

THE PLAUSIBILITY AND IMPLAUSIBILITY OF USING CORPORA

Siaw-Fong CHUNG

Department of English, National Chengchi University,
No. 64, Zhi-Nan Road Sec. 2, Wen-shan District, Taipei, Taiwan R.O.C.

sfchung@nccu.edu.tw

Manuscript received 16 January 2018

Manuscript accepted 11 April 2018

ABSTRACT

Corpora are resourceful tools for linguistic observation, as they can provide quick statistics about a word compared with manual analysis. However, in some circumstances where qualitative analysis is needed, the use of a quantitative method may not produce the required results. Based on the keyword *love*, four types of materials were used to show the plausibility and implausibility of using corpora. First, romantic English songs were used to illustrate the kind of input students received from these songs. Second, we compared the songs to students' essays on general essay topics. Then, we contrasted the first two sets of data to a native speakers' corpus. Lastly, we showed how the creativity element can be reproduced by students through extended use of metaphors "Love is X". While the students may have been exposed to more colloquial uses of *love* in love songs, their general essays showed a more neutral use, with fewer occurrences of *love*. In comparing the different texts, this paper will show how a corpus-based approach can be complemented by a qualitative examination of texts. Regarding pedagogical implications, corpora are a good resource of authentic materials and should be encouraged in the classroom as a reference for recurring language patterns.

Keywords: corpus, creative writing, metaphors, students' essays, love, romantic songs

Introduction

As time goes on there will be alternative theoretical positions formulated, and new descriptions begun, but for the present, any teacher or student can readily enter the world of the corpus and make the language useful in learning (Sinclair, 2004, p. 297).

The use of corpora and concordances in the English classroom is not revolutionary. Advancements in technology, with easy access to the Internet, have accelerated the

use of corpora in the classroom. Many have written about using concordances in class, among which are O’Keeffe, McCarthy, and Carter (2007), Tsui (2005), Todd (2001), and Cobb (1997). Among these, Tsui (2005) used corpus evidence to answer three frequently asked questions by students in Hong Kong. These questions include: (a) synonyms (e.g., *big* versus *large*); (b) “grammar rules and conflicting evidence—countable and uncountable nouns” (p. 348) (e.g., *less than* versus *fewer than*); and (c) “prescriptive stylistic rules” such as whether “‘because,’ ‘and,’ and ‘but’ can be used to start a sentence” (p. 349). On the other hand, Todd (2001) found that “students who are able to induce valid patterns from a self-selected concordance are also more likely to be able to self-correct” (p. 100), but the relation between these two needs further observation.

As Svartvik (1996) stated in his article, “corpora are becoming mainstream,” predicted that “corpora will be commoner” in language learning, as “[c]orpora...will improve our understanding of the language learning process and result in better methodologies and teaching materials” (p. 11). According to Aston (1997), the easiest way is “to treat them [corpora] as sources of classroom materials which the teacher can select from and adapt according to requirements” (p. 52). Corpora have also been used in disciplines such as English for Specific Purposes (ESP) (Gavioli, 2005) and literature (Adolphs, 2006). Despite these benefits, many EFL practitioners are still more comfortable with the use of qualitative materials in class, mainly because they “may not be immediately clear what this [corpus linguistics] approach offers that other methods of language description do not” (Renouf, 1997, p. 259). This paper will show the plausibility and implausibility of using corpora by examining the word *love* as a case study.

Four types of materials were used in this study to show the plausibility and implausibility of using corpora. These four types of materials are (a) romantic songs, (b) students’ essays; (c) native speakers’ general data; and (d) students’ creative essays. In each of these materials, we used the word *love* to explore the advantages and disadvantages of using corpus linguistic skills in dealing with the data.

According to corpus linguistics, a general corpus is “a corpus of texts of many types” and is “often used as a baseline in comparison with more specialized corpora” (Hunston, 2002, p. 15). Corpora such as romantic songs or students’ essays are called specialized corpora. A specialized corpus is a “corpus of text of a particular type” which is “used to investigate a particular type of language” (Hunston, 2002, p. 14).

In this study, the uses of *love* in romantic songs is to illustrate the kind of input the students received from English songs. Lems (2001), for instance, enlisted several uses of music in teaching ESL adults: “Music can be used in the adult ESL classroom to create a learning environment; to build listening comprehension, speaking, reading, and writing skills; to increase vocabulary; and to expand cultural knowledge” (online paper). Littlewood (1998) also argued against the use of controlled materials in second language learning:

Most teaching approaches to date have been based on the assumption that the learner’s input should be carefully controlled and graded for structural complexity. However, it is now sometimes suggested that this assumption is

mistaken, and that classrooms should concentrate on providing the kind of input...[that is]...comprehensible, interesting, relevant, but not strictly graded.” (Littlewood, 1998, p. 60)

However, to what extent romantic songs can provide vocabulary to students is unknown, although it is known that some songs contain more difficult vocabulary than others. The use of music and authentic materials are also not new in TESOL, yet it must be cautiously pointed out here that not all outside input forms good resources for teaching. We will discuss some of the features of romantic songs, and show that despite some colloquial features that are unwanted in classroom teaching, we can still make songs a good resource for learning. One of the features that can be introduced in the classroom is the creativity. In this paper, we inspect creativity through analysing students’ creative writing on extended metaphors.

When dealing with creative writing, statistics may not help much because creative thoughts need no quantification unless one is aiming to find out whether a similar thought is repeated. If one only wants to ascertain whether a student is creative, it is less likely that he or she will look for a repetition of ideas. In other words, creative tasks go in the opposite direction of quantifying whether a keyword is recurring in the text. A keyword list generated from corpus, which provides the listing of words arranged from those most frequently appearing to those least frequently used is certainly not required in assessing creative writing. In such a case, it does not make sense to quantify how many times a similar word is mentioned when students talk about a topic such as *love*. However, quantitative methods could be applied if a researcher wanted to find out how many students came up with a similar thought when the same creative prompt was given. This result is different from analysing the most frequent combinations of words, which can be carried out with most concordancers or corpus tools. A quantitative analysis of a collection of creative writing will probably return the number of hits for *my love*, *gives love*, *our love*, *love eating*, etc., but these do not reflect the essence of a creative writing. The use of such a measure loses the creative thought hidden in the text. Therefore, for a TESOL practitioner, it is important to know the application of corpus tools—what worked and what did not. In this paper, we show these by using different types of materials.

Methodology

In the sections that follow, the word *love* will be used as an example to explain how a corpus can be utilised to make linguistic observations. As the issue of “love” remains an interest for many students, the examination of the word *love* crystallised how the students in this study encountered this word in authentic settings. Several types of materials were examined, including (a) romantic songs; (b) students’ essays on general topics (without focusing on metaphors); (c) native speakers’ data from a general corpus (British National Corpus); and (d) students’ extended use of metaphors (LOVE IS X) in a short task.

First, the use of *love* in romantic songs was examined, a resource that students often rely on for entertainment in which they are exposed to them

intentionally or unintentionally. Then, the students' writing in the classroom was examined to determine whether their uses of *love* were the same as those found in the romantic songs. This established how *love* appears in the language when no emphasis was given. Next, a native speaker corpus was used to show how *love* is portrayed in a corpus that focuses on neither romantic song lyrics nor the students' works. Finally, a creative writing task on *love* was presented to see whether the students were able to produce uses of *love* that were novel and creative.

For the romantic songs, 100 randomly selected lyrics from English love songs were collected from the website Romantic Song Lyrics.¹ AntConc 3.2.4 (Anthony, 2005), a freeware concordancer, was used for the calculation of the features of the corpora.² The total word count was 21,502 for the 100 files, indicating that each file contained 215 words on average. The results returned 1,851 word types (e.g., distinctive words—*love* and *loves* are two distinctive types) in the corpus. The search for *love* in the love songs returned 395 instances (for *love*, *loves*, *loved*, *loving*, and the form *lovin'*).

For the general essay topics, 41 essays written by undergraduate sophomores at a university in Northern Taiwan were examined. Most of the essays were based on writing strategies rather than topics. The purpose of observing the use of *love* in the general essays was to see how the word *love* appeared in these essays. It was hypothesized that the word *love* would appear in rather similar patterns (*I love you*, *you love me*, *our love*, etc.) as those in the love songs, but it would be used less in ordinary essays. The 41 essays amounted to 21,941 words with 3,516 types.

In order to see the general linguistic behaviours of *love*, the results from the British National Corpus (BNC) were further extracted through the BNCweb (cf. Hoffmann et al., 2008), an online interface with additional analytical functions designed specifically for the BNC.³ The BNC is a 100-million-word corpus of written (90%) and spoken (10%) language from a variety of sources. The corpus represents British English from the late twentieth century. The search for all forms of the word *love* in the BNC showed that *love* returned 28,429 instances (or 289.17 instances per million words).

The use of analogies of metaphors was introduced to three different classes of undergraduate students. An essay requiring the students to write LOVE IS X, in which X was restricted to sports or transportation devices, was conducted. It was hypothesized that, upon receiving specific instructions, the students would be able to write about *love* creatively. Unlike romantic songs, which consist mainly of conversational uses of *love*, the last task in the study used *love* in extended metaphors. This analogy exercise also showed how the students were able to associate elements in two different domains to create novel metaphors. As these creative metaphors were unique, they were examined as a complete discourse rather than using a statistical analysis. About 50 students received instructions on analogies and how to use them in writing.

Results

Romantic Songs

It was found that the form *love* was the most frequently appearing content word in the romantic song corpus, ranked number eight after seven pronouns and function words (*the, and, and to*). In these romantic songs, *love* was most frequently found in *love you* (61), *I love* (44), and *I love you* (29). Other examples were *don't lay our love to rest, as long as you love me, with loving eyes*, etc. To show how *love* was used in the love songs, Table 1 below lists the two- to five-word (i.e., n-grams) combinations of expressions containing *love*.

Table 1
List of two- to five-word combinations of the word *love* in love songs

2-grams	3-grams	4-grams	5-grams
love you	I love you	ever been in love	you ever been in love
I love	been in love	remember me with love	in the arms of love
in love	you love me	the arms of love	do is make love to
my love	arms of love	after we make love	is make love to you
love me	me with love	baby I love you	one and all for love
love is	all for love	do is make love	wanna do is make love
of love	all the love	is make love to	all in love is fair
the love	and I love	make love to you	all the love in the
love to	baby I love	and all for love	as long as you love

Going through the list, one might agree with the author that these expressions did not bring many learning benefits to the students, except for the few that might be considered for multiword teaching (shaded). Therefore, it could mean that some romantic songs are colloquial and bring less learning effects to the students, an observation partly made by Beasley and Chuang (2005).

Students' Essays on General Topics

When comparing the essays in the learner corpus with the romantic songs, it was found that the word types were more varied in the learner corpus than in the romantic songs, indicating that romantic songs repeated similar words more often than the essays did because songs often repeated the same verses. In most romantic songs, words are repeated and there is little variation of patterns, although the author is aware that this is also part of the characteristics of songs.

Love in the learner corpus returned only 12 instances. Most of the instances of *love* were nouns (*love story, romantic love, forbidden love, children without love, etc.*). The students also used *love* to describe their favourite actions, such as *love shopping, love organic food, whom he loved most, and love videos and music*, three of which indicated favourites and preferences and one which referred to an intimate love relationship. Table 2 shows a list of n-grams taken from the learner corpus.

Note that because the instances of *love* were few, these multiword combinations were likely to be single instances only or they may have been repeated based on a different boundary (e.g., *and spread love and care* versus *spread love and care to* are both 5-grams but they were from the same sentence).

Table 2
List of two- to five-word combinations of the word love in the students' essays

2-grams	3-grams	4-grams	5-grams
love and toward love	and spread love attitude toward love	and spread love and attitude toward love and	about out of his love and spread love and care
and love	emotion toward love	for his romantic love	depict less emotion toward love
his love	his love and	his love and despair	for his romantic love story
I love	his romantic love	his romantic love story	his romantic love story and

Based on Table 2, one can see that the uses of *love* in these essays are of a rather different nature than those from the romantic songs, as these uses are less conversational.

Comparisons to the Native Speakers' Data

The distributions of *love* in both specialized corpora were compared to the BNC, as shown in Table 3:

Table 3
Distributions of word forms of love in romantic songs, students' essays, and the BNC

	Romantic Songs	Students' Essays	British National Corpus
love	370 (93.7%)	11 (91.7%)	22,265 (78.3%)
loved	5 (1.3%)	1 (8.3%)	4,313 (15.2%)
loves	4 (1.0%)	0 (0.00%)	1,340 (4.7%)
loving	9 (2.3%)	0 (0.00%)	510 (1.8%)
lovin'	7 (1.7%)	0 (0.00%)	0 (0.00%)
lovest	0 (0.0%)	0 (0.00%)	1 (0.0%)
Total	395 (100.0%)	12 (100.0%)	28,429 (100.0%)

In Table 3, more than one-third of *love* appears as the form *love*, which includes both nouns and verbs. The romantic songs, in particular, contained 1.7% of *lovin'*, a feature often found in song lyrics but is considered informal in written texts.

Table 4 further compares the distributions of parts-of-speech with the BNC⁴:

Table 4
Distributions of parts-of-speech of love in romantic songs, students' essays, and the BNC

	Romantic Songs	Students' Essays	British National Corpus
Noun	268 (67.8%)	8 (66.7%)	14,190 (49.9%)
Verb	126 (31.9%)	4 (33.3%)	14,195 (49.9%)
Modifier	1 (0.03%)	0 (0.00%)	1 (0.00%)
Others (e.g., Mr. Love)	0 (0.00%)	0 (0.00%)	43 (0.2%)
Total	395 (100.0%)	12 (100.0%)	28,429 (100.0%)

In Table 4, both the romantic songs and the students' essays contained more uses of *love* as nouns than as verbs. Even though the data for *love* was small in the learner corpus and their formality differed (from the comparisons in the last section), it displayed the same distribution as the romantic songs did, with the exception that the learner data did not contain examples of *love* as a modifier. The BNC, however, displayed a different pattern of distribution, with approximately half of its instances as nouns and the other half as verbs. The comparisons in Tables 3 and 4 with the BNC thus provide a contrast between the purpose-specific corpora and a general corpus.

A further investigation into the types of genres *love* may appear in was also examined using the BNCweb. The noun and verb forms of *love* were searched separately. The results from the BNCweb showed some interesting findings, as shown in Table 5, which presents the overall distribution of *love* in a specific text types in the BNC. In Table 5, first, it was found that *love* (noun) appeared more often in written texts (148.36 per million words) than in spoken texts (114.31 per million). In contrast, *love* (verb) appeared more often in spoken texts (175.89 per million) than in written texts (140.65 per million). Comparing all four, *love* (verb) in the spoken genre constituted the highest frequency. For both verb and noun, most of the spoken context took place within the scope of the "leisure" topic (see bottom of Table 5).

Table 5
Use of love in different text types in the BNC

Love (Noun)		Love (Verb)	
Genre	Frequency per Mil.	Genre	Frequency per Mil.
Written	148.36	Spoken	175.89
Spoken	114.31	Written	140.65
Total	144.75	Total	144.39

Table 5 (continued).

Derived Text Type	Frequency per Mil.	Derived Text Type	Frequency per Mil.
Fiction and verse	335.85	Fiction and verse	374.38
Spoken conversation	187.77	Spoken conversation	305.86
Newspapers	139.29	Newspapers	136.31

Text Domain	Frequency per Mil.	Text Domain	Frequency per Mil.
Informative: Belief and thought	472.42	Imaginative prose	375.05
Informative: Arts	348.14	Informative: Belief and thought	268.31
Imaginative prose	338.07	Informative: Arts	215.06

Age of Author	Frequency per Mil.	Age of Author	Frequency per Mil.
0-14	520.49	0-14	251.85
15-24	283.83	25-34	232.46

Age of Audience	Frequency per Mil.	Age of Audience	Frequency per Mil.
Teenager	313.67	Teenager	415.05
Adult	146.83	Child	298.85
Child	129.5	Adult	134.82
Any	101.81	Any	93.61
Total	148.36	Total	140.65

Type of Interaction	Frequency per Mil.	Type of Interaction	Frequency per Mil.
Dialogue	120.71	Dialogue	189.2
Monologue	78.1	Monologue	100.51
Total	114.31	Total	175.89

Domain (spoken context-governed texts):	Frequency per Mil.	Domain (spoken context-governed texts):	Frequency per Mil.
Leisure	122.58	Leisure	130.2
Public/Institutional	60.38	Public/Institutional	95.66
Educational/Informative	51.02	Educational/Informative	71.07
Business	13.26	Business	42.11
Total	63.96	Total	86.79

Both noun and verb forms appeared most frequently in “fiction and verse,” “spoken conversation,” and “newspapers.” However, *love* (noun) appeared most often in texts that were informative in nature, especially in the topic “belief and thought” (472.42 per million), while *love* (verb) appeared more often in “imaginative prose” (375.05 per million). As for the age of the authors, it was found that both noun and verb forms were used most often by teenagers in the BNC. Thus, the corpus provided statistical information, which helped in visualising the use of a particular word quickly.

Students’ Extended Use of Metaphor “Love is X”

This section aims to keep the creativity element of romantic songs by asking students to produce analogies in writing. In texts where creativity is high, the frequency count of a corpus may not have helped because creative writing is often novel and atypical.

Not all students used metaphors effectively and not all essays were included. The students were required to produce a writing task comparing *love* to a different thing. It was found that the students became more creative when they were “required” to produce an analogy. Many students responded that writing analogies were difficult, but their writing products were better than what they originally imagined. Some creative excerpts are given in (1):

(1) Love is a Transportation Device

- (a) To me, **love** is similar to a motorcycle. When you ride the motorcycle, you can’t see every thing [sic] clearly with its high speed.
- (b) **Love** is a train on the rail.... There seems to be no end of the rail. Pushed by the thirst, we keep **loving** and **being loved** with someone and never stop.
- (c) **Love** is a train. It supplies the place for passengers to rest on.... **Love** isn’t selfish, since it doesn’t care about what the passengers are.
- (d) I think **love** is like a train...and on the way you will see beautiful views like when you fall in **love**, everything in yours eyes is all colorful and brilliant, but when also go through the cave like when you and your **lover** have an argument, you will feel everything is dark and helpless.... **Lover** is like motorcycle you can’t stand more than two people’s weights.
- (e) **Love** can also be a trunk. In a trunk, there is always a great load of things in it. It brings them from place to place, and sends them to the destination required.
- (f) Some say **love** is car racing. When two people are in **love**, they persuade [pursue] excitement and their emotions move too fast. It’s easy to enjoy the fun in racing cars, but such activity comes with a risk. **Love moving too fast** is always at risk of having a crash.
- (g) **Love** is a big, large truck. It may look heavy and stupid, run slowly, look like a little dangerous but it can carry the most things.
- (h) But sometimes with a long time **love** is getting similar to an old car. Outside it might look nice with all its gorgeous shows [hard to read] and fancy paint, but the engine used all of resources and you need all the time maintain and

report it for it not to fail.

Here, different perspectives of love could be seen. The students were more creative in their uses of *love*, and sometimes a metaphor was created not by a single instance of *love* but instead it ran through the whole text.

The examples in (2) are from the second metaphor “Love is Sports”:

(2) Love is Sports

- (a) **Love** could also be compared to playing volleyball.... Each one of the members could be players. **Love** is sometimes not just one-sided. It must be played by a group with strong unity.
- (b) **Love** is like playing sports, you need time, passion and companions. **Love** makes you happy, the awesome feeling is close to what you feel when you win a sport game.... **Love** is like playing sports, you know the rules, but there are still many things you didn't expect will happen; **love** is like playing sports, because you will try your best to win the game.
- (c) **Love** is playing baseball. **Love** comes suddenly and strongly like a straight ball, you try to bat it but usually you just miss it.... But it's a game; we should enjoy it, try to have fun, and when we see **love** come, just catch it or bat it. Seize the opportunity and don't let it go.
- (d) In the tennis game, when we start to hit the ball, it's like that we take courage to show our **love**. If he or she hits the ball back, that means we get the responses.
- (e) Catching every kinds [sic] of balls is like solving every kinds [sic] of problems that appear in the **lovers'** daily life.
- (f) Bowling is a sport which use [sic] a big ball to strike many bottles. Making all of the bottles down is the biggest prize. So I think it is similar to those people who just want to conquer.
- (g) Romantic **love** is like table tennis, it takes two to create romance. Two players are separated at a fair distance by a net, not to [sic] far as to lose affection, but also not too close and not to [sic] physical to loses [sic] its beauty.

Holmes (2004) stated, “[m]etaphorical extension develops language away from the bedrock of common responses to the experience” (p. 147). By using metaphors in class, students were able to become more creative and imaginative. If the above excerpts were inserted into a concordancer, the following collocations would be obtained: *love is* (13 instances), *in love* (2), *love can/come(s)/could/isn't/makes/moving* (1 each).

The results, nonetheless, showed that none of the uses of *love* was creative or special, but by reading the excerpts in (1) and (2), one can see that this is untrue, as stated by O'Keeffe (2007):

It is difficult to identify creative uses of language a priori and then search a corpus for them.... It is also difficult to recognise what may or may not be creative in the lists of words and chunks extracted from a multi-million-word corpus. To make such assessments we have to read the corpus screen by

screen and make judgements and evaluations of purposes and functions. (p. 197)

This part of the analysis provided evidence that showed how quantitative data look for repeated patterns in a collection of texts, while creative writing goes slightly beyond the accessibility of an ordinary concordancer. When extended metaphors occur, quantitative analysis cannot effectively show the patterns of how a metaphor works because the functions of a concordancer are not designed for this purpose.

Further Discussion of Using Romantic Songs and Extended Metaphors in Classroom

The corpora of romantic songs and the students' essays serve as tools for observing how a word such as *love* might be used. As a medium frequently accessed by students, the role of romantic songs does deserve some attention. It was observed that romantic songs showed an excessive use of first- and second-person pronouns in noun phrases such as *my love*, *love you*, etc. Earlier studies such as Murphey (1992) also found that love songs were full of personal pronouns, and the language was usually conversational. This was also seen in the corpus of romantic songs, in which *you* (1,169 instances) and *I* (926 instances) were the two most frequently found words in the whole corpus. It was found that the romantic songs were generally filled with the repeated uses such as *love you*, *I love you*, *love me*, *loving eyes*, etc., but these uses did not help the students to expand their vocabulary. Other vocabulary items frequently found in love songs were *baby*, *heart*, *forever*, *cry*, etc. Contrastively, the students' essays focused on writing strategies that avoided the conversational use of *love*.

The implication for teaching is that, when students listen to English romantic songs, it is likely that they will be exposed to conversational language. Even though Beasley and Chuang (2005) found that, it was unlikely that listening to songs alone increased students' vocabulary level; they agreed that written lyrics do play a role in improving their learning. The author believes that this depends on the selection of songs and the type of course design used. It is true that the current study showed a similar observation because the song lyrics were more conversational. The vocabulary task in (3) shows how a song was changed in a lesson for the students. (3a) shows the first verse of the song "I Believe" and (3b) shows a vocabulary task given to the students before they listened to the original song. Students were asked to fill in the slots with nouns (underlined), verbs (boxed), an adjective (shaded), and an adverb. Although the parts-of-speech were predetermined, the students invented creative poems (one example is given in [3c]).

(3)

(a)	<p style="text-align: center;"><i>I believe for every drop of rain that falls A flower grows, I believe that somewhere in the darkest night A candle glows, I believe for everyone who goes astray, Someone will come to show the way, I believe, I believe.</i></p>
(b)	<p>I <input type="text"/> for every <input type="text"/> of <input type="text"/> that <input type="text"/> A <input type="text"/> grows, I <input type="text"/> that somewhere in the <input type="text"/> A <input type="text"/> glows, I <input type="text"/> for everyone who goes <input type="text"/> (adv), Someone will come to <input type="text"/> the <input type="text"/>, I <input type="text"/>, I <input type="text"/>.</p>
(c)	<p style="text-align: center;">I sing for every cities [sic] of people that sorrows [sic] A hope grows, I see that somewhere in the deepest desperation A hope glows, I pray for everyone who goes hopelessly, Someone will come to save the world, I pray, I pray.</p>

The students reflected that they felt satisfied after completed this task. A student wrote: “But after I finish, there is big satisfaction filled in my mind. It’s like I build a Taipei 101 by myself. And the song is another struck to me. Every words [sic] used in the song are [sic] really simple but moving”. Another student said “[a]lthough it’s hard to fill every blank because finding a suitable word and making the sentences coherent are difficult, I like this exercise.” These reflections suggest that using songs in the classroom can be meaningful when well-designed.

Some songs, especially those that use metaphors extensively, should be analysed individually. The controversy between typicality presented in a corpus via quantitative measures and individualized instances is best captured in the statement by Renouf (1997):

It may not be obvious to the teachers at the outset that corpus-based study is essentially quantitative—i.e. its purpose is to identify the generalities of language use.... [L]anguage teachers are inveterate snippers of juicy instances of usage from newspapers, but these individual occurrences cannot be evaluated or relied on in terms of typicality. (p. 258)

It is therefore important to show to TESOL practitioners what corpora can do and what they cannot do, as presented in this work. However, this does not indicate that corpora have no role in analysing literary pieces or creative works. For instance, Burrows, as early as 1987, used corpora to create a frequency list for the novels by Jane Austen. Adolphs (2006) also showed that analysis of a literary character and/or point of view in a work of fiction can easily be done with quantitative methodology. What is important here is what kind of analysis is needed. In the following sections, we summarise the advantages and disadvantages of corpora.

Advantages of Using Corpora

The advantages of using corpora are many, as showed in previous literature. For example, Timmis (2015), in his book *Corpus Linguistics for ELT*, emphasized the significance of corpora to ELT practitioners:

Ideally, it would take us to a point where we stop talking about the *potential* of corpora to be applied to ELT, which has been proclaimed for many years, to a point where the ability to carry out corpus research and/or interpret the significance of corpus findings for the classroom is regarded as the norm for ELT practitioners. (p. 8)

However, Timmis (2015) also mentioned that many have commented that “only a minority of students, teachers and materials writers directly engage with or in corpus research in practice” (p. 8). Since this is true in many classrooms, it is therefore crucial to clarify the advantages and disadvantages of using corpora.

It is important to know what a corpus can do. A corpus provides better contextual information that is sometimes missing in dictionary examples. In corpora, the contexts can be expanded according to the needs of the user. In doing so, learners can be trained to predict the meanings of an unknown word by reading the contexts on the left and right of a keyword. This training is essential in making a student an independent reader as a means to life-long learning:

Since the main rationale of corpora in teaching is their immediate availability for students’ use, it is important that the students should be able to acquire the necessary ‘hands on’ know-how, so that they can explore corpora for their own purposes. (Leech, 1997, p. 8)

In addition to serving as a context provider, corpora also includes frequency counts in the linguistic information. Jones and Waller (2015) suggested that dictionaries that include “frequency data and information on grammatical patterning” are “excellent, evidence-based resources for information about language” (p. 13). A dictionary that employs corpus information “will show us how often a word is used, the common contexts of use, the other words it commonly forms partnerships with and the grammatical patterns it is part of” (Jones & Waller, 2015, p. 13). These are among the advantages of using corpora for EFL purposes.

The authentic application of corpora is beneficial to both teachers and students. From the corpus, the context can be obtained by enlarging the contexts surrounding the keyword according to the needs of a user. With the complementation of contexts, teachers and learners can be more informed in terms of the meaning of a word and how and when to use it. In such case, corpora can be a good resource for exposing learners and teachers to authentic materials.

In addition, corpora can also be used to teach grammar. For example, the construction *love was ___ed away* can be taught as a unit to students; as Lewis (1997) suggested in his lexical approach, collocated phrases should be taught rather than single units. In addition, Ellis (1988) emphasized the importance of teaching formulas in the classroom. Frequently seen phrases such as *as long as you love me* and *romantic love* can be adapted to the teaching of formulas such as *long as you + VERB + me* and *ADJECTIVE + love*, or other forms generated from a particular pattern.

Since these formulas can be stored as templates by the learner, they “help to reduce learner’s processing difficulties” (Ellis, 1988, p. 81). Statistics in corpora can also be used to teach multiword expressions such as collocations. For example, one can see that the noun *love* can act as an agent in sentences such as *love took over my heart* and *love will make it alright* (examples taken from the collection of romantic songs used in this study). When *love* acts as a patient, it can form sentences such as *share love* and *express love*. In sum, one can say that corpora are resources that can present systematic data to learners and teachers. However, corpora are not omnipotent in all aspects. In view of this, the shortcomings of corpora will be discussed.

Shortcomings of Corpora

In some cases, the use of corpora is not as effective as qualitative analysis. In the following paragraph, for instance, a corpus was less effective in presenting the creativity in it:

(4) *My love is the garden window, protect you and surround you. It gives love a boundry [sic], as also focus the object/people inside the garden window...the development of our love is like Louvered windows. First you just open little of it, then you see little thing, it like the first stage of our love development, we love eating, we love icecream [sic]. Then you open more of it, you see more, it is like I love my parents, my friends. Open more, you see more, like you start to love someone. In the end, you open fully of the window, and you can see the full view from the window, it's the stage you love the whole world.* (student’s essay)

The analogy between *love* and the *garden window* forms a novel metaphor that does not conform to the systematic mappings of a conventional metaphor, a type of metaphor that consists of linguistic expressions that represent the connection between two mapped domains (Lakoff & Johnson, 1987). Conventional metaphors are expressions that exist in one’s language, but they are seldom noticed as metaphors (e.g., *we are at a crossroads* and *our marriage is at a bumpy road* are

representations of the conventional metaphor LIFE IS A JOURNEY). When two mapped domains (LOVE and WINDOW) do not form a conventional use, they are called novel metaphors. Novel metaphors like LOVE IS X are creative and less repetitive in use.

Conclusion

This study focused on the importance of using corpus-based research in teaching and learning. At the same time, it aimed to remind readers that corpora are not the only solution to all linguistic observations. Using *love* as an example, this paper demonstrated how corpora can be used for observing language patterns. The methodology demonstrated that specific norms of words could be shown by examining their distributional information in a collection of texts. For instance, the norms of *love* or any other words can be retrieved through running data from a general corpus.

Considering these advantages, the use of corpora in the classroom is a feasible idea. The major obstacle to the use of corpora is probably the skills-training of instructors and students. If this aspect can be solved, a step toward a statistical analysis of words can bring learners closer to the authentic use of words. Therefore, this paper serves as a recall of these beliefs about the usefulness of corpora in classroom teaching. The methodologies suggested herein can be utilised to raise students' awareness regarding the language itself using corpora.

Nonetheless, at the same time, this paper also pointed out the limitation of applying corpora in teaching. This was done by observing extended metaphors in a creative writing task. When instructed to do so, the students produced language not commonly featured in the corpora. It was also shown that the creativity in these writings may have been minimized if a quantitative measure was utilized. Since creative language is atypical, it cannot be quantified. Therefore, the other aim of this study was to encourage and at the same time to serve as a reminder to caution all readers regarding the use of corpora in TESOL.

Endnotes

¹ <http://www.romantic-lyrics.com/lyricsae.shtml>

² <http://www.antlab.sci.waseda.ac.jp/software.html>

³ <http://bncweb.lancs.ac.uk/bncwebSignup/user/login.php>

⁴ These percentages include probable cases. NN1-VVB (11.22%) or “ambiguous but most probably nouns” was grouped under “nouns” in Table 4. Similarly, VVB-NN1 (5.62%) or “ambiguous but most probably verbs” was grouped as “verbs.”

Acknowledgements

The author would like to thank the Ministry of Science and Technology, Taiwan, project grant 106-2410-H-004-109-MY2 for supporting the research herein. Advice from the editorial team of *Issues in Language Studies (ILS)* is much appreciated.

References

- Adolphs, S. (2006). *Introducing electronic text analysis: A practical guide for language and literary studies*. London, UK/New York, NY: Routledge.
- Ahrens, K. (2010). Mapping principles for conceptual metaphors. In L. Cameron, D. Alice, L. Graham, & T. Zazie (Eds.), *Researching and applying metaphor in the real world* (pp. 185-207). Amsterdam, NL: John Benjamins Publishing Company.
- Anthony, L. (2005). AntConc: Design and development of a freeware corpus analysis toolkit for the technical writing classroom. In the *Proceedings of the International Professional Communication Conference (IPCC)*, pp. 729-737.
- Aston, G. (1997). Enriching the learning environment: Corpora in ELT. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 51-64). London, UK/New York, NY: Longman.
- Beasley, R. E., & Chuang Y.-S. (2005). The effects of web-based American music, lyrics, definitions, and explanations on Taiwanese ESL learners. *Journal of Educational Technology Systems*, 34(4), 461-471.
- Burrows, J. F. (1987). *Computation into criticism*. Oxford, UK: Clarendon.
- Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System*, 25(3), 301-315.
- Ellis, R. (1988). *Classroom second language development*. London, UK/New York, NY/Toronto, ON: Prentice Hall.
- Gavioli, L. (2005). *Exploring corpora for ESP learning*. Amsterdam, NL/Philadelphia, PA: John Benjamins.
- Hoffmann, S., Evert, S., Smith, N., Lee, D., & Prytz, Y. B. (2008). *Corpus linguistics with BNCweb—A practical guide*. Frankfurt am Main, GE: Peter Lang.
- Holmes, R. (2004). *Mind, metaphor and language teaching*. New York, NY: Palgrave.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge, UK: Cambridge University Press.
- Jones, C., & Waller, D. (2015). *Corpus linguistics for grammar: A guide for research*. London, UK/New York, NY: Routledge.
- Lakoff, G., & Johnson, M. (1987). *Metaphors we live by*. Chicago: University of Chicago Press.
- Leech, G. (1997). Teaching and language corpora: A convergence. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles, (Eds.), *Teaching and language corpora* (pp. 1-23). London, UK/New York, NY: Longman.
- Lems, K. (2001). Using music in the adult ESL classroom. *ERIC digest*. Washington, DC: National Clearinghouse for ESL Literacy in Education. (ERIC Document Reproduction Service No. ED459634.)
- Lewis, M. (1997). *Implementing the lexical approach: Putting theory into practice*. Hove, UK: Language Teaching Publications.
- Littlewood, W. (1998). *Foreign and second language learning*. Cambridge, UK: Cambridge University Press.
- Murphey, T. (1992). The discourse of pop songs. *TESOL Quarterly*, 26(4), 770-774.
- O’Keeffe, A., McCarthy M., & Carter R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge, UK: Cambridge University Press.