The Use of Corpus-based Techniques in Literary Analysis: Exploring Learners' Perceptions[1]

Ridwan Wahid[2]
International Islamic University Malaysia

Abstract

Although corpus linguistics is now known to broaden our understanding of various phenomena in linguistics and literature, little research has been undertaken to examine its effects on the learning experience of learners in the literature class. This study was aimed at exploring how learners respond to the adaptation of corpus-based techniques in their literary analysis. Specifically it tried to determine (1) how the participants perceived the use of corpus-based techniques in their literary analysis and (2) if these perceptions were influenced by their preference for literature or linguistics as a field of study. Participants for the study were 39 undergraduate English majors taking a language-based computer applications course at a public university in Malaysia. Data were collected from a specially designed questionnaire and group interviews. Findings indicated that more than 50% participants perceived all three techniques favourably in terms of learning enjoyment, improvement of understanding and wider applicability. The study also found that preference for literature or linguistics had no effect on their perceptions. To further appreciate their usefulness, the next step would be to apply these techniques in actual literature courses.

Abstract in Malay

Walaupun kini diketahui bahawa linguistik korpus berupaya untuk meluaskan lagi pemahaman kita tentang pelbagai fenomena linguistik dan sastera, hanya sedikit kajian telah dilakukan untuk menganalisa kesannya ke atas pengalaman pelajar yang menggunakannya di dalam kelas sastera. Kajian ini bertujuan untuk meneroka respon pelajar terhadap adaptasi teknik berlandaskan korpus di dalam analisa sastera. Secara khususnya, ia cuba memastikan (1) bagaimana peserta-peserta menanggap penggunaan teknik berlandaskan korpus di dalam analisa sastera mereka dan (2) jika persepsi ini dipengaruhi oleh pilihan bidang pengajian mereka iaitu sastera atau linguistik. Peserta kajian ini ialah 39 mahasiswa/mahasiswi dari pengajian bahasa Inggeris yang mengambil kursus aplikasi komputer berlandaskan bahasa di sebuah universiti umum di Malaysia. Data diperolehi dari soal selidik direka khas untuk kajian ini dan temuduga secara berkumpulan. Kajian ini mendapati lebih dari 50% peserta mempunyai persepsi yang

---

[1] A version of this paper was presented at the 2005 KATE (Korean Association of Teachers of English) conference at Korea University, Seoul.

[2] Ridwan Wahid (Ph.D.) teaches linguistics in the Department of English Language and Literature, International Islamic University Malaysia. His research interests include descriptive grammar of new varieties of English and second language acquisition.

---

baik ke atas kesoronokan belajar, penambahbaikan kefahaman dan keluasan kepenggunaan. Kajian juga mendapati pilihan antara sastera atau linguistik tidak mempengaruhi perspepsi peserta. Untuk meningkatkan lagi kepenggunaan teknik-teknik ini, langkah berikutnya ialah pengaplikasian di dalam kelas sastera yang sebenar.

Keywords
Corpus-based techniques, literary analysis, learners' perceptions, literature learning, corpus stylistics, CALL (computer assisted language learning)

Keywords in Malay
Teknik berlandaskan korpus, analisa sastera, persepsi pelajar, pembelajaran sastera, stilistik korpus, PBBK (pembelajaran bahasa berbantukan komputer)

1. Background
Corpus linguistics is currently enlarging our understanding of various phenomena in linguistics and literature. By way of computer automation, patterns of usually large numbers of words are analysed with increased efficiency and accuracy. Corpus-based methods have also been adapted both directly and indirectly in language teaching (Tribble and Jones; Leech; Römer; O'Keefe et al.). Data-driven learning (Johns), which requires learners to utilise corpora to make discoveries about language by themselves, is an example of a direct application of corpora. The use of corpus data in lexicography, which results in more accurate dictionaries to be used by learners in the classroom, is an example of an indirect application.

The use of corpus linguistics, or in particular its research techniques, has also been extended to literary analysis. These techniques, either quantitative or qualitative or both, can result in very detailed stylistic descriptions of literary texts that can either complement conventional interpretations or enable insights not possible through intuition alone (Adolphs 64; Stubbs, *Conrad* 22). Using such techniques, researchers have examined, among others, authorship (Holmes and Forsyth), rhythm (Roberts), generic structure (Mealand) and word clusters as a stylistic device (Rommel). Despite these great advances, one needs to note that efforts like these are made for research purposes and have not, in general, been adapted for literary pedagogy.

Among a growing number of teachers, literature is seen as an essential component of language education. Despite this healthy trend, there exists a gap between the view held by those teachers and their actual teaching practice (Paran 10). It is therefore imperative that greater efforts are made to ensure that literature instruction is effective for students at all levels (Paran 10). This, and the fact that higher-level criticisms can often benefit from corpus-based perspectives (Louw 250), should justify the use of corpus linguistics in the study of literature by university undergraduates. This argument is also corroborated by

the need to ease difficulties faced by many tertiary-level ESL learners when they are engaged in literary analysis (e.g. Nora Nasir; Butler).

At this juncture, it is worthwhile to bear in mind that while helping students to improve their way of analysing literary texts is a valid pedagogical agenda, a wholesale import of corpus linguistics into the literature class is not necessarily the right action to take. Studies in computer-assisted language teaching (CALL) frequently report various kinds of tension between learner traits and the inherent characteristics of the newly introduced technology (Salaberry; Scholfield and Ypsilandis). Though such problems must not deter teachers from using technology, it does imply that for any adaptation of technology to be successful, proper efforts must be made to better understand how users can best benefit from it.

The few existing studies that are found to demonstrate possible benefits of corpus-based techniques were mostly approached from the perspective of the teachers. Ma, for example, looked at the learning strategies of learners using a concordancer in learning to write in an ESP class. Although he concluded that the use of the concordancer was useful in some ways to the students, many of them also admitted that they found certain difficulties in exploiting the technology. Nevertheless, how these learners perceived these difficulties as part of their overall learning experience was not explored. Nuraihan Mat Daud and Zamnah Husin also found that the use of the concordancer was useful to their learners, specifically in improving critical thinking skills in an extensive reading class. Their quantitative study indicated that exposure to and use of collocation analysis could lead to more effective problem solving strategies. The study, however, presented a limited assessment of the learners' own experiences in using the concordancer and it is not known if the learners themselves had perceived a significant learning effect on their part. One study that has been approached from the perspective of the learners is by Abusa', in which he found the use of the concordancer and its printed output as an instructional material in his class generally well received by the learners. This study, however, only investigated the teaching of grammar. Finally, a study by Vethamani is an investigation of learners' perceptions of the adaptation of computer technology in literature learning. Nevertheless, it did not involve corpus-related instruction as it focused on how the participants responded to a few web-based presentations of their analyses. This review thus points to a need to explore how learners respond to the use of techniques derived from corpus linguistics in analysing literary texts.

The present study aims to determine (1) how the participants perceive the use of corpus-based techniques in their literary analyses and (2) if these perceptions are influenced by their preference for literature or linguistics as a field of study. Following Holahan (qtd. in Barkhuizen 90), *perception* is defined as a "process of apprehending through sensory input" the use of corpus-based

techniques in the participants' literary analyses. One assumption of the study is that if these techniques are positively viewed by the learners, they will continue to use the former in their future literary analyses.

2. Research Design
2.1 Corpus-based Techniques
Three corpus-based techniques were investigated in this study. These are by no means the only techniques that can be used for analysing literary texts. They were selected for the study because (1) a number of published studies had made use of them, (2) they were techniques that were commonly described in introductory textbooks on corpus linguistics and (3) they could be demonstrated to the students using the resources available to the researcher. In the course taken by the students, they were identified as:

   (i) word frequency analysis
   (ii) word collocation
   (iii) part-of-speech analysis.

"Word frequency analysis" is a technique that requires analysts to examine how frequent or infrequent a particular word (or a phrase) appears in a corpus (or text) (O'Keefe et al. 11-12). "Word collocation" is an investigation of the tendency of two or more words to co-occur in a text (Stubbs, *Words and Phrases* 29-30). "Part-of-speech analysis" requires analysts to examine the frequency and/or the collocates of a given lexical category (or a combination of categories) (McEnery and Wilson 135).

   In a word frequency analysis, one is given information about the number times each word is used in the text. The list in figure 1, containing the top 20 words in Lewis Carroll's *Alice's Adventures in Wonderland* (henceforth *Wonderland*), is an example.[3]

Figure 1: Top 20 words in Wonderland

N	Word	Frequency
1	the	1,635
2	and	875
3	to	732
4	a	628
5	it	595
6	she	553
7	I	544
8	of	513
9	said	462

---

[3] All examples are not given as full lists due to space constraints.

10	Alice	415
11	you	412
12	in	373
13	was	357
14	that	315
15	as	265
16	her	249
17	at	214
18	on	195
19	with	186
20	all	182

This list, however, is not very informative for students in that it does not give them much information about the story. One reason is that it contains a large number of function words (e.g. "the," "and," "to," "a") which, despite their necessity for structural purposes, do not contain much content (Stubbs, *Words and Phrases* 40). If these words are removed so that only the content words remain, a different picture begins to emerge. The new list is shown in figure 2.

Figure 2: Top 20 content words in Wonderland

N	Word	Frequency
1	said	462
2	Alice	415
3	little	127
4	out	113
5	down	101
6	up	98
7	know	88
8	went	83
9	queen	75
10	thought	74
11	time	68
12	see	67
13	king	66
14	turtle	60
15	hatter	59
16	began	58
17	gryphon	58
18	mock	57
19	head	53
20	think	53

From this list one can identify some of the central characters in the novel (through such nouns as "Alice," "queen," "turtle," etc.) and many other words

that are probably used to describe them. Understandably, at this stage any information that one gets by studying the frequency list before reading the text is tentative. However, teachers can turn this uncertainty to their advantage. A learner, for example, can be asked to study the frequencies of these words and conjecture about the plot of the story. This may serve as a useful pre-reading activity (Römer 95-96).

The word frequency analysis can also suggest aspects of the story or text that can be further investigated. In their analysis of *Wonderland* and its sequel *Through the Looking Glass and What Alice Found There* (henceforth *Looking Glass*), Inaki and Okita analysed the "verbs of saying" and their contents used by Alice to confirm their hypothesis that the protagonist had different statuses in the two novels. This analysis was prompted by their observation that the verb "said" was almost equally highly used in both books, suggesting that Alice's conversations (as opposed to, e.g., descriptions of her actions) in each story were significant (286-290). Indeed, they found that Alice tended to be "passive" in the first book but "active" in its sequel (291).

The word collocation analysis is demonstrated by a concordance of "Alice" from *Wonderland* in figure 3 and a concordance of "the Queen" from the same book in figure 4 below. The concordances were created by sorting the lines according to the first word to the right of "Alice" and "the Queen" in alphabetical order (symbolised by "N + 1" in the figures). Due to space considerations only the first 25 lines are shown. Notice that "Alice" tends to collocate, or co-occur, with the verb "said" more than "the Queen" does. Although "the Queen" is also found to collocate with "said," the association is not as strong. Firstly, based on its left collocates, the queen also "shout[s]," "shriek[s]" as well as "roars." Secondly, looking at the right collocates, when the Queen indeed says something, she sometimes does so "furiously" and "in a shrill, loud voice." The concordances in figures 3 and 4 can be used to reinforce the analysis of these characters. The lines can be sorted in ascending or descending alphabetical order according to any nearby word to the left or right of the search word. Rearranging the concordance lines in different assortments can uncover new ways of reading the patterns (Römer 12).

Figure 3: The first 25 lines in the concordance of "Alice" from Wonderland*, N + 1 sorted*

N Concordance

```
1        top of his shrill little voice, the name `Alice!'
2                  from this morning,' said Alice a little timidly: `but it's no use
3          `That was a narrow escape!' said Alice, a good deal frightened at the
4      Alice `Please would you tell me,' said Alice, a little timidly, for she was
5    its great eyes half shut. This seemed to Alice a good opportunity for making her
6    ought to be ashamed of yourself,' said Alice, `a great girl like you,' (she might
7        bowed low. `Would you tell me,' said Alice, a little timidly, `why you are
8      the Cat. `--so long as I get somewhere,' Alice added as an explanation. `Oh,
9        watch In another moment down went Alice after it, never once considering how
10   happened. `How am I to get in?' asked Alice again, in a louder tone. `Are you to
11    riddle yet?' the Hatter said, turning to Alice again. `No, I give it up,' Alice
12        mice--oh, I beg your pardon!' cried Alice again, for this time the Mouse was
13      your feelings may be different,' said Alice; `all I know is, it would feel very
14   a friend of mine--a Cheshire Cat,' said Alice: `allow me to introduce it.' `I don't
15    ordered'; and she walked off, leaving Alice alone with the Gryphon. Alice did
16   I had our Dinah here, I know I do!' said Alice aloud, addressing nobody in
17   sharply and very angrily. `A knot!' said Alice, always ready to make herself
18     another hedgehog, which seemed to Alice an excellent opportunity for
19        `Come, it's pleased so far,' thought Alice, and she went on. `Would you tell
20      the setting sun, and thinking of little Alice and all her wonderful Adventures,
21   much under the sea--' (`I haven't,' said Alice)-- `and perhaps you were never
22      to day.' This was quite a new idea to Alice, and she thought it over a little
23        `How dreadfully savage!' exclaimed Alice. `And ever since that,' the Hatter
24   a sad tale!' said the Mouse, turning to Alice, and sighing. `It is a long tail,
25     on with the game,' the Queen said to Alice; and Alice was too much frightened
```

Figure 4: The first 25 lines in the concordance of "the Queen" from Wonderland*, N + 1 sorted*

N Concordance

```
1        `--and just take his head off outside,' the Queen added to one of the officers:
2      the order of the words a little, `From the Queen. An invitation for the Duchess
3      at the picture.) `Up, lazy thing!' said the Queen, `and take this young lady to
4    one,' said Alice. `Come on, then,' said the Queen, `and he shall tell you his
5     and all the players, except the King, the Queen, and Alice, were in custody
6    go and get ready to play croquet with the Queen,' and she hurried out of the
7      foot to the other, looking uneasily at the Queen, and in his confusion he bit a
8        the garden, called out `The Queen! The Queen!' and the three gardeners
9    shouted Alice. `Come on, then!' roared the Queen, and Alice joined the
10     had not as yet had any dispute with the Queen, but she knew that it might
11   shouted in reply. `That's right!' shouted the Queen. `Can you play croquet?' The
12    and Alice looked round, eager to see the Queen. First came ten soldiers
13      he said to the Queen. `Never!' said the Queen furiously, throwing an
14   I wish you would have this cat removed!' The Queen had only one way of settling
15   the other side of the court. All this time the Queen had never left off staring at the
16   unhappy at the number of executions the Queen had ordered. Gryphon asleep
17      Mock Turtle Soup is made from,' said the Queen. `I never saw one, or heard of
18   just been picked up.' `What's in it?' said the Queen. `I haven't opened it yet,' said
19    carefully, with one foot. `Get up!' said the Queen, in a shrill, loud voice, and the
20        Alice looked up, and there stood the Queen in front of them, with her arms
21      said--' `Get to your places!' shouted the Queen in a voice of thunder, and
22   that day. `That proves his guilt,' said the Queen. `It proves nothing of the sort!'
23   the first verse,' said the Hatter, `when the Queen jumped up and bawled out,
24   and under sentence of execution. Then the Queen left off, quite out of breath,
25      her, they hurried back to the game, the Queen merely remarking that a
```

The part-of-speech analysis usually requires additional software where each word is "tagged" or assigned a code indicating its lexical class. In the course, the students were asked to use the "demo" version of CLAWS 4,[4] a piece of software that tags each word according to its own set of codes.[5] To illustrate the utility of the part-of-speech analysis, let's consider the study by Inaki and Okita again.

One of the findings from their initial investigation is that Alice in *Wonderland* is more passive than she is in *Looking Glass* (286). To provide further support to this claim, they generated concordances for (a) all adverbials modifying verbs of saying uttered by Alice and (b) all modifiers of "Alice" in both novels. Although (a) and (b) were accomplished by Inaki and Okita using the word collocation analysis (see above), the process was a lengthy one. For example, to carry out (b) one has to "first… search *Alice*, *girl*, *little thing*, and *child* in the corpus and then prune lists of the occurrences of given words to include only the case of Alice, and then furthermore prune to concordance lines, which include modifiers" (291). With a tagged corpus, on the other hand, most of these steps can be skipped by querying the corpus for adjectives that collocate with *Alice/girl/little thing/child*. Using the chosen tagset (see footnote 5), the symbols to be searched for are "AJ0" (unmarked adjective), "AJC" (comparative adjective) and "AJS" (superlative adjective). This query would retrieve all (adjectival) modifiers for "Alice." Figure 5 shows the first ten lines of the result of this search in *Wonderland*.

Figure 5: The first unsorted 10 lines in the concordances of   "AJ0/AJC/AJS   +
Alice/girl/little thing/child" in Wonderland

```
N Concordance
1    what_DTQ an_AT0 ignorant_AJ0 little_AJ0 girl_NN1 she_PNP 'll_VM0
2              ,_, '_POS thought_NN1 poor_AJ0 Alice_NP0 ,_, `_" it_PNP
3         but_CJC the_AT0 wise_AJ0 little_AJ0 Alice_NP0 was_VBD not_XX0
4    ;_; but_CJC ,_, alas_ITJ for_PRP poor_AJ0 Alice_NP0 !_! when_CJS she_PNP
5            ,_, for_PRP this_DT0 curious_AJ0 child_NN1 was_VBD very_AV0
6    now_AV0 ,_, '_POS thought_NN1 poor_AJ0 Alice_NP0 ,_, `_" to_TO0
7         garden_NN1 door_NN1 ._. Poor_AJ0 Alice_NP0 !_! It_PNP was_VBD
8          Alice_NP0 ,_, `_" a_AT0 great_AJ0 girl_NN1 like_PRP you_PNP ,_,
9    he_PNP could_VM0 go_VVI ._. Giant_AJ0 Alice_NP0 watching_VVG
10   words_NN2 ,_, '_POS said_VVD poor_AJ0 Alice_NP0 ,_, and_CJC her_DPS
```

---

[4] CLAWS stands for "Constituent Likelihood Automatic Word-tagging System," developed at Lancaster University, England.

[5] The codes used in this study were taken from a collection or "tagset" called CLAWS 5.

Using both qualitative information (i.e. the *types* of adjectives that collocate with "Alice") and quantitative information (i.e. the *frequency of tokens* of each of those adjectives) in comparing the data from the two novels, Inaki and Okita were able to "clarify the distinctive positions of Alice" (292). For example, they claimed that:

> In *Looking-Glass*, Alice goes to the looking-glass world intentionally and actively, which signifies Alice's active attitude to the adventures and her situation in the story. Another is the use of some modifiers which do not occur in *Wonderland*, such as 'thoughtful' and 'reasonable'. These imply positive assessment of Alice considering her age and then enhance some of Alice's roles in *Looking-Glass*. (292)

The descriptions of the analyses in this section are necessarily brief and only intended to illustrate how the corpus-based techniques were taught and subsequently used by the participants.

2.2 Data Collection
Data for analysis were collected through the use of a specially designed questionnaire and group interviews.

2.2.1 Participants
The participants in this study were 39 final-year undergraduates (34 females, five males) in the English Language and Literature programme at a public university in Malaysia. 35 of these learners were in their final semester. 36 of them were Malaysians and three were international students. While all of them were ESL learners, the time spent on learning English differed greatly between these two groups. Whereas the average time spent on learning English for the Malaysians was 17 years, the length for the non-Malaysians ranged from four to 13 years. All of them had studied English literature for an average of five years. Enrolled in an introductory course called "Computer Applications in Language Studies" (CALS), the participants learned how to use the MonoConc Pro concordancer and a number of other concordancers that were web-based. They also learned how to make use of information provided by word frequencies and collocations (see previous section). Other than that, they learned how to "tag" a text for parts of speech using several "demo" versions of tagging software available on the internet (see previous section). As a partial requirement of this course, in groups of five the participants were asked to review a published computer-based literary analysis and complete two computer-based literary analyses. The first of these analyses required them to conduct a word frequency analysis on Ernest Hemingway's "Hills like White Elephants," a text that was selected due to (1) its availability in the electronic form on the internet and (2) the writer's status as a major literary figure in American literature, which was also a separate, full length

course for these students. The second was an analysis of a literary text or texts of their choice utilising corpus-based techniques that they thought would best answer their research questions.

2.2.2 Questionnaire

At the end of the course, the participants were asked to complete a questionnaire. This questionnaire began by asking for information about the participants' learning background of language and literature. Next it asked them to consider three types of perception that were adapted from Barkhuizen (107): (1) expressing a feeling about a corpus-based technique used in the literary analysis (resulting in a question that basically means "Did you like using this particular technique in studying literature?"), (2) making a judgement about the use of this technique ("Did this technique help you to understand a literary text better?") and (3) making a prediction about how useful the technique would be beyond the context of the course ("Do you think the technique will be useful for literary analyses outside this class?"). For convenience, the rest of the paper will refer to Questions 1, 2 and 3 as "Enjoyment," "Understanding Literature" and "Beyond CALS" respectively (see Appendix for a sample question).

Following Barkhuizen (92), responses to these questions were analysed in terms of positive perception percentages. The positive perception percentage was created by combining responses indicating *definitely* and *quite a bit* on the agreement scale in the questionnaire. Responses indicating *just a bit* and *not at all* were collapsed and considered negative. Based on these percentages the three techniques were rank-ordered for each type of perception. The participants were also invited to give comments for each technique.

2.2.3 Group Interviews

Two semi-structured group interviews were subsequently carried out to explore these perceptions further. The first group consisted of five participants while the second group consisted of six. Some questions included:

> (1) Did you find the word frequency/word collocation/part-of-speech analysis useful? Why?
> (2) Did you have any problems in using it? What were they?
> (3) The word collocation analysis was ranked highest in the survey for Enjoyment and Beyond CALS but lowest for Understanding Literature – do you have any comment on that?

The interviews were tape-recorded and transcribed. Using the Nud*ist software, analysis was made for recurrent themes in the learners' responses. Interviews with the two groups are identified below as "Interview 1" and "Interview 2" respectively.

3. Findings and Discussion
3.1 Perceptions of the Techniques
Table 1 shows the ranking of the techniques according to the three types of perception. As can be seen from the table, the positive perception percentages are very high, with all but one nearing or exceeding 75%. This means that approximately three-quarter of the participants perceived each technique positively in terms of their enjoyment of use, understanding of the subject and prediction of its usefulness beyond the course. Interestingly, the rankings for Enjoyment and Beyond CALS are nearly identical, while the ranking for Understanding Literature is generally a reverse pattern of the other two. The one technique that scored far below 75% is "word collocation", under the category of Enjoyment (66%).

The study was originally intended to investigate learners' perceptions of each technique in detail. However findings from the questionnaire and group interviews were quite limited in this sense. In the former, the majority of the participants chose not to comment on each technique while in the latter, when prompted to comment on individual techniques, the participants indicated that they felt that their responses were true for all techniques. The almost homogenous percentages for all techniques in all types of perceptions parallel the participants' invariable responses. As a result the following discussion in this section presents the learners' perceptions of all the three techniques in terms of Enjoyment, Understanding Literature and Beyond CALS.

Table 1: Rank Order of Positive Perception Percentages for Each Technique

Technique	Enjoyment		Understanding Literature		Beyond CALS	
	Rank	%	Rank	%	Rank	%
1. Word frequency	1	76	3	74	1	79
2. Part-of-speech	1	76	2	82	2	77
3. Word collocation	3	66	1	84	3	72

3.1.1 Enjoyment
Table 1 indicates that "word frequency" and "part-of-speech analysis" each had the highest positive percentage for enjoyment at 76%. The participants'

comments on the questionnaires reveal that the novelty of these techniques was often what they found most enjoyable:

- Learning literature the traditional way can be boring and dull. We need something different. This is a good way to do it.
- It's unconventional but it's interesting.
- Something new in the linguistics course.
- This is in fact such an interesting way to understand more about literature compared to the traditional way.
- This is very fun.

Other participants may be more tentative in expressing their enjoyment as these comments in the questionnaires reveal:

- It should be fun to use this (word frequency) but it also depends on my mood.
- It's okay at times, but it's not always that I like it (part-of-speech analysis).

In one interview, two participants revealed that the slightly unfavourable impression given by some of the learners was due to the manual nature of certain frequency counts involved in "word frequency" and "part-of-speech analysis":

> Participant 6: There's just more work to this frequency thing.
> Participant 7: It's like doing grammatical analysis in another class of ours.
> Participant 6: Exactly. (Interview 2)

There were two incidents that happened in class that led them to give these comments. First, one of the web-based concordancers used by some of them did not feature a word count facility. Thus they had to count the occurrences of a word in a concordance by themselves. Second, at the time of the analysis they were describing, the part-of-speech tagger was not available to the learners due to technical difficulties (although they were able to use the tagger a week before). As a result they also had to identify the parts of speech manually. The tasks of identifying and counting words and/or parts of speech through manual labour proved to be very unappealing to many of the participants.[6] Despite the above comments, the high positive perception percentages for "word frequency" and "part-of-speech analysis" still suggest that

---

[6] Although one may argue that carrying out these tasks manually would defeat the purpose of introducing corpus linguistic techniques to the students, it succeeded in demonstrating to them how computer technology had enhanced certain ideas in early corpus linguistics research.

the majority of the learners enjoyed their learning experience with these techniques.

"Word collocation," on the other hand, scored relatively low for Enjoyment at 66%. Two comments were given for this technique:

- It (word collocation) doesn't help a lot.
- I don't quite get the significance of using it.

Findings from the interview suggest that the lack of enjoyment in the use of this technique may have to do with its difficulty. This will be elaborated towards the end of the next section.

3.1.2 Understanding Literature

Table 1 shows that, of the three types of perception examined in the study, the techniques were most highly rated in the category of Understanding Literature. "Word collocation" was positively perceived by 84% of the participants and "part-of-speech" by 82% of them. "Word frequency" received the lowest percentage at 74% but even this figure suggests that nearly three quarters of participants perceived it positively as well. Findings from the interviews show that the participants felt the three techniques had helped them to understand literature better in a variety of ways.

Firstly, in common with the usual claims made by researchers using the techniques in literary analyses for the first time (e.g. "Stanford Students Use Digital Tools"), several participants mentioned the new perspective through which they came to see the literary text, as the following interview excerpts exemplify:

> Interviewer: Were there a lot of new things that you learned in this course?
> Participant 6: Oh yes, definitely. We've learned a lot of new things.
> Participant 7: I feel we've learned a lot, like, I didn't know the computer can, like, analyse some texts. In my group presentation (review of published study), we did on John Milton, I think, and, and using the technique is actually like another way of looking at the poem, a whole new way of doing it.
> Participant 6: Yeah, that is very useful. (Interview 2)

For some others, this perspective was not entirely new but made clearer by corpus linguistics:

> Participant 1: Some writings have their own style. If we look at the genre of literature, it's like, there are some words that are used by the writers – nouns, pronouns – in horror, romance and so on. So, using these techniques…

> Participant 2: … make us understand better.
> Interviewer: Were you already aware of the different kinds of words in different kinds of writing?
> Participant: Yes.
> Interviewer: Ok. But would you make use of such knowledge if this course didn't ask you to?
> Participant 1: Well, probably not.
> Participant 2: You know, this is a lot like "Linguistic Approaches to Literature" (the title of another course)
> Participant 3: Ah, yes!
> Participant 4: We already did something like this but not as detailed as this course.   (Interview 1)

As Participant 4 describes it, they knew a literary text could be analysed linguistically – a fact they learned from a different course – but they never attempted to do so on their own.

Secondly, the participants clearly appreciated the efficiency of the computer in carrying out the analytical tasks. Several of them mentioned this in the interviews. The following are two examples:

> Participant 3: With the computer, it's faster. With the concordancer and all, everything is quick. You just click, click, click. In a conventional way, if you were to do this, you'd have a lot of paper, written here and there with charts and diagrams everywhere. That is, if you want to do some comparison between words. (Interview 1)
> Participant 9: If you use the computer, it can give you many details because of its word analysis. So, you know, it could help you. Because normally, when you read, there's a huge amount of words, vocabulary. You cannot go to these words one by one, or to see which words have more frequency in a text. Only by a computer, then you can know exactly. Then you will start to think whether this word is significant, you know, maybe to contribute to the character or the theme. (Interview 2)

Thirdly, a few participants felt that the corpus-based techniques had helped to make better sense of their analysis. The participants who mentioned this benefit pointed to the fact that they may already have a vague idea about something but it was the computer analysis that provided them with the evidence.

> Participant 3: The corpus shows what I already have in mind, sometimes. But it makes it clearer. Sometimes you have a hard time describing something but the corpus helps you to identify certain words and make it clearer. (Interview 1)
> Participant 9: You know, in a conventional analysis you might miss some tiny details. You may already know what you think about the novel, or

what you think about the characters. But because words are so many, you can miss the important ones. But by the computer, these words can be noticed.
Interviewer: How?
Participant 9: When they come up with other words that you look at.
(Interview 2)

It was shown earlier that the novelty of the technology had led many participants to enjoy the use of the techniques in their literary analyses. Where Understanding Literature was concerned, however, the new technology was found to give a number of negative effects as well. Although the positive perception percentages for this category (see table 1) show that the participants in general agreed that they had learned literature better by using the techniques, some of them pointed out that they only did so with initial difficulties. The following comments from the questionnaires are shown as examples:

- At first, it seems quite difficult to understand the relevance. But later on, I could see the significance.
- I didn't have any idea at all what I was studying at first. Later, gradually I started to understand what was being taught.
- I think the techniques may help a bit, but if we are not familiar with it, it would be troublesome to get the analysis done.
- At first I found it quite boring, but as time went by, I began to see the importance and benefits of it.

A similar response was also found in one of the interviews.

Participant 7: It was difficult at first.
Interviewer: How was it difficult at first?
Participant 7: I think because it was technical and because there was a lot of new things. (Interview 2)

For one participant, some of these difficulties were never overcome as her comment in the questionnaire shows:

- At this moment, perhaps corpus-based study in literature is very new for me, and I think it's supportive in certain ways only.

Her feelings are reflected in a response given by a participant in one of the interviews:

Interviewer: Did you get anything out of the analysis?
Participant 11: No, not really, actually. To me, this is a whole new thing to analyse literature, but no…. (Interview 2)

Recall that "word collocation," at 66%, was not as highly rated as the other techniques for Enjoyment. It is possible that "word collocation" was difficult due to its unfamiliarity and the learners were still trying to come to terms with the new technique. Thus it was rated relatively low in this category. A piece of evidence is offered in Participant 9's description of her feelings about analysing Joseph Conrad's *Heart of Darkness* using the computer in which she describes her "open" position as a researcher as well as her difficulties in "trusting" the patterns of words she found. In doing so, the participant gives an account of how she became ambivalent about using the technique.

> Interviewer: We also analysed Conrad's *Heart of Darkness* using corpus-based techniques such as "word collocation." Was that difficult?
> Participant 9: I liked that, but the effect, I didn't feel that the effect, the effect was obvious from the analysis, you know.
> Interviewer: Were you able to get a clear picture?
> Participant 9: No, not really.
> Interviewer: So, was it difficult for you then?
> Participant 9: Er, I mean I tried, but it wasn't easy.
> Interviewer: What did you try to do actually?
> Participant 9: I wanted to get something like, maybe, something I never felt before. Something is there, but you're not so sure whether it can, or it is very trustworthy, you know, from the patterns that I see. (Interview 2)

However, in other cases, "word collocation" certainly made an impact on the learners as it was rated most highly for Understanding Literature (84%).

Participant 9's difficulty in "trusting" the patterns that she saw in the concordances is echoed in the other participants' concerns regarding the role of the computer in helping them (re)construct a meaning. In the interviews, some participants suggested that their literature instructors viewed "meaning" in a text differently from one another. The following excerpt illustrates:

> Participant 6: In "Hills," I know it's about abortion, but you can also see it as premarital sex, or we can see it as a lot of different things. Like, you can say drug abuse and so on.
> Interviewer: Did your teacher allow this?
> Participant 7: Yes. She says that it's good that we think of different ideas.
> Interviewer: All your literature teachers are like this?
> Participant 7: Of course not!
> Interviewer: So?
> Participant 7: There's only one way that they (some other teachers) accept.
> (Interview 2)

The different views held by the participants' teachers were concerned with the meaning or meanings that a text may embody. It is a reflection of the conflict

between two traditions of literary studies, New Criticism and transactional theory (Addington; Gilroy and Parkinson). Addington points out that this conflict is typical of larger literary pedagogical settings (such as the English department in which this study took place) (219). New Criticism proponents (e.g. Brooks and Warren) believe that meaning should be viewed from inside the text with no or little external elements brought into the reading. In contrast, transactional theorists (e.g. Rosenblatt) propose that meaning is what the reader brings to the text and is essentially an interaction between the two. The different views of meaning that the learners were asked to see in their different literature courses were clearly oriented towards either one of the two traditions. In this regard, it was found that the learners felt that the use of corpus-based techniques was more suitable when meaning was seen as "fixed" as compared to "flexible":

> Interviewer: Some of you talked about feelings in such classes (where meaning is not fixed). How important are feelings to a corpus-based analysis? Do they play a role at all?
> Participant 5: Not really.
> Participant 4: They do, but they are more objective.
> Participant 5: It's more difficult to do that.
> Interviewer: Do you think you would use these techniques in such classes?
> Participant 5: Hmmm. Maybe not.
> Participant 4: No.
> Interviewer: In the other classes (where meaning is fixed)?
> Participant 4: Come to think of it, yeah, maybe. (Interview 1)

This is an excerpt from the other interview:

> Interviewer: How does the computer help you in such a situation (where meaning is not fixed)?
> Participant 11: Okay, when I was reading the concordances, it opens a lot of possibilities. Maybe I'll analyse the verbs or the nouns, what kind of nouns are used and so on. So from there, you can analyse.
> Interviewer: Any meaning is possible in this interpretation?
> Participant 11: Yes.
> Interviewer: Does anyone else feel the same way?
> Participant 7: I don't.
> Participant 8: I don't.
> Interviewer: Why?
> Participant 7: It's just difficult for me. I see, perhaps, less meaning through that.
> Participant 8: When the words are there, it's hard to say some other thing.
>
> (Interview 2)

The learners' views are interesting because the dichotomy between reader response theory and New Criticism in their teachers' instruction appears to have divided their perceptions of the use of corpus-based techniques as well. In advocating the use of such techniques, Stubbs (*Conrad* 6) and Inaki and Okita (293) clearly call for an integration of intuition and computer technology in the analysis. Nevertheless, it was not easy for the novices in this study to find the right balance between the two. In the interviews many of them relied more heavily on existing knowledge than they did on what they saw (or did not see) in the concordances although, in fact, both sources of knowledge could be used to support each other. This finding ties in with the above argument that some of them were still trying to get used to the use of corpus-based techniques in their analyses. The difficulties faced by the learners in reconciling intuition and corpus-based technology will be discussed again in section 4 below.

3.1.3 Beyond CALS

With all techniques getting more than 70%, table 1 shows that all three were perceived positively by more than half of the participants for Beyond CALS. Given the many benefits that they themselves identified, it is not surprising that the participants did not give much elaboration on this category in their interview responses. When asked if they would use the techniques for future literary analyses, many of the participants answered in the affirmative. Further, when prompted to explain why, they pointed to the earlier responses that they made. The excerpt below serves as an example:

> Interviewer: Now that we've discussed what the computer can or cannot do, do you think you will continue these techniques for, say, the literature courses you take next semester?
> Participant 3: Yes.
> Participant 2: Yes.
> Participant 4: Yes.
> Interviewer: Why?
> Participant 4: For all the things we mentioned just now. (Interview 1)

In Interview 2, however, Participant 10 indicated that he would be more willing to use them for research in linguistics. When asked about his preference for a study field, he chose linguistics.

> Interviewer: Do you think what you learned here will be useful for future literary analyses?
> Participant 10: For literature, no. For linguistics, yes.
> Interviewer: Do you prefer linguistics as a field of study, actually?
> Participant 10: Yes. (Interview 2)

At the beginning of the course, a number of learners voluntarily declared their preference for either literature or linguistics as a field of study when informed that the corpus-based techniques could be used for research in both. Their declaration may have to do with a desire to inform the instructor of the kind of research projects they would prefer to do in the course. Although the learners were made to undertake research both in literature and linguistics (not discussed in this paper), preference for either one of these study fields may have influenced their perceptions of each project, as Participant 10's response appears to be suggesting. This possibility is explored through the study's second objective. The findings are presented next.

3.1.4 Students' Field of Study Preferences and their Perceptions

To find out if there was a possible interaction between preference for literature or linguistics and a learner's decision to use the corpus-based techniques in literary analyses, a Pearson chi-square test was conducted on these variables. Three participants indicated that they had no preference and these were excluded (N = 36).

It was found that in all cases, there was no significant relationship between preference for linguistics or literature and their decision to use the techniques in the future. The decision to use the essentially quantitative technique "word frequency" ($\chi2$ (1) = 0.296, $p < .05$) was not related to the participants' preference for literature or linguistics. This was also true for the largely qualitative technique, "word collocation" ($\chi2$ (1) = 0.590, $p < .05$) as well as the technique that can straddle both paradigms, "part-of-speech frequency" ($\chi2$ (1) = 0.177, $p < .05$). This also means that, despite the obvious linguistic orientation of some of the ways the techniques were used, learners who indicated that they did not prefer to study linguistics found the use of corpus linguistics in analysing literature favourable. The fact that literature-minded students are not necessarily averse to corpus-based techniques should thus encourage teachers to adopt them in their instruction.

4. Broader Issues

This section discusses other problems that were identified through the findings although not from the viewpoint of the learners. First, in the interviews it was found that many of the difficulties that were mentioned by the participants were, directly or otherwise, the result of a lack of proper context for their research skills to develop. When discussing their project on "Hills like White Elephants," the interview findings reveal they were not always clear on what they were using the techniques for.

> Participant 10: Well, when we did "Hills," I did some research on Hemingway and found that at one point, he said we must remove some

of the important words in the text, or, or, we should know where to stop. This means that, in his story, he did not mention the word 'abortion'. Although the story…

Participant 6: is about abortion.

Participant 10: is about abortion. So if you use this technique without, like, myself, before this I've never read the story, so I didn't know the story is about abortion. But if you start using this technique, you start analysing, you will get nothing…

Interviewer: Nothing?

Participant 10: Nothing about abortion, yeah. (Interview 2)

For some others, the use of the techniques was seen as a means to simplify their research. The following is a continuation of the above discussion.

Interviewer: Who feels the same way?

Participant 9: I think it is the same whether you're using corpus linguistics or the conventional way. It's the same. You have to know the background. It will help you to analyse. If you don't know the background, even when you use the conventional analysis, you won't know anything about it.

Participant 6: Yes. You have to know the background.

Participant 7: Then maybe we should start with the conventional analysis and maybe after that we can start using corpus linguistics.

Participant 8: But I think, er, I agree with you, but what we did, the word frequency and everything, it actually, sort of, narrows the analysis.

Participant 6: Yes, narrows it.

Participant 7: So we can actually see; ok, at least we know a few things.

(Interview 2)

Despite the enthusiasm that was shown in the arguments for the use of corpus linguistic techniques, it was evident in this discussion that their research objectives were not properly formed. The learners were in fact reminded repeatedly in class that a good research question is crucial for an analysis to be meaningful. In this case while it may be easy to attribute their ignorance of such an important reminder to carelessness, it is felt that the context of the course that did not allow a more elaborate treatment of the text would be a more pertinent factor. With an essentially non-literature oriented syllabus and severe time constraints, there were few opportunities to explore other important aspects of the story, e.g. Hemingway's style of writing and the broader context of the issues found in the story. The fact that the word frequency analysis could also highlight words that are rarely or never used (provided that one has a good understanding of the story) had escaped many of the participants. Left to their own devices, some of the learners were confused over how corpus-based techniques should be used in literary analysis. This confusion may also further

explain the difficulties that they faced in using word collocation described in section 3.1.2 above.

From the perspective of the teacher, it was also found that the lack of suitable software had limited the learners' analysis. This is especially true in the case of parts of speech. Although they were introduced to the concept of tagging and could often appreciate its usefulness, the use of "demo" versions of such software was obviously limiting in many ways. It was observed that in their second projects many of the learners did not make use of the part-of-speech frequency technique. For the same reason, none of them was found to venture into investigations of grammatical patterns that could be uncovered in their chosen text.

4. Conclusions and Recommendation

To conclude, the use of corpus-based techniques in literary analysis was found to be very positively perceived by the learners. Overall, it was found that the majority of the participants believed that the techniques were enjoyable to use, helpful in understanding literature and useful for future literary analyses. It was also found that learners who preferred to study literature to linguistics had no problems in adapting these techniques in their analyses.

Although the learners' perceptions were mostly positive, some experienced difficulties in using the techniques. Some of these difficulties were overcome while some others were not. Interestingly, some participants indicated that the techniques were more suitably used when meaning was seen as "fixed." In addition, many of the difficulties were found to be the result of lack of proper context for the learners' research skills in corpus linguistics to develop.

It is recommended that the use of these techniques should be adopted in actual literature courses so that learners can fully appreciate their usefulness. In a single course that aims to introduce theoretical groundings as well as practical skills for both linguistics and literature such as CALS, coming to a full appreciation of what corpus-based techniques can do to the study of literature has not proved to be easy. This can be seen in the lack of context for the learners' research skills to develop and the confusion that it has resulted in. Perhaps with exposure to the theoretical background to corpus linguistics in a course such as CALS and simultaneous practical training provided by "proper" literature courses, the experience of using corpus linguistics in the study of literature will be far more rewarding to both learners and teachers.

Another reason why corpus linguistic techniques should be used in such courses is that the novelty of a technology like this will diminish with time (Salaberry 52). It is not impossible that many of the enthusiastic remarks made here by the participants will no longer be true for them in the future. In an introductory course such as CALS, there are not many opportunities to sustain this level of interest and further challenge the learners to deepen their

understanding of a literary text. Only by embedding the use of these techniques in a deeper context of a literature class, can the benefits of such efforts be genuinely reaped.

This study has been based on a group of 39 learners. There is an obvious need for further investigation, perhaps including a broader spectrum of learners than that investigated here and examining techniques that were not considered in this paper.

Works Cited

Abusa', Naser. "The Response of ESL Students to the Use of Concordance Output as Instructional Material." MA Thesis. International Islamic University Malaysia, Kuala Lumpur, 1997.

Addington, Ann. "Talking About Literature in University Book Club and Seminar Settings." *Research in the Teaching of English* 36 (2001): 212-48.

Adolphs, Svenja. *Introducing Electronic Text Analysis*. London and New York: Routledge, 2006.

Barkhuizen, Gary. "Discovering Learners' Perceptions of ESL Classroom Teaching/Learning Activities in a South African Context." *TESOL Quarterly* 32.1 (1998): 85-108.

Brooks, Cleanth and Robert Warren. *Understanding Poetry: An Anthology for College Students*. New York: Henry Holt, 1938.

Butler, Ian. "A Brighter Future? Integrating Language and Literature for First-year University Students." *Literature in Language Teaching and Learning*. Ed. Amos Paran. Mattoon, IL: TESOL Inc., 2006. 11-25.

Gilroy, Marie and Brian Parkinson. "Teaching Literature in a Foreign Language." *Language Teaching* 29 (1997): 213-25.

Holmes, D. and R. Forsyth. "The 'Federalist' Revisited: New Directions in Authorship Attribution." *Literary and Linguistic Computing* 10.2 (1995): 111-27.

Inaki, Akiko and Tomoko Okita. "A Small-corpus-based Approach to Alice's Roles." *Literary and Linguistic Computing* 21.3 (2006): 283-94.

Johns, Tim. "Whence and Whither Classroom Concordancing?" *Computer Applications in Language Teaching*. Ed. Theo Bongaerts et al. Foris: Dordrecht, 1988. 9-27.

Leech, Geoffrey. "Teaching and Language Corpora." *Teaching and Language Corpora*. Ed. Anne Wichmann et al. London and New York: Longman, 1997. 1-23.

Louw, Bill. "The Role of Corpora in Critical Literary Appreciation." *Teaching and Language Corpora*. Ed. Anne Wichmann et al. London and New York: Longman, 1997. 240-51.

Ma, Bruce. "Learning Strategies in ESP Classroom Concordancing: An Initial Investigation into Data-driven Learning." *Entering Text*. Ed. Lynn Flowerdew and Anthony Tong. Hong Kong: University of Science and Technology, 1994. 197-214.

McEnery, Tony and Andrew Wilson. *Corpus Linguistics*. Edinburgh: Edinburgh UP, 2001.

Mealand, David. "Measuring Genre Differences in Mark with Correspondence Analysis." *Literary and Linguistic Computing* 12.4 (1997): 227-45.

Nora Nasir. "ESL Learner Difficulties in the Malaysian Literature Classroom." PhD Thesis. University of Strathclyde, Scotland, 1997.

Nuraihan Mat Daud and Zamnah Husin. "Developing Critical Thinking Skills in Computer-aided Extended Reading Classes." *British Journal of Educational Technology* 35.4 (2004): 477-87.

O'Keefe, Anne, Michael McCarthy and Ronald Carter. *From Corpus to Classroom*. Cambridge: Cambridge UP, 2007.

Paran, Amos. "The Stories of Literature and Language Teaching." *Literature in Language Teaching and Learning*. Ed. Amos Paran. Mattoon, IL: TESOL Inc., 2006. 1-10.

Roberts, Alan. "Rhythm in Prose and the Serial Correlation of Sentence Lengths: A Joyce Cary Case Study." *Literary and Linguistic Computing* 11.1 (1997): 33-39.

Römer, Ute. "Where the Computer Meets Language, Literature, and Pedagogy: Corpus Analysis in English Studies." *How Globalization Affects the Teaching of English: Studying Culture through Texts*. Ed. Andrea Gerbig and Anja Müller-Wood. Lampeter: E. Mellen Press, 2006. 81-109.

Rommel, Thomas. "'So Soft, So Sweet, So Delicately Clear.' A Computer-assisted Analysis of Accumulated Words and Phrases in Lord Byron's Epic Poem *Don Juan*." *Literary and Linguistic Computing* 9.1 (1994): 7-12.

Rosenblatt, Louise. *Literature as Exploration*. New York: D. Appleton-Century, 1938.

Salaberry, Michael. "The Use of Technology for Second Language Learning and Teaching: A Retrospective." *The Modern Language Journal* 85 (2001): 35-56.

Scholfield, Peter and George Ypsilandis. "Evaluating Computer Language Learning from the Learners' Point of View." *Evaluating Language*. Ed. David Gradoll and Jan Swann. Clevedon: Multilingual Matters, 1994. 62-74.

"Stanford Students Use Digital Tools to Analyze Classic Texts." *Human Experience: Humanities at Stanford*. Stanford University, n. d. Online. 2 January 2011. http://humanexperience.stanford.edu/digitaltexts.

Stubbs, Michael. "Conrad in the Computer: Examples of Quantitative Stylistic Techniques." *Language and Literature* 14.1 (2005): 5-24.

-------. *Words and Phrases: Corpus Studies of Lexical Semantics*. Cornwall: Blackwell, 2002.

Tribble, Chris and George Jones. *Concordances in the Classroom*. London: Longman, 1990.

Vethamani, Malachi Edwin. "'CALL: Computer Applications for Literature Learning.'" *Internet Journal of E-Learning and Teaching* 1.2 (2004). Online. 30 June 2005.

Appendix

Sample Question on Section B of Questionnaire

Section B

This questionnaire is designed to help your teacher understand your experience in using corpus linguistics techniques in your literary analyses. Please think about the various analyses you have done using the three techniques you have learned in CALS (and remember that it shouldn't matter whether your analysis was successful or not). The three techniques are (i) word frequency analysis, (ii) part-of-speech analysis, and (iii) word collocation. For each technique, you must give three responses:

a. ENJOYMENT: Did you like using this particular technique in studying literature?
b. UNDERSTANDING LITERATURE: Did this technique help you to understand a literary text better?
c. BEYOND CALS: Do you think the technique will be useful for literary analyses outside this class?

Please circle your chosen answer. If you like, you can add your comment for the technique in the space provided.

1) word frequency analysis

ENJOYMENT

definitely	quite a bit	just a little	not at all

UNDERSTANDING LITERATURE

definitely	quite a bit	just a little	not at all

BEYOND CALS

definitely	quite a bit	just a little	not at all

Comment:_____
_____
_____
_____