s)?
   Please write down the phrases.

3). Which preposition commonly follows objection(s)?
   objection(s)_____

About The Author

Huang Zeping is a doctoral student at the Chinese University of Hong Kong. Her interests are in second language writing and corpus linguistics. She has taught English in mainland China and in Hong Kong.